SUPEREROGATION FOR A VIRTUE ETHICIST

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Claire Michelle Brown

________________________________________

W. David Solomon, Director

Graduate Program in Philosophy

Notre Dame, Indiana

January 2011
SUPEREROGATION FOR A VIRTUE ETHICIST

Abstract

by

Claire Michelle Brown

The contemporary discussion of supererogation and how it fits within particular moral theories is strikingly thin when it comes to the question of whether or how virtue ethics can make room for supererogation. Indeed, whereas the debates about how utilitarianism and Kantianism ought approach the supererogatory are well-developed, although by no means exhaustive, to date there is but one article-length treatment of the topic of virtue ethics and supererogation.

Given the above gap in the literature, the primary goal of this dissertation is to develop a satisfying account of supererogation that fits well within a virtue ethical framework. A second (and secondary) aim is related and enables the dissertation to make a modest contribution to the larger debate in moral philosophy about the relative strengths and weakness of the major moral theories. This second aim is to argue that the developed virtue ethics account is, as an account of the supererogatory, superior to the accounts of supererogation that have been offered on behalf of both Kantianism and utilitarianism. A third and minor aim is to show how virtue ethics can deepen our
understanding of the supererogatory, shedding light on certain aspects of actual cases of supererogation that the contemporary discussion of the topic tends to miss.

Chapter One provides an introduction to the concept of supererogation with a brief history of the academic discussion of it up to the present, introduces five “central cases” of supererogation, argues for the import of the supererogatory, and offers an outline of the project. Chapters Two and Three assess some of the more prominent utilitarian (Chapter Two) and Kantian (Chapter Three) accounts of supererogation that appear in the literature. Chapter Four introduces and provides a defense a new virtue ethics account of supererogation. Chapter Five brings together the accounts of supererogation introduced and assessed in Chapters Two, Three, and Four for the purposes of determining how successful utilitarians, Kantians, and virtue ethicist have been at accounting for supererogation.
For my parents
# CONTENTS

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

Chapter One: Introduction and the Idea of Supererogation .................................................................1
  1.1 A Brief History of Supererogation .............................................................................................. 5
  1.2 Supererogation Today ..............................................................................................................19
  1.3 Five Central Cases ..........................................................................................................................24
    1.3.1 The Grenade Case ..............................................................................................................24
    1.3.2 The Torture Case ..............................................................................................................25
    1.3.3 Paul Kocheleff's Work in Africa .....................................................................................25
    1.3.4 Babette's Feast ..................................................................................................................27
    1.3.5 Bishop Myriel and the Candlesticks ..............................................................................28
  1.4 Is Supererogation Possible? ...........................................................................................................32
  1.5 The Import of the Central Cases for Moral Theory ......................................................................37
  1.6 Thesis and Outline of the Project .................................................................................................43

Chapter Two: Beyond Maximizing: Utilitarianism and Supererogation ...........................................46
  2.1 Urmson and Basic Duty ..............................................................................................................47
  2.2 Attfield and Low Stakes Cases ...................................................................................................61
  2.3 Portmore and Total Adjusted Wellbeing .....................................................................................65

Chapter Three: Kantianism and Supererogation .................................................................................80
  3.1 Hill and Wide Imperfect Duties ..................................................................................................83
  3.2 Jokic and Moral Luck ..................................................................................................................96

Chapter 4: Virtue Ethics, Optional Virtue, and Supererogation ......................................................113
  4.1 Defining Virtue Ethics ................................................................................................................115
    4.1.1 The Priority Claim ...........................................................................................................115
    4.1.2 Two Types of Virtue Ethics ..............................................................................................123
  4.2 Supererogation for the Moderate Virtue Ethicist? ....................................................................126
    4.2.1 Jason Kawall's Ideal Observer Account .........................................................................127
    4.2.2 Better than Virtuous? ......................................................................................................135
    4.2.3 Lowering the Obligation Bar ............................................................................................139
    4.2.4 Magnificence and Optional Virtues .................................................................................144
  4.3 Virtue as Optional .......................................................................................................................152
  4.4 Defending the Optional Virtue Account ......................................................................................157
Chapter Five: Recognizing the Morally Extraordinary ...................................................173

5.1 Paring Down the Accounts .................................................................173
   5.1.1 The Utilitarian Accounts ............................................................174
   5.1.2 The Kantian Accounts .................................................................178
   5.1.3 The Virtue Ethics Accounts .........................................................184

5.2 A Side-by-Side Comparison ...............................................................186
   5.2.1 Fit with Moral Theory .................................................................187
   5.2.2 Sensitivity to the Central Cases ..................................................190
   5.2.3 Specificity ..............................................................................199
   5.3.4 Additional Considerations .........................................................204

5.3 A Victory for Virtue Ethics .................................................................206

5.4 What about the Radicals? .................................................................209

5.5 Extravagant Love in the Morally Extraordinary .................................212

Bibliography .................................................................................................219
ACKNOWLEDGMENTS

This dissertation would not have been possible without the contributions of many others. I thank my parents for introducing me to the beauty of books, valuing my education and encouraging me to do the same, and sacrificing so that I could receive a Christian liberal arts education. My high school English teacher, Mona Beverly, did more than anyone else to help me to learn how critically to edit my own writing, a skill not easily taught to high school honors students who often feel entitled to praise and exempt from serious criticism. I thank my Asbury professors for introducing me to the wonders of the liberal arts. My philosophy professors, Mike Peterson and Ray VanArragon, deserve particular thanks for inculcating in me a love of philosophy, for encouraging me to go to graduate school, and for giving me the tools to do so. Once I was there, the philosophy department at Notre Dame gave me an outstanding education, all the while making clear I was valued as both a student and a person. The support that my church and my friends gave me over the past several years has been invaluable, but I am particularly and deeply grateful for all that they did for me in the summer of 2010, when I needed and received encouragement and love of a sort I had never before known.

To turn more directly to the development of the actual document, I am extremely grateful to Robert Audi and Jim Sterba for so carefully reading multiple drafts of the chapters here included and, each time, offering new, insightful questions, criticisms, and
suggestions aimed at making my work both more cogent and more philosophically interesting. Their repeated generosity with their time and expertise is most appreciated. Many thanks, also, to Neil Delaney for agreeing to serve as a reader so late in the game.

Finally, the debt of gratitude I owe to my advisor, David Solomon, I will likely never be able to repay. When I started working with David about four years ago, my philosophy classes had done much to hone my analytic skills, but my philosophical work and thinking was, I now know, woefully shallow, and I was unable to appreciate much beyond a very narrow style of analytic philosophy. Through classes, reading suggestions, and, above all, much careful, patient discussion, David enabled me to see that, while analytical rigor is indispensible to good philosophy, it does not suffice for it. I am most grateful.

The above list is, of course, incomplete, but it does at least indicate that this single-authored work is by no means the product of only one person’s efforts. Still, in writing, there is an asymmetry between accomplishments and errors: for the former, others often deserve partial or even full credit; for the latter, the author has—and in this case explicitly recognizes—full responsibility.
CHAPTER ONE

INTRODUCTION AND THE IDEA OF SUPEREROGATION

The term, “supererogation,” is not one that English speakers frequently use. In my own experience, non-philosophers who hear that I am interested in “supererogation” are more likely to think that my work has something to do with agriculture and droughts than that it concerns moral excellence. Yet the concept of going above and beyond the call of moral duty and the associated concepts of moral saints and heroes are not foreign to contemporary English speakers. Gratitude toward the stranger who “went out of her way to do more than anyone could ask” and admiration of the strength of character of such people as Mother Theresa are common sentiments. Of course, sometimes respect for saints and heroes is of a non-moral variety—as is plausibly (but not obviously) the case when members of a society honor the patriotism of a war hero\(^1\) or when a religious community commemorates the way a martyr held fast to the faith. But on other occasions respect for saints and heroes is, at least apparently, of the moral variety, which

\(^1\) As when, for instance, the United States Congress bestows the “Congressional Medal of Honor,” the traditional purpose of which is to honor those who have “[distinguished themselves] conspicuously by gallantry and intrepidity at the risk of [their lives] above and beyond the call of duty while engaged in an action against an enemy of the United States...”. Department of the Army (2002-07-01), Section 578.4 “Medal of Honor.” Code of Federal Regulations, title 32, Vol. 2. Popular discussions of recipients (at least in the United States) suggest that the recipients have done more than gone above and beyond the call of military duty; they are morally exemplary as well.
would explain, in the case of Mother Theresa, why she is admired not only by Roman Catholics but also by people who are not at all religious. What these people most admire about her is not her religious devotion but her compassion for the suffering. Recognition of something very much like supererogation even shows up in the popular television series, *The Simpsons*. In one episode, Homer arrives at the gates of heaven only for St. Peter to tell him that if he wants to gain admission, he must go back to earth and do one good deed. When wife Marge suggests that he clean the garage and paint the house, Homer’s response is typical Homer: “Whoa, whoa, whoa! I’m just trying to get in. I’m not running for Jesus!” The joke, of course, is not, to use the language of moral philosophers, that Jesus is a moral saint whose actions were supererogatory; the joke is that cleaning the family garage and painting the family house might count as supererogatory.

Acceptance of the possibility of supererogation, then, appears to be widespread, even if use of the term is not. Yet moral philosophers have struggled to provide an account of it. In recent years, two distinct ways of providing one have emerged. One approach attempts to develop a generic account of supererogation—generic in the sense that the account does not presuppose that any particular normative theory (Kantianism, say) is correct. Rather, the account of supererogation developed is supposed to be correct regardless of which normative theory (if any) turns out to be true. The problem with this

---


general approach is that the accounts advocated tend to be too general to be useful, and it is often difficult to see what would be involved in meeting the conditions of the accounts. To take one example, Chisolm and Sosa’s generic account of supererogation says only that a supererogatory act is one whose “intentional performance is morally good” and “intentional nonperformance is morally neutral.”\(^4\) Two key questions, of course, are “under what conditions is an action’s intentional performance morally good?” and “under what conditions is an action’s intentional nonperformance morally neutral?” Without answers to these questions (which Chisolm and Sosa do not attempt to provide), the Chisolm/Sosa account is not particularly helpful. One might be tempted to think that it is the job of particular normative theories to provide answers to the previous questions. Unfortunately, most prominent normative theories do not have ready-made answers to both questions. Utilitarianism, for instance, does not have an obvious account of “morally neutral intentional nonperformance.” Other generic accounts encounter similar problems.

The second way that philosophers have attempted to develop accounts of supererogation produces accounts that are unique to particular normative theories. Thus, utilitarians have developed uniquely utilitarian accounts of supererogation, and Kantians have developed uniquely Kantian ones. On one Kantian account, for instance, supererogatory acts constitute a particular subset of acts of imperfect duty.\(^5\) The advantage of this second approach is that, in general, once we understand the normative


theory, if its account of supererogation is coherent, we should be able to apply it fairly easily to particular cases, borderline cases excepted. More significantly, ideally, the particular normative theory also provides the justification of the supererogatory status of the acts in question, telling us, for instance, why the acts that turn out to be supererogatory are so spectacular as to be “above and beyond the call of moral duty.” Thus, whereas the generic Chisolm/Sosa account could only insist that a given supererogatory action—a particular instance of Mother Theresa’s service to the dying, say—is morally good when intentionally performed and morally neutral when intentionally not performed, a utilitarian account could speak of the extraordinary utility of the particular action.

The advantages of the second approach, the one that develops an account of supererogation that is unique to particular normative theories, persuade me that this second approach is a helpful way to develop an account of supererogation. As someone with strong virtue ethics sympathies, I am interested in the particular account of supererogation that might be offered on behalf of virtue ethics, but to date little progress has been made in developing and defending such an account of supererogation unique to virtue ethics, although the topic has not been ignored completely. The primary goal of this dissertation is to provide such an account. Before I begin that project, however, it will be helpful first to gain a better understanding of what is involved in supererogation

---------------------------------------------------------------------

6 By contrast, I know of no cases that obviously satisfy the Chisolm/Sosa account of supererogation if either standard consequentialism or Kantianism is presupposed.

7 For a discussion of the recent progress made on this topic, see section 1.5 of this chapter.
by looking at the history of the concept and the contemporary discussion of it and by reflecting on some concrete cases that plausibly involve supererogation.

1.1 A Brief History of Supererogation

The first known use of the term, ‘supererogation,’ (or, more correctly, the first known use of the Latin verb from which the English term is derived) occurs in the Vulgate New Testament in the parable of the Good Samaritan. Notably, the term is not used in appraisal of the hero’s action. Rather, the Samaritan himself uses the term when he gives the innkeeper advance payment for tending to the injured traveler. The Samaritan tells the innkeeper that whatever he overspends (i.e., spends beyond the advance payment) to care for the injured man, the Samaritan will reimburse the innkeeper. The term for overspending is, ‘supererogaveris,’ a form of ‘supererogare,’ and is being used, in the context of the parable, in a purely financial—and not at all a moral or even a religious—sense.8

The moral/religious sense is thought first to have emerged in Christianity. Traditionally, it is associated with a distinction between God’s commands (or precepts) and God’s counsels (often called “the counsels of Christian perfection”). Obeying the former was thought to be required while fulfilling the latter was optional and supererogatory. Roman Catholics trace this distinction between command and counsel and the resulting possibility of supererogation (in the moral and the religious sense) to the

New Testament. They cite, for instance, the following excerpt from the Gospel of Matthew in which a wealthy man questions Jesus about the requirements of eternal life:

Then someone came to him and said, "Teacher, what good deed must I do to have eternal life?" And he said to him, "Why do you ask me about what is good? There is only one who is good. If you wish to enter into life, keep the commandments." He said to him, "Which ones?" And Jesus said, "You shall not murder; You shall not commit adultery; You shall not steal; You shall not bear false witness; Honor your father and mother; also, You shall love your neighbor as yourself." The young man said to him, "I have kept all these; what do I still lack?" Jesus said to him, "If you wish to be perfect, go, sell your possessions, and give the money to the poor, and you will have treasure in heaven; then come, follow me."9

Some Catholics have seen Jesus as, in this passage, making a distinction between keeping God’s commandments—obedience to which is required for eternal life—and living a life of poverty—which is necessary only for perfection, thereby allowing for supererogation in the case of the latter.10 Other Biblical passages cited in support of the possibility of supererogation are those in which the apostle Paul claims that, although he has a right to demand material sustenance from those to whom he is ministering, he chooses not to demand it so that the gospel of Christ will not be hindered (see, e.g., 1 Corinthians 9:1-19). Paul’s choice to forego such a right and to support himself by making tents even while he ministers to others is then interpreted as an instance of supererogation.

Of course, even if the Catholics in question are correct and the New Testament does distinguish between commands and counsels and imply the possibility of supererogation,11 it does not follow that the concept of supererogation was a common one

---

9 Matthew 19:16-21, NRSV. Italics mine.
11 There are reasons for questioning the traditional Catholic interpretation of the passages in question. To take the Matthew 19 passage, for instance, the remainder of the passage reveals that the rich
in early Christian thought; the possibility of supererogation could have been implied by scripture but not explicitly recognized until much later. In fact, some early church fathers did recognize the possibility of non-obligatory, meritorious acts that surpass God’s commands.\textsuperscript{12} We can see one example of such recognition in a homily by John Chrysostom on I Corinthians 9. Chrysostom says that Paul edified the Corinthians “by a sort of supererogation” when he declined to take anything from them (though he lawfully could have done so).\textsuperscript{13} Applying the passage to the people he is addressing, Chrysostom urges his listeners to avoid being satisfied with doing what is permitted, even if “Christ Himself” has granted the permission. Chrysostom wants his audience to follow Paul’s example and do more, to “go beyond the precepts.”\textsuperscript{14} Explaining the reason for the permissions, Chrysostom says,

Thus has our Lord in His mercy mingled much gentleness with His precepts that it might not be all merely of commandment, but that we might do much also of our own mind. Since it was in his power, had He not been so minded, to extend the

\begin{itemize}
\item man went away disappointed because he had much money, suggesting that he was not prepared to part with it. If parting with it were necessary only for being perfect and not also for eternal life, we might well wonder why the rich man was so disappointed. Reinforcing this reading is the fact that Jesus himself tells his disciples, following the rich man’s departure, “it is easier for a camel to go through the eye of a needle than for someone who is rich to enter the kingdom of God.” (v. 24, NRSV) The disciples then ask, “Then who can be saved?” (v. 25, emphasis mine) Jesus’ reply is “For mortals it is impossible, but for God all things are possible” (v. 26). All of this suggests that selling one’s goods and giving the proceeds to the poor is, according to Jesus, a requirement for salvation, not merely perfection. Those who are still not convinced should recall that Luke’s version of this story does not include reference to perfection at all; rather than saying, “if you wish to be perfect,” Jesus says “there is still one thing lacking” (Luke 18:22). To give the Roman Catholic interpretation its due, however, Jesus does make clear in the Luke passage that if the rich man does all Jesus says, including giving his wealth to the poor, then he will have “treasure in heaven.”
\end{itemize}

\textsuperscript{12} Heyd, \textit{Supererogation}, 18.


\textsuperscript{14} Chrysostom, \textit{Homily 21}, section 8.
commandments further... But he did not do so, giving you occasion, if you will, to be forward in doing more.  

To Chrystostom, then, we have the opportunity to do better than what is commanded, and this opportunity is a gift from God.  

While the early church fathers recognized the possibility of going beyond the commands, according to David Heyd, St. Ambrose and St. Augustine were the first explicitly to distinguish commands from counsels (of which poverty, chastity, and obedience were the most frequent examples).  

Such a distinction is philosophically significant because it provides theoretical resources for justifying the further distinction between required acts and the supererogatory acts that surpass them. The church now had an explanation available for (some) non-commanded but permitted acts having positive (as opposed to neutral) value: some non-commanded, permitted acts could have such value by virtue of being in accord with divine counsel.  

In the thirteenth century, Alexander of Hales is significant because he is recognized as the first to connect supererogation with the (already existent) church practice of indulgences.  

Indulgences are church-sanctioned exemptions from the penitential punishment required for venial sin. Hales suggests that the church is able to grant these exemptions because it has access to a treasury of merits (sometimes called the treasury of the saints) that has accumulated as a result of the supererogatory acts of Christ.  


and the church. The pope can then tap on this treasury to grant an indulgence, thereby ensuring that the recipient’s penitential debt for venial sin is reduced.\(^\text{18}\)

The most well-known theological discussion of supererogation belongs to another thirteenth-century theologian, Thomas Aquinas. Aquinas affirms the traditional view that supererogatory acts are ones that, while not commanded, are in accord with counsels.\(^\text{19}\)

For Aquinas, both commands and counsels are expressions of the will of God,\(^\text{20}\) but commandments, unlike counsels, are “obligatory” and, in a certain sense, “necessary.”\(^\text{21}\)

The key to understanding Aquinas’ distinction between commands and counsels—and by extension his distinction between obligatory acts and supererogatory ones—is to get a grip on the multiple senses in which commanded acts are necessary. Aquinas says both that their necessity is due to their obligatory nature as commanded\(^\text{22}\) and that fulfilling them is necessary for attaining our final end, salvation.\(^\text{23}\)

\[\text{Counselled acts, by contrast, are not necessary in either of these ways: God does not require that we fulfill his counsels,}\]

\[\text{———}\]

\(^{18}\) Hales, \textit{Summa Universae Theologiae}, IV, q. xxiii, m. 3, n. 6, referenced by Heyd, \textit{Supererogation}, 18-9.


\(^{20}\) Aquinas, \textit{Summa}, I, q. 19, a. 12.

\(^{21}\) Aquinas, \textit{Summa}, II-I, q. 108, a. 4.

\(^{22}\) Aquinas, \textit{Summa}, II-II q. 58, a. 3.

\(^{23}\) Aquinas, \textit{Summa}, II-II, q. 187, a. 3. Aquinas’ discussion of 1 Corinthians 9 suggests a third sort of necessity that qualifies an act as being good but non-supererogatory: necessity for survival. Thus, Aquinas writes that whether or not Paul’s supporting himself turned out to be supererogatory depends on the actions of some of his contemporaries. If other people had not provided his livelihood, it would not have been supererogatory for him to work as a tentmaker to do so; instead, he would have been required to do so. See also Aquinas, \textit{An Apology for the Religious Orders}, trans. by J. Proctor (Westminster, MD: Newman Press, 1952), pt. II, chapter 1.
and we can attain our final end of eternal bliss without fulfilling them. By extension, supererogatory actions are neither required by God nor necessary for salvation. On other occasions, we learn that supererogatory actions are also not necessary for possession of virtue, for the soul’s union with God, or for merit. Commanded acts can serve all of these ends.

Despite the fact that supererogatory acts are not required for any of the above goods, Aquinas does provide considerations in their favor. In the first place, although supererogatory acts are generally non-obligatory, on some occasions supererogatory acts are obligatory: if the supererogatory act or way of life in question (e.g., poverty) is one that the agent has vowed to perform, then the act or way of life is obligatory even while (Aquinas seems to imply) supererogatory. Additionally, Aquinas sees special benefits in supererogation. Following Alexander of Hales, Aquinas believes that supererogatory works of penance contribute to the treasury of the saints on which the church can draw by

\[ \text{\footnotesize (References omitted for brevity.)} \]

\[ \text{\footnotesize \text{\textsuperscript{24}} Aquinas, Summa, II-II, q. 186, a. 5.} \]

\[ \text{\footnotesize \text{\textsuperscript{25}} Aquinas, Summa, II-II, q. 12, a. 1.} \]

\[ \text{\footnotesize \text{\textsuperscript{26}} Aquinas, Summa, II-II, q. 58, a. 3 and II-II, q. 81, a. 6.} \]

\[ \text{\footnotesize \text{\textsuperscript{27}} Aquinas, Summa, II-II q. 185, a. 6 and II-II q. 186, a. 2. When speaking of cases in which the agent undertakes an obligation to perform an action that, had the agent not undertaken the obligation, would have been supererogatory, Gregory Mellema comes to a different conclusion than Aquinas, saying that in such cases the would-be supererogatory action is not supererogatory (because it is obligatory) but that, nonetheless, the very act of undertaking the obligation is supererogatory. Mellema, Beyond the Call of Duty, 177. My own inclinations say that Aquinas is right, that performing the action itself—not promising to perform it—is what is supererogatory. The problem with Mellema’s view is that promising is too easy and insignificant an act to count as supererogatory. At most, the agent, in promising, is making himself liable to moral guilt if he does not perform the difficult action. But undertaking such a liability is hardly an extraordinary action. See also Aquinas, An Apology for the Religious Orders, pt. II, chapter 1.} \]

issuing indulgences. Moreover, the agent reaps significant benefits from supererogation because, according to Aquinas, God will reward supererogatory acts in a special way. Aquinas also seems to think that while the soul can be united to God without the aid of supererogation, a life of supererogation produces a special kind of union with God not otherwise available. Most famously, with regard to our final end of happiness, supererogation makes “the gaining of this end more assured and expeditious.” Despite all of these benefits, Aquinas cautions that supererogation is not for everyone, and for those who cannot handle abiding by the counsels, attempting to do so is counterproductive.

Supererogation came under attack during the Protestant reformation. Some of the Protestant objections to supererogation were aimed not at supererogation itself but at associated practices and beliefs, most notably the sale of indulgences, which makes sense only if the treasury of merit of the saints—and by extension the possibility of supererogation—is presupposed. But even the wholesale abandonment of indulgences would not have satisfied all of the Reformers’ objections to supererogation; some Protestants saw the very possibility of supererogation as incompatible with scripture. Given the overlap in reasons for objecting to supererogation among the Reformers—and virtually all of the major Reformers did object to the possibility of ordinary (i.e., non-


30 Aquinas, *Summa*, II-II q. 12, a. 1.


divine) humans performing supererogatory acts\textsuperscript{33}—I will not rehearse each of the Reformers’ individual positions on why supererogation is impossible. I will, however, attempt to explain each of the major Protestant objections to supererogation.

**Protestant Objection 1 (PO1): Belief in supererogation is motivated by sin (in particular, pride, arrogance, or sloth).**\textsuperscript{34} The thought here is that people who believe in supererogation come to this belief because, consumed by pride, they believe that they do more than they are obligated to do, or because their arrogance prevents them from holding that they are sinful when they fail to do what is best, or because they are slothful and do not want to admit that they are obligated to do a number of very difficult tasks. At first blush, this objection appears to be an *ad hominem* attack on those who believe that supererogation is possible. And so it is. But, notably, PO1 also undercuts the justification for belief in supererogation. Pride, arrogance, and sloth, after all, are unreliable guides with respect to truth. A slothful or arrogant person’s beliefs about supererogation run a serious risk of being instances of wishful thinking. Pride is a particularly epistemically dangerous character trait because it involves a false belief about the self—a belief that the self is better than it is. Hence, it behooves those who

\textsuperscript{33} A notable exception is Melanchthon, who, according to Mellema, thinks celibacy can be supererogatory. *The Expectations of Morality*, (Amsterdam: Rodopi, 2004), 109-110. According to Melanchthon’s *Apology for the Augsburg Confession*, trans. by F. Bente and W. H. T. Dau, in *Triglot Concordia: The Symbolical Books of the Evangelical Lutheran Church*, (St. Louis: Concordia Publishing House, 1921), however, “obedience, poverty, and celibacy, provided the latter is not impure, are, as exercises, adiaphora [in which we are not to look for either sin or righteousness].” Article XXVII.

have these misleading character defects to think twice before endorsing initial inclinations in favor of supererogation. If the Reformers are right, we all have serious problems with these vices, so belief in supererogation should be subjected to scrutiny. Of course, it does not follow that all belief in supererogation is in error.

**Protestant Objection 2 (PO2): The distinction between commands and counsels fosters immorality.** The Reformers recognized two ways in which the distinction between commands and counsels—a distinction that under-girded the Catholic Church’s doctrine of supererogation—promotes immorality. First, the distinction makes us fail to recognize the gravity of sin. Those who accept the distinction are likely to assume that required actions are only recommended and so excuse themselves from them and thereby fall into sin. The second way that the distinction fosters immorality is that it produces pride, which is essentially sinful. When people believe in supererogation, they are likely to take themselves, at least on some occasions, as having performed supererogatory acts or as having fully satisfied God’s requirements. Calvin goes so far as to categorize such behavior as blasphemy. On the face of things, both forms of PO2 beg the question against the supererogationist: if the distinction between commands and counsels is legitimate, it will not lead to sin if it simply leads to a lax attitude toward the counsels or to the agent’s true belief that she has, in some limited way, done more than God has required. But we can also see how, even if the distinction is legitimate, belief in


36 Calvin, *Institutes*, Vol. 1, 780-81. Calvin does consider the response that supererogation might be possible even if one is not perfect. Calvin’s reply is that one still cannot supererogate because one cannot do anything good that is not commanded: all good things are commanded because all good things fall under either love of neighbor or God, a point I discuss later under (what I call) PO4.
it may lead to a lax attitude toward commands that individuals falsely perceive to be matters of counsel or to pride when individuals overestimate the extent of their supererogatory acts. In any case, PO2 does not challenge the truth of the doctrine of supererogation; it rather points out that acceptance of the doctrine may be morally dangerous.

**Protestant Objection 3 (PO3): The distinction between commands and counsels is arbitrary.** As we have seen, the Catholic Church supported the distinction by appeal to scripture. In response, the Reformers tried to show that the Catholic interpretation of the relevant portion of scripture was incorrect\(^\text{37}\) and that, even if a distinction between the supererogatory and the obligatory were legitimate, scripture provides no way of knowing, of all the good acts it endorses, which are required and which are supererogatory.\(^\text{38}\) The Roman Catholic Church, then, the Reformers insisted, makes its distinction not by following scripture but out of convenience, deeming the so-called “Counsels of Christian Perfection” supererogatory because doing so elevates the status of the clergy.\(^\text{39}\) Calvin, Luther, Melanchthon, and Zinzendorf went so far as to say that many of the particular actions and ways of life that the Roman Catholic Church saw as matters of counsel (ones, e.g., having to do with pilgrimages, penance, and monasticism), far from being supported by scripture as supererogatory, were often not


\(^{38}\) To see how pressing a problem this is for anyone who takes the Christian scriptures as a source of moral knowledge, consider Matthew 5:48, in which Jesus is reported as saying, “Be perfect, therefore, as your heavenly Father is perfect,” NRSV. Neither the context nor the words make clear whether Christ is commanding or only counseling perfection.

even good but rather useless, foolish, silly, or even sinful and thus were
incorrectly deemed “supererogatory.”

Protestant Objection 4 (PO4): The doctrine of supererogation contradicts
scripture. We have already seen (in our discussion of PO3) that the Reformers directly
responded to claims that particular passages in the Christian Bible imply the possibility of
supererogation. Not satisfied with a negative argument from scripture, the Reformers
went on offence, claiming that scripture clearly implies that supererogation is impossible.
One of the passages they frequently cite is Luke 17: 7-10, in which Jesus teaches the
following:

Who among you would say to your slave who has just come in from plowing or
tending sheep in the field, 'Come here at once and take your place at the table'? Would you
not rather say to him, 'Prepare supper for me, put on your apron and serve me while I eat and
drink; later you may eat and drink'? Do you thank the slave for doing what was commanded? So you also, when you have done all that you were ordered to do, say, 'We are worthless slaves; we have done only what we ought to have done.'

Given that the above passage speaks of commands, teaching that followers of Christ
should not boast or expect any thanks when they do what they are commanded to do, it
may seem odd that the Reformers would take it as contradicting the possibility of

40 Zinzendorf: “Hence come the works of supererogation, the many ceremonies, ecclesiastical
regulations which are unquestionably twice as difficult as to follow the Saviour; and they have no promise
for this life or the life to come,” quoted in William Harvey-Jellie and Frederick William Brown, The

41 Calvin, Institutes, Vol. 1, 782; Calvin, Tracts and Treatises, Vol. 1, 164.

42 Melanchthon, Apology for the Augsburg Confession: Art. XVII.

43 Luther, “Sermon for the Eighth Sunday after Trinity; Matthew 7:15-23,” Luther’s Church
“Of Monastic Vows.”

44 Luke 17:7-10, NRSV.
supererogation, which the tradition taught is a matter of fulfilling the counsels, not merely the (obligatory) commands. Yet the Reformers did appeal to this passage to argue against supererogation. Calvin takes it as teaching that we should examine ourselves on the question of whether we give God more than we owe him; when we do that honestly, we will see that the answer is “no.”45 So, we should all recognize ourselves as “unworthy servants,” not as agents of supererogation.46 Calvin recognizes the objection that a person could be guilty of sin and yet still perform acts of supererogation on particular occasions, but his response is that a good but not required action is in principle impossible. Calvin provides two (scripture-based) reasons for the in principle impossibility: (1) all good actions either fall under the rubric of honoring God or of loving neighbor, and both types of action are commanded47 and (2) God owns everything, so we could never give him more than he is due.48

45 Calvin, Institutes, Vol. 1, 781.
46 Calvin, Institutes, Vol. 1, 781.
47 Calvin, Institutes, Vol. 1, 781. Calvin’s actual words are: “there is nothing that can come to mind which contributes to the honoring of God or the love of neighbor that is not comprised within God's law” and hence obligatory. Although Calvin does not mention this point, it is worth drawing attention to the extent of love of God and neighbor that the Christian scriptures make clear is a matter of command:

One of the scribes came near and heard them disputing with one another, and seeing that he answered them well, he asked him, “Which commandment is the first of all?” Jesus answered, “The first is, 'Hear, O Israel: the Lord our God, the Lord is one; you shall love the Lord your God with all your heart, and with all your soul, and with all your mind, and with all your strength.' The second is this, 'You shall love your neighbor as yourself.' There is no other commandment greater than these” (Mark 12:28-31, NRSV, emphasis mine).

Thanks to James Sterba for drawing my attention to the radical nature of the second commandment.

48 Calvin, Institutes, Vol. 1, 782.
Protestant Objection 5 (PO5): The doctrine of supererogation leads to the abuse of indulgences. The connection between supererogation and indulgences may be the primary reason that Luther and his followers so strongly opposed supererogation. Note that, for Lutherans, “the abuse of indulgences” would not be limited to events as vulgar as, say, the auctioning off of get-out-of-Purgatory-free cards to the highest bidders. Due to their commitment to salvation by grace through faith and not by works, Luther and his followers opposed any scheme that claimed to transfer merit to guilty sinners from a church-controlled treasury of merit. For Luther, then, indulgences were an affront to the redeeming work of Christ on the cross and hence essentially abusive.

Now, the possibility of supererogation does not imply the legitimacy of indulgences; supererogation is possible even if it does not contribute to a treasury of merit to which the church can appeal to reduce penitential debt. So, even if Luther were right about salvation by faith and its implications for indulgences, PO5, like PO2, does not challenge the reality of supererogation; it merely points to a bad consequence associated with its acceptance.

Supererogation fell on hard times following the Protestant Reformation and was largely ignored in the moral philosophy of the modern period, occasional mentions of it


50 Calvin agreed and even thought that supererogation itself—not merely the associated practice of indulgences—was incompatible with salvation by grace, *Institutes*, Vol. 1, 779-82.

51 In light of this fact, it is interesting that one of the many derogatory names given to Wesley and his followers (alongside “Bible-Moths” and the longer-lasting “methodists”) was “Supererogation-men.” Mark Noll, *The Rise of Evangelicalism: The Age of Edwards, Whitefield, and the Wesleys*, (Downers Grove, IL: InterVarsity Press, 2003), 68.
A resurgence of interest in supererogation occurred in 1958 with the publication of J.O. Urmson’s “Saints and Heroes,” an article that saw a problem in contemporary moral philosophy’s lack of attention to morally saintly and heroic action as a result of its focus on the permissible, the obligatory, and the wrong. Urmson noted that Kantianism, utilitarianism, and intuitionism initially seem to crowd out moral saints and heroes: on Kantianism, the only acts that have moral worth are required by the categorical imperative; intuitionists focus exclusively on concepts of duty; and to utilitarians, extreme self-sacrifice is often obligatory. While Urmson was unconvinced that those theories could not (with acceptable alterations) accommodate supererogatory acts, he was emphatic that the major ethical theories had thus far neglected to attempt to do so and would remain unacceptable until they did. Urmson’s article sparked a discussion of supererogation that continues to the present day. Unsurprisingly, this discussion has been, unlike the discussions of supererogation that preceded it, largely secular.


53 In fact, Urmson was interested not only in morally saintly or heroic actions, which are by definition extraordinary, but also any action whatsoever that surpasses moral duty, Urmson, “Saints and Heroes,” in Essays in Moral Philosophy, A.I. Meldon, ed., (Seattle: University of Washington Press, 1958), 205.

54 Urmson, 214-5.

55 Heyd (as well as others) have pointed out that the word, “supererogation,” does not appear in Urmson’s article (see Heyd, Supererogation, 114). They are correct. Urmson does, however, use the phrase, “supererogatory act,” 214, which is why I have characterized him as saying something about supererogatory acts.

56 See Urmson, 205; 207-9; 214-15. Urmson was particularly optimistic that utilitarianism could accommodate morally saintly and heroic action.
1.2 Supererogation Today

In the contemporary discussion of supererogation, the consensus is that

(1) The supererogatory is not a matter of the agent’s moral duty.\(^{57}\)

And

(2) The supererogatory somehow surpasses ordinary morality, i.e., acting in a supererogatory manner in a particular situation is somehow morally better\(^{58}\) than doing only what (if anything) is morally required in that situation.\(^{59}\)

Even (1) and (2), however, have their few detractors. What is far more controversial is whether the following is true:

(3) Supererogation is possible.

Other controversial claims include:

(4) Trivial or nearly trivial examples of supererogation are possible.

(5) All acts of supererogation are done for the good of others.

(6) All supererogatory acts are freely performed.

\(^{57}\) How best more-precisely to formulate (1) is a difficult issue. The point is not simply that supererogatory acts are not specifically required, for very few acts are specifically required even though many more are “merely” fulfill duty. We can discharge most of our obligations in a number of mundane ways, no particular way being specifically required. But (1) also does not imply that a supererogatory act never fulfills a duty, for a supererogatory act can fulfill a duty if it fulfills it in a supererogatory way (as, e.g., if I discharge my duty to do you a favor by doing you a particular sort of favor). See footnote 27 of this chapter for more on this issue.

\(^{58}\) Note that this implies that supererogatory acts have moral value, a point with which Mellema (Beyond the Call of Duty, 29) and Urmson (199) both agree. It is also a thesis David Heyd’s “continuity condition of supererogation” (Supererogation, 5-6 and 131-6) entails. Heyd’s continuity condition states that the value of supererogatory acts is continuous with that of morally obligatory acts. It is unclear whether Heyd’s continuity condition is stronger than the simple claim—which it clearly entails—that supererogatory actions, like morally obligatory actions, have moral value. He may mean that supererogatory and morally obligatory actions share a more specific type of moral value.

\(^{59}\) As Julia Driver points out, the statement that supererogatory actions are always morally superior to obligatory actions is not only non-trivial; it is not even true. Sometimes, enduring torture is obligatory, but that does not mean that such endurance is morally inferior than every possible supererogatory action. Nonetheless, in a particular situation, acting in a supererogatory manner is morally superior to merely doing what duty requires—which, depending on the situation, may be nothing. “The Subererogatory,” Australasian Journal of Philosophy, 70 (1992): 290.
(7) Only actions are capable of falling under the heading, ‘supererogatory.’

(8) Agents performing supererogatory acts always regard their own actions as obligatory.\(^{60}\)

According to (4), an act need not be extraordinary in order to qualify as supererogatory; actions that are only minimally morally superior to what (if anything) is morally required in a given situation can count as supererogatory. Thus, letting a stranger cut line at the grocery store or donating an extra dollar to famine relief might count as an instance of supererogation.\(^{61}\) Chisolm, Sosa, Driver, Attfield, and Kawall all endorse (4),\(^{62}\) and in a similar vein Urmson says that what interests him is the neglected category of action that surpasses duty—not necessarily only the sort of action that we would normally consider so great as to be morally saintly or heroic.\(^{63}\) By contrast, Patricia McGoldrick insists that “the distinguishing feature of a supererogatory act is that it is performed at extreme risk to one’s own life and well being,” which hardly allows for supererogatory door-openings.\(^{64}\) Moreover, the common, quick definition of supererogation as “going above and beyond the call of moral duty” connotes more than slightly surpassing moral requirements: when we say that someone goes “above and

\(^{60}\) Conversely, the moral saint as self-deprecating servant who insists she has done no more than her duty may be more a creation of idealistic thinking than a reflection of sociological reality.

\(^{61}\) I do not mean to imply that those who hold (4) are committed to non-vague accounts of moral obligation so specific that there is a specific dollar amount each person is obligated to give to famine relief.


\(^{63}\) Urmson, 205.

“beyond the call of duty” in some context we mean that, with reference to duty in that context, the action in question is extraordinary. Thus, putting in five minutes of unpaid overtime on a single occasion does not count as going above and beyond the call of duty at work. So, if the quick definition of supererogation is a good approximation of the concept we have in mind when we speak of supererogation, (4) is false. Of course, one option is that we have two different concepts—one that allows for trivial (or nearly trivial) examples and one that does not—and neither of the two is the only important one. Another option is that we have a single threshold notion that is such that the exemplars of this notion, but not all the acts that fall under it, are extraordinary.

Several philosophers have defended (5), some saying simply that supererogatory acts must be intended to benefit others. Others who endorse (5) go further and classify supererogatory acts as purely altruistic. On the other side of (5), Gregory Mellema and Jason Kawall have both provided examples of actions that, they claim, are supererogatory despite the fact that they are not done for nor do they benefit other people. Mellema’s example has no intended or actual beneficiary; Kawall’s examples are actions that require a great deal of self-sacrifice (for a time) but are still done for the agent’s own (long term) good, raising the question of whether self-regarding supererogatory actions are possible.


When (6) states that acts of supererogation are always freely performed, it means that whenever an agent’s action is supererogatory, the agent is exercising freedom of the will; (6) is thus not a simple restatement of the claim that agents are under no moral obligation to perform supererogatory acts, which is what some moral philosophers occasionally seem to mean when they say that acts of supererogation are “free.” While those moral philosophers who explicitly discuss the issue of whether freedom of the will is required for supererogation are more likely than not to say that it is,\(^{68}\) several of the accounts of supererogation put forward in the contemporary literature nonetheless appear to license as supererogatory actions that are not free.\(^ {69}\) Moreover, at least one writer on the subject—Bashar Haydar—has explicitly stated that forced supererogation is possible.\(^ {70}\)

With regard to (7), Heyd, at least, has pointed out that “supererogation is primarily attributed to acts or actions rather than to persons, traits of character, motives, intentions, or emotions.”\(^ {71}\) The question remains, though, whether any such non-actions

\(^{68}\) See, for example, how Heyd has insisted on the freedom requirement on multiple occasions, including in Supererogation, 115; and in “Moral Subjects, Freedom, and Idiosyncracy,” in Human Agency: Language, Duty, and Value, Dancky, Moravcsik, and Taylor, eds (Stanford, CA: Stanford University Press, 1988), 166.

\(^{69}\) If an account makes no mention of freedom, frequently the question of whether or not it allows for non-free supererogatory action depends on whether freedom is required for moral praiseworthiness or positive moral value. If freedom isn’t so-required, it will be possible to have an act that is supererogatory but not free.

\(^{70}\) “Forced Supererogation and Deontological Restrictions,” The Journal of Value Inquiry, 36 (2002): 445-54. Nancy Stanlick has also claimed that some supererogatory actions are non-voluntary, but Stanlick is presupposing an unusual understanding of supererogation: while she never explains what she means by “supererogatory,” she says that some supererogatory actions are not morally good and in fact, are not even permissible. In any case, Stanlick holds that for a supererogatory act to be good, the agent who performs it must be free. “The Nature and Value of Supererogatory Action,” Journal of Social Philosophy, 30 (1999): 217-8.

\(^{71}\) Heyd, Supererogation, 1.
could ever qualify as supererogatory. In the end, the dispute over (7), like the dispute over (4), may simply be a reflection of the fact that we have (at least) two concepts to which we refer by using the single word “supererogatory.” One such concept applies only to actions; the other may apply to actions, character traits, ways of life, whole lives, persons, and so forth.

When Urmson introduces discussion of saints and heroes, he admits what is not at issue in the controversy over (8)—that, sometimes, and perhaps often, moral saints and heroes see the extraordinary actions that they perform as obligatory even if their actions are actually beyond duty. What is at issue in (8) is whether this is always the case. Note that, to endorse (8), one needn’t have the view that moral saints and heroes always see their (in fact) supererogatory actions as obligatory for everyone, the type of action that all people in similar situations would be obliged to perform. One could instead hold with Heyd that in cases of supererogation the agent has higher moral standards for himself than he does for others, that he sees himself as having moral obligations that others do not, including the obligation to perform the act in question. The important point is that those who agree with (8) believe that agents who perform supererogatory acts always see themselves as under an obligation to perform those acts, whether or not they see others as under a similar obligation. For his part, Urmson is confident that even though saints and heroes often see their extraordinary acts as obligatory, this view of their own actions is only temporary and prospective; after performing such great actions, “only


a modesty so excessive as to appear false could make [the moral saint or hero] say, ‘I only did my duty,’ for we know, and he knows, that he has done more than duty requires.”

1.3 Five Central Cases

I am not prepared to settle the debates surrounding (4)-(8), but I am interested in getting a stronger grip on supererogation and what it involves than we currently have, which may put us in a better position to evaluate (4)-(8). One way to get a stronger grip is to reflect on apparent cases of supererogation. Consider the following cases, which I will call ‘central cases’ of supererogation because if the supererogatory is a non-empty class, these cases are ones that common sense strongly recommends be included in that class.

1.3.1 The Grenade Case

Urmson himself is responsible for introducing the Grenade Case (which I will call “Grenade,” for the sake of brevity). It involves a group of soldiers conducting a training exercise using live grenades. During the exercise, one of the grenades accidentally lands close to the group, threatening to explode and kill the entire group in short order. One of the soldiers smothers the grenade with his own body, thereby protecting the other soldiers from the blast. We can make the soldier’s action even


75 Urmson, 202-3.
more plausibly supererogatory by (a) adding that the soldier who falls on the grenade has an escape route, but he forgoes the escape route to protect his comrades \(^76\) and (b) specifying that the soldier who falls on the grenade is not the only one with the opportunity to do so.

1.3.2 The Torture Case

Gregory Mellema has developed this case (*Torture*, for short) in which terrorists capture a man and insist that he renounce his own country and swear allegiance to them. When the prisoner refuses, they torture him. Despite the torture and despite the fact that the prisoner knows both that the terrorists are the only ones who will know of his decision and that his refusal to do as they say will not lead to better actions on their part, the prisoner remains true to his principles, never renouncing his country or swearing allegiance to the terrorists.\(^77\) The prisoner’s resistance in the face of extreme pressure makes this case such a strong candidate for supererogation.

1.3.3 Paul Kocheleff’s Work in Africa

One of Urmson’s own examples of supererogation involves a doctor who travels to a plague-ridden city to help the inhabitants. Urmson’s doctor is hypothetical, but we

---

\(^{76}\) This additional feature of *Grenade* was inspired by an actual event. Michael Monsoor was a Navy SEAL who was awarded the Congressional Medal of Honor posthumously on April 8, 2008, for smothering a live grenade with his body in order to protect others. Monsoor was a combatant in so-called “Operation Iraqi Freedom.” When a live grenade thrown in his direction hit him in the chest, he was close enough to an exit to escape, but others were not. Before the grenade exploded, he called out to warn those around him and then fell on the grenade, an act that resulted in his own death but saved the lives of two of his fellow SEALs and three Iraqi soldiers, Mike Mount, “Navy SEAL Paid Ultimate Price to Save Buddies,” April 9, 2008, cnn.com.

can point to an actual person, Paul Kocheleff, who has done something similar. Kocheleff, a physician originally from Belgium, saw his first patient with AIDS when he was thirty-eight years old and working in Burundi in the 1970’s. In subsequent years, he would see and treat more HIV and AIDS patients there. Once he recognized the cases for what they were, he worked both to persuade others that AIDS had a real presence in the area that demanded a response and to develop a program offering medical and psychological help for the victims. The effectiveness of this program has been documented. Meanwhile, those in power denied that the problem existed, even prohibiting Kocheleff and his colleagues from publishing the results of their 1985 study, which suggested that 10.2% of Burundians over the age of fifteen were HIV positive. When civil war broke out in 1994, Kocheleff finally left Burundi.

Kocheleff’s fight against AIDS in Africa continued in 1996, when he moved to Pietermaritzburg, South Africa. In Pietermaritzburg, he again faced opposition, this time in the form of bureaucracy, prejudice (of multiple forms), apathy, and denial. Such problems combined with the devastating nature of the disease and the poverty in the region resulted in his feeling, at times, like he was (in his words) “hunting dinosaurs with a watergun.” Yet Kocheleff’s efforts have not been in vain. He has not saved South

---

81 Paul Kocheleff, “AIDS in Burundi and South Africa,” 149.  
82 Gerald Oppenheimer and Ronald Bayer, *Shattered Dreams*, 196.
Africa from AIDS, of course, but his impact there has been significant: his achievements include establishing two HIV clinics, helping those clinics attract other long-term physicians, and playing a role in hundreds of poor patients beginning antiretroviral therapy.\textsuperscript{83}

1.3.4 Babette’s Feast

In the film, \textit{Babette’s Feast}, the title character serves an elaborate dinner for a small religious congregation of which she herself is not a member. The circumstances surrounding that meal—and features of the characters involved—are what make Babette’s course of action in preparing the feast such a strong candidate for supererogation. The film is set in a small, 19\textsuperscript{th}-century, Danish village. Babette is an accomplished French chef who, fleeing political persecution in her home country, shows up on the doorstep of two sisters and asks to work as their maid. The sisters, members of a very strict religious congregation, spend most of their time and money caring for the infirm in their community and cannot afford to pay Babette, but they agree to take her in. The sisters’ decision proves to be a fortuitous one as Babette’s skills and labor significantly improve both the economy of the sisters’ household and service and the recipients’ experiences of that service.

After living with and working for the sisters for fourteen years, Babette learns that she has won the French lottery. She then asks the sisters for permission to prepare a dinner for their upcoming church celebration. The sisters consent, but when they and

\textsuperscript{83} See Gerald M. Oppenheimer and Ronald Bayer, \textit{Shattered Dreams}, esp. 70-1, 140, 196, 207-8, 225-6, and 232-3.
their fellow church members see some of Babette’s preparations over the next several days and suspect that her plans involve more extravagance than their puritanical impulses deem appropriate, they all decide amongst themselves to eat the meal stoically pretending “as if [they] never had the sense of taste.” They hold to their plan at first, but thanks to an unexpected guest who, unaware of the plan, is vocal in expressing his appreciation for the quality of the food and drink, all of the guests eventually come to enjoy the meal. What is more, the feast proves to be an impetus for reconciliation within the congregation, which had been plagued by internal squabbles and grudges for some time. After the guests go home, the sisters ask Babette when she plans to leave them, assuming she will return to France; Babette responds that she cannot leave—she has nowhere to go and no money with which to go, having spent all of her winnings on that night’s dinner.84

1.3.5 Bishop Myriel and the Candlesticks

*Les Miserables* holds a number of memorable characters, and not the least of these is Bishop Myriel, also called Monsieur Bienvenue, a man who is the epitome of kindness and generosity. The Bishop devotes over 94% of his income and travel allowances to various charitable purposes, living on only one thousand francs a year,85 but he does permit himself one extravagance—a valuable set of silver utensils and candlesticks. One evening, Bishop Myriel hears a knock on his door. The source is Jean Valjean, a recently-released convict passing through Myriel’s village looking for lodging.

84 Gabriel Axel, director, *Babette’s Feast* [Babettes gaestebud], (Copenhagen: S Panorama Film International, 1987).

No one else in the town will take him in because, although Valjean has money, his passport calls him—“not unfairly,” as we are told—“highly dangerous.” Valjean explains his situation, and Myriel invites him in for the night, offering him dinner (served with the silver, of course) and a bed, both of which Valjean accepts. At two o’clock in the morning, Valjean wakes up, steals the silverware, and leaves with it. When Myriel learns what Valjean has done, he immediately acquits him of any wrongdoing, saying “I have long wrongfully held back this silver, which belonged to the poor. Who was this person? evidently, a poor man.”

Later that morning, the police bring Valjean to Myriel: they have found Valjean with Myriel’s silverware and have come to him to verify that it was stolen. When Myriel sees Valjean with the police, however, the bishop’s reaction recalls that of the father in the parable of the prodigal son: “‘Ah! there you are,’ [Myriel] said, looking at Jean Valjean. ‘I am glad to see you. Why, I gave you the candlesticks too, which are also silver and will fetch you 200 francs. Why did you not take them away with the rest of the plate?’” The bishop goes on to press the candlesticks into Valjean’s hands.

After the officers leave, Myriel’s interaction with Valjean becomes at once more urgent and more intimate:

…the bishop walked up to [Valjean] and said in a low voice:--
“Never forget that you have promised me to employ this money in becoming an honest man.”
Jean Valjean, who had no recollection of having promised any thing, stood silent.
The bishop, who had laid a stress on these words, continued solemnly:--

86 Hugo, 73.
87 Hugo, 78.
88 Hugo, 79.
“Jean Valjean, my brother, you no longer belong to evil, but to good. I have bought your soul of you. I withdraw it from black thoughts and the spirit of perdition, and give it to God.”

The candlestick incident proves transformative for Valjean, and he does indeed go on to live up to Myriel’s exhortations. It is Myriel’s manner of forgiving Jean Valjean by lavishing on him the additional gift of the candlesticks—and not merely the fact that Myriel forgives him—that prompts me to call this case, which I shall call Candlesticks, a central case of supererogation.

Reflection on these central cases is useful for a reason that I have already intimated: doing so provides some resources useful for settling the truth (or falsity) of (4)-(8). To be sure, reflection on these cases will probably not be helpful for generating a verdict on each of (4)-(8). Nonetheless, when we consider (5), for instance, reflection on one of the cases—Torture—suggests that it is false. Recall that (5) states that all acts of

---

89 Hugo, 80.

90 Some may complain that, far from being an instance of supererogation, Myriel’s manner of forgiving Valjean is condescending, particularly Myriel’s comments about buying Valjean’s soul and giving it to someone else (God). Two features of the story argue against that interpretation, although neither feature provides a decisive argument. First, Jean Valjean does not himself interpret Myriel’s actions as condescending. In fact, he so values what Myriel does for him that he refuses to sell the candlesticks and keeps them for the rest of his life. Second, Myriel is not insensitive to the problem of “helping” someone in a patronizing way, a fact that we see the night that the two meet. While they are eating dinner, Myriel gives Valjean some useful information about the part of the country to which he is headed, and “dwell[s] a good deal on the comfortable position of the grurin, as if wishful to make the man understand, without advising him directly and harshly, that it would be a refuge for him,” 62, emphasis mine. In addition, we learn that Myriel, “during the whole of supper and, indeed, of the evening, did not, with the exception of some words about the Saviour, when he entered, utter a word which could remind [Valjean] of what he was, or tell him who [Bishop Myriel] was. It was apparently a good opportunity to give him a little lecture, and to let the bishop produce a good effect on the galley-slave. It might have seemed to anyone else that, having this wretched man in hand, it would be right to feed his mind at the same time as his body, and to address him with a few reproaches seasoned with morality and advice, or at any rate a little commiseration, with an exhortation to behave better in the future,” 62-3. Bishop Myriel, however, saw fit to do none of these things.

91 We cannot expect the central cases to help us with (4) for we cannot expect any central case of supererogation to be trivial.
supererogation are done for the good of others, but the prisoner’s extraordinary resilience in adversity and opposition to terrorism are what make *Torture* such a compelling case for inclusion among the central cases of supererogation. The prisoner’s lack of an intention to benefit others is irrelevant. The central cases *may* also help us with (7), which says that only actions fall under the label, ‘supererogatory.’ It is, after all, Paul Kocheleff’s decades-long way of life that appears supererogatory, not some discrete action such as boarding an airplane to South Africa or treating his 211\textsuperscript{th} patient. Still, the defender of (7) might simply say that what is supererogatory is a (large) set of Kocheleff’s actions and that the label “supererogatory” applies to these actions collectively, if not to all of them individually. For now, I leave it to the reader to judge the success of such a reply.\textsuperscript{92}

In the case of (8), the claim that agents performing supererogatory acts always regard their own actions as obligatory, a number of the central cases are unhelpful because they are insufficiently worked out; we simply do not know how the agents involved regard their own actions (or, in the cases that are hypothetical, how real persons who performed such actions “would” regard them, if indeed they all would have the same attitude). In *Myriel* we get a hint that Myriel sees forgiving Valjean as obligatory—since the bishop reasons that he should not have held on to the silver utensils in the first place—but even in this case we are never told about how Myriel views the subsequent gift of the candlesticks. Conversely, *Babette’s Feast* (*Babette*, for short) may give us some indication that (8) is false for we are given no indication in the film that Babette

\textsuperscript{92} Some of what I say in Chapter Four and Chapter Five will be relevant to this issue.
feels obligated to do anything as elaborate as she does. Indeed, it would be quite surprising if she really felt that it was her duty to spend all of her winnings—and a week’s worth of labor, to boot—on a single meal for two of her friends and the quarrelsome members of their church.  

1.4 Is Supererogation Possible?

In addition to providing some help with (4)-(8), reflection on the central cases also provides motivation for developing an account of supererogation. As I mentioned above, (3), the claim that supererogation is possible, is itself controversial. The central cases speak to this controversy in that they are hard to read as anything other than as instances of supererogation. It is difficult to say, for instance, that the soldier in Grenade was merely doing his duty when he had an avenue of escape and some of the beneficiaries of his action who lacked such an escape could have fallen on the grenade themselves. But if the soldier was not merely doing his duty, given the extent of his sacrifice for others, how are we to read that sacrifice as having the merit that it does without saying that it counts as supererogatory? To take up Myriel, we can perhaps view the Bishop as obligated both to forgive Valjean for the theft and to try to protect Valjean from the police. Yet surely the Bishop’s doing this by giving him his only remaining possession of value goes above and beyond the call of moral duty. The appearance of supererogation in these central cases, then, is strong. Because it is so strong, the central

93 It is also not plausible that Babette sees herself as having a disjunctive duty to do this or else use all money to help others in some way. At most, she might, given her attachment to the sisters, see taking her winnings and returning to France as unthinkable, not because it would be base or wrong, but because France is no longer her home; her home is with the sisters.
cases provide support for the claim that supererogation is genuinely possible, even if instances of it are rare.

Despite the availability of the above examples, many moral philosophers have denied the possibility of supererogation. The most commonly-offered argument against supererogation is one that I will call the Moral Reason Argument (MRA). According to MRA, both moral requirements and moral merit are determined by a preponderance of moral reasons: we are morally obligated to do the morally best, which is whatever we have most moral reason to do. Since we are obligated to do the best, no one can do more than she is morally obligated to do. Because supererogation requires going beyond duty, supererogation is thus impossible.94

MRA is not, however, without problems. The most obvious worry for MRA is that it is not at all clear that morality is so demanding as to require that we do what is best, that is, that we be moral maximizers. Part of this worry comes from the intuition that, however demanding morality may be, it is not so demanding as to require our best at all times. Although I do not think we should assume, without argument or discussion, that this intuition is correct, it is an intuition that the defender of MRA must address if MRA is to be plausible; stipulating that we are required to do what is morally best will not do, given the existence of the intuition about the demandingness of morality.95


95 Even if we assume that the defender of MRA has some way of explaining away general intuitions about the limits of what morality can or does require of us, MRA’s assumption that we are obligated to be moral maximizers faces some further, less obvious worries, that are brought to light when we reflect on Robert Adams’ discussion of God’s obligations with respect to creation. If one believes that God has moral obligations, it is tempting to say that God was/is obligated to create the best possible world, assuming (as we will) that there is such a world and that God can create it. The strangeness of this belief
A second problem for MRA is a technical (but real) one: in the form presented, MRA only shows that supererogation is impossible in cases in which some action open to the agent is the morally best act, as determined by a preponderance of moral reasons. If in a given situation no potential act has a preponderance of moral reasons in its favor, then MRA, as formulated, says nothing about whether supererogation is possible in that situation and so cannot be used to show that supererogation is impossible simpliciter. Now, there are three reasons why, in a given situation, no potential act might have a preponderance of moral reasons in its favor. First, two or more potential acts might be tied for first in terms of moral reasons in their favor. Second, we could have a problem that is an analogue to the problem of the largest possible number or the best possible world: an agent’s options might be such that, for each option X open to her, she has some other option Y such that her moral reasons for Y are stronger than her moral reasons for X. Third, the agent’s moral reasons for some of her various options might be incommensurable with one another.

that God must create the best emerges when we try to articulate why God has this supposed obligation and to whom God owes the creation of this best possible world. At the time of creation, no one exists to whom God could “owe” creation, and that includes the potential inhabitants of the best possible world. If God does not create anyone, then God has not violated any obligations to anyone, even if the best possible world is not the one that is devoid of creatures. This does not mean that, morally speaking, God can choose to create any world whatsoever; God may have obligations not to create worlds containing persons whose lives are wretched on the whole; God may even have an obligation not to create any world containing any person whom God could have, by creating a better world, have given a better life. These obligations, however, are far from an obligation to create the best possible world. See Adams, “Must God Create the Best?” The Philosophical Review, 81 (1982): 317-332.

All this is to say that the principle, “God must create the best,” is not nearly as non-controversial as it initially appears. Indeed, apart from the obvious technical problems (with the coherence of the concept, “best possible world”) that we are sidestepping for now, the principle is quite likely false. This realization should give us pause in affirming the related principle that we are obligated to do whatever is morally best, assuming there is such a thing. Before endorsing the principle that we are always obligated to do the best, we should ask, among other questions, to whom we have this standing obligation to do what is best, and why.
The defender of MRA, in order to maintain that supererogation is impossible, must show, for each of the above three types of situations, either that (i) the situation type is not (in fact) possible or (ii) the situation type is not one that allows for supererogation. The defender of MRA is not without resources for making such an argument. Looking at situations of the first type, one could say that in the case of ties, the agent has a disjunctive duty to perform one of the multiple actions that is, given the agent’s reasons, morally best, rendering no acceptable option superior to another acceptable option. This sort of strategy may be adaptable to address cases of the second sort as well. One could thus (a) invoke a disjunctive duty to perform one of the infinitely-many actions surpassing some high level of moral merit and (b) simply deny that satisfying this duty by performing any of the ‘higher’ disjuncts rather than one of the ‘lower’ but acceptable ones counts as supererogatory on the grounds that the difference between the disjuncts is too small to be significant.96 Defenders of MRA could handle cases of the third sort in one of two ways. The first option would be simply to deny that such incommensurability ever occurs, saying, perhaps, that while we often have difficulty comparing our moral reasons for various options, nonetheless, these reasons always, in principle, admit of such comparison so that, for any two potential actions we might consider, either the preponderance of moral reason favors one of them simpliciter or else the moral reasons in favor of each are simpliciter equally strong. The second option would be to say that, in the case of incommensurability, one has a disjunctive duty to perform one of the incommensurable options. Disjunctively obligatory options are limited to those the moral

96 Alternatively, a defender of MRA could simply deny that cases of the second sort are ever possible, that is, she could deny the possibility of a person being in a situation in which the list of the actions open to her includes infinitely many actions of ever-increasing merit.
reasons in favor of which are not outweighed by any other option. The MRA defender could then say that, with the obligatory options so limited, no acceptable action could count as having more moral reasons than—and hence as morally superior to—any other acceptable action.

Some of the above responses to the three problem cases are themselves far from convincing, which provides something of a problem for MRA. What is more significant, though, is that MRA suffers from a final, devastating problem quite apart from the worries raised thus far. At the heart of MRA is the assumption that moral reasons play a two-fold role with respect to action assessment: they fully determine (by virtue of preponderance) which action or disjunctive set of actions an agent is morally obligated to perform in a given situation and they also fully determine the moral worth of an agent’s performed action. But, in fact, the features of a situation that determine which action or disjunctive set of actions, if any, is obligatory are not always the same features that determine the moral worth of a given action in that situation.97 Suppose that there is a particularly disagreeable, but important, job to do, and I am one of two people capable of doing it. The other capable person and I are both willing to do the job, but neither of us wants to. We agree to flip a coin to decide who will do it. The coin lands in my favor, and the other person assures me that she will take care of the job. If, at this point, I do the job anyway, not because the other person neglects her duty but simply because I want to give her a break,98 then I have performed an action deserving of higher moral praise than

---


98 I am not assuming that the other person is more deserving of a break than I am or that she would enjoy such a break more than I would.
if I had let the other person do what she was willing to do. Yet, at least some of what makes my action so praiseworthy does not determine what is obligatory in the situation in question. If it did, then presumably I would be obliged to volunteer to take over the job despite winning the coin toss, and surely I am not so-obliged given that the procedure we followed for dividing up the labor was a fair one and given that the job would have been done even if I had not done it. Contra MRA, then, an action can be non-obligatory and morally superior to an alternative permissible action. Even if a preponderance of moral reasons makes an action morally required, a preponderance of moral reasons is not the sole factor determining how morally good a particular action is.

MRA thus fails to establish that supererogation is impossible. Until MRA is repaired or another argument successfully shows that supererogation is impossible, given the central cases and the associated common assumption that people can, in some situations, act in a morally extraordinary, but not morally obligatory, way, we should assume that supererogation is genuinely possible. Until evidence is given for doubting our reflections on the central cases, then, we should assume that the appearances of supererogation in the central cases are for the most part reliable.99

1.5 The Import of the Central Cases for Moral Theory

The central cases constitute an important part of the ethical world. When an actual person appears to do what the soldier in Grenade does, the story is widely published and society honors the apparent hero. The widely-popular television program,

99 The qualifier, “for the most part,” grants that some apparent cases of supererogation may turn out to be mere appearances.
24, is centered almost entirely on a fictional hero, Jack Bauer, whose interactions with terrorists recall what occurs in *The Torture Case*. Whole books are written to gain insight into the lives of people like Paul Kocheleff. The success of the film, *Babette’s Feast*, cannot be explained without discussion of the aspects of its plot that I have described. Similarly, larger-than-life characters like Bishop Myriel constitute an important part of the appeal of *Les Miserables*. What explains the widespread interest in the sorts of things going on in these central cases is that these cases represent an important part of human experience or, at the least, an important part of human aspiration. As such, what goes on in these cases is an important part of the ethical world, broadly construed. A moral philosophy that neglects such cases is, as Urmson saw, incomplete.

Given, then, the importance of the central cases, the fact that they seem to involve supererogation, and the failure of MRA (the only philosophical argument for the in-principle impossibility of supererogation of which I am aware), particular moral theories must either make room for supererogation or else explain how to account for the importance of the central cases without appeal to the category of the supererogatory. Moreover, because the ways in which particular moral theories account for supererogation (or the central cases) are likely to vary widely, comparing the quality and plausibility of each moral theory’s account of supererogation (or the central cases) can play a positive role in helping us to adjudicate debates about the strengths of the particular theories. If a moral theory can provide a better account of supererogation than its rivals can, it scores a point against its rivals. Of course, the help that comparing accounts of supererogation provides in adjudicating among moral theories is not without restriction. Moral theories vary widely with respect to their theoretical foundations, and
we should expect that these divergent theoretical commitments will result in different analyses of purported instances of supererogation. The typical utilitarian, for instance, will hardly consider Babette a case of supererogation; rather, to the utilitarian, Babette’s actions will be positively wasteful. We should not privilege our pre-theoretical intuitions about particular cases to such an extent that the utilitarian’s theoretical reasons for rejecting Babette as an example of supererogation (e.g., that serving so lavish a feast when the money for it is needed elsewhere is wasteful) count for nothing. So, the fact that utilitarians, for reasons foundational to their moral theory, consider Babette’s actions immoral should not count against them, or, at the least, the extent to which it counts against them should be minimal. Nonetheless, once appropriate allowances for the divergent theoretical foundations of the various moral theories are made, we can assess the plausibility of those theories in part by looking at how they handle the central cases and their apparent supererogatory status.

The questions, “are the major ethical theories of Kantianism, utilitarianism, intuitionism, and virtue ethics capable of accommodating supererogation?” and, “if so, which theory or theories accommodate it best?” are thus worth asking. Much work has been done on a beginning of an answer to these questions, but what we have so far is just a beginning. I say, “beginning,” because, while many have tried to show how utilitarianism and/or Kantianism either allows or excludes supererogation and what that says about the plausibility of those theories, much less has been said about how virtue ethics stands with respect to supererogation.
Indeed, to the best of my knowledge, only one philosopher has provided a proposal that is accurately construed as a virtue ethics account of supererogation, and that philosopher’s account is not very informative: Jason Kawall’s virtue ethics account of supererogation simply tells us that an action is supererogatory iff a fully-informed virtuous agent would consider it supererogatory. The consensus of the rest of the literature on the issue of whether virtue ethics and supererogation are compatible is that they are not. Thus, Julia Annas has said (rather quickly) that, because everyone should strive for full virtue, virtue ethics leaves “no obvious room for supererogation.” Similarly, both David Heyd and Nancy Sherman, addressing the question, “is supererogation compatible with an Aristotelian ethic?” give negative answers. For Sherman, the problem for the Aristotelian is that right action is defined in terms of virtue, and as long as someone does not have a virtue to the highest degree, one cannot be said to

______________________________

100 To say that only one philosopher has attempted to provide an account of supererogation for the virtue ethicist is not to say that virtue and vice are not frequently mentioned in discussions of supererogation. In fact, one view of supererogation, one advanced by Gregory Trianosky, is that what distinguishes the person who performs supererogatory actions from the one who merely performs her duty and does what is permissible is the score of virtue and vice: a person who merely does her duty and does what is permissible may, despite the fact that she does nothing morally wrong, be vicious and lack complete virtue. “Supererogation, Wrongdoing and Vice: On the Autonomy of the Ethics of Virtue,” Journal of Philosophy, 83 (1986): 26-32. A similar (and popular) view is that the supererogatory is praiseworthy, not because it fulfills an obligation (it does not), but because it arises from a virtuous character. David Heyd, “Supererogation,” in Stanford Encyclopedia of Philosophy. Despite their mention of virtue (and vice), these explanations of supererogation are not the sort that is likely to appeal to a virtue ethicist. The explanations, as they stand, do not embody the definitive features of virtue ethics, and they appear to presuppose accounts of moral obligation that it is unclear whether a virtue ethicist can or will want to adopt.


have done more than what ought be done. Heyd’s conclusion about Aristotelian ethics is based on similar reasoning: for the Aristotelian, the virtues are means between vices; hence, excess is a fault and the virtues cannot be improved upon. Heyd sees an additional problem if virtue ethics is what he calls “non-deontological”:

Now a completely non-deontological theory—that is to say, a theory based on a concept other than duty (virtue, sympathy, etc.)—cannot accommodate supererogation because if there is no duty then a fortiori there cannot be action which transcends duty. 105

When the issue of supererogation and virtue ethics is raised, then, the consensus is that virtue ethicists will have nothing to say about supererogation.

One important departure from this consensus occurs in the work of Andrew Flescher. Flescher’s home discipline is not philosophy but religious studies, but his Heroes, Saints, and Ordinary Morality has much to teach moral philosophers. While the goal of the book is not to provide an account of supererogation for virtue ethics, Flescher does consider himself to be a virtue ethicist who is wrestling with the question of what moral saints and heroes teach us about morality. Given the saints and heroes that Flescher considers, we can say that Flescher is not ignoring some of the sorts of cases that I call central cases. In that respect, Flescher has made important headway on the issue of how virtue ethics might accommodate the central cases. Flescher’s contribution to this issue is, however, limited. For although Flescher considers himself a virtue ethicist, and although he has much to say about what saints and heroes are, he does little to connect what he says about saints and heroes to virtue ethics as such. Indeed, the deepest


104 Heyd, Supererogation, 36-7.

105 Heyd, Supererogation, 3-4.
connection he draws is that virtue ethics teaches us that we are obligated to improve our character, and this helps us to see why are obligated to make ourselves more like heroes and saints. Swanton and Slote, virtue ethicists who have expressed some amenability to supererogation, do even less than Fletcher to show how virtue ethics can make room for supererogation.\textsuperscript{106}

While the consensus of the literature is that virtue ethics cannot accommodate supererogation, too little has been done on the topic to warrant this conclusion. To the best of my knowledge, the only substantive challenges to a virtue ethics account of supererogation that have thus far been raised are the ones that Heyd and Sherman raise for the Aristotelian and that Annas raises for virtue ethics more generally. But Heyd and Sherman’s criticisms should not be assumed to be fatal for the virtue ethicist who wants to allow for supererogation for Aristotelian ethics and ethics of virtue are not equivalent. What is more, and what can be said in the wake of Annas’ criticisms as well as Sheman’s and Heyd’s, Kantians and utilitarians also face initial problems accommodating supererogation, but that has not stopped adherents of both ethical theories from attempting to construct accounts of supererogation. Not every theory can be expected to accommodate supererogation as readily and naturally as a Divine Command Theory that has recourse to a command/ counsel distinction,\textsuperscript{107} but that does not mean that we should assume that the initial difficulties that other moral theories face are insurmountable.

\textsuperscript{106} See Flescher, \textit{Heroes, Saints, and Ordinary Morality}, chapter 4.

\textsuperscript{107} Thanks to Jim Sterba for pointing out that supererogation is likely most at home within divine command theory.
Careful exploration of accommodation strategies is what is needed to determine what, if anything, a moral theory facing initial challenges can say about supererogation.

1.6 Thesis and Outline of the Project

My own diagnosis of the lack of attention that has been paid to the intersection of supererogation and contemporary virtue ethics is that it is due less to the philosophical hurdles facing a virtue ethics account of supererogation than it is to historical contingency. Anscombe’s 1958 “Modern Moral Philosophy” put down roots for contemporary virtue ethics, but the school of thought did not come to be seen as a significant alternative to Kantian and utilitarian ways of thinking until the publication of After Virtue in 1981. But while moral philosophers—Peter Geach and Phillipa Foot excepted—were largely ignoring the virtue alternative, much work was being done, in response to 1958 Urmson, on the topics of supererogation, moral saints, and moral heroes.108 Given that the bulk of contemporary work on supererogation and its place within particular normative theories was done prior to the time when virtue ethics was seen as a genuine alternative to Kantian and utilitarian views, the neglect of the intersection of virtue ethics and supererogation is unsurprising, even if that neglect is a poor indicator of virtue ethics’ ability to accommodate supererogation.

108 Thus, a look at the dates of the books and articles included in the bibliography for the entry, “Supererogation,” in the Stanford Encyclopedia of Philosophy reveals the following trend: the number of articles and books mentioned, looked at by decade, grows steadily from the 1950’s to the 70’s, when it peaks; the numbers decline slightly starting in the 80’s, and the mild decline continues up until the first decade of the twenty-first century. By contrast, the bibliography of the Stanford entry on “Virtue Ethics,” Rosalind Hursthouse, first published 2003, revised 2007, suggests a very different trend in work on that topic: the bibliography includes only two entries from the 50’s, none from the 60’s, and only a handful from the 70’s (which, recall, was the decade of supererogation’s peak); significant growth occurs in the 80’s, however, and extends to the 90’s; the number of entries for the first decade of the twentieth century is only slightly lower than the 90’s peak.
Not enough, then, has been said with regard to how virtue ethicists can accommodate supererogation or the central cases. The primary goal of this dissertation is to address that neglect and to develop an account of supererogation—or, at least, something like it that explains the draw of the central cases—for virtue ethicists or a significant subset of them. A secondary aim of my dissertation is related: just as a moral theory’s ability (or inability) to account for supererogation says something about the viability of that theory, so too does the way in which it accounts for it (and how its account fits with our beliefs about supererogation) have implications for the plausibility of the moral theory. My secondary aim, then, is to show not only that virtue ethics can offer an account of supererogation, but also that this account is, as an account of what is of significance in the central cases, superior to the accounts of supererogation that have been offered on behalf of both Kantianism and utilitarianism. Assuming that our intuitions about the central cases are generally reliable, this feature of virtue ethics is a mark in its favor on a comparison of its plausibility to that of Kantianism and utilitarianism.

My aim, then, is to show that virtue ethicists—or at least a significant portion of them—have access to a viable account of supererogation, one that accounts better for supererogation than the accounts of Kantians and utilitarians do. Along the way, I hope to shed additional light on the nature of supererogation itself, revealing features of supererogation and of moral saints and heroes that, while acceptable to many who are not in the virtue ethics camp, may nonetheless be difficult to see (at first) if one is not looking at those topics through the eyes of the virtue ethicist.
Now for a sketch of the remainder of the dissertation. In the next two chapters I explain how utilitarians (Chapter Two) and Kantians (Chapter Three) have attempted to accommodate supererogation, offering criticisms of these accounts. In Chapter Four, I provide an overview of the field of virtue ethics and develop my own virtue ethics account of supererogation that is available to a significant subset, although not all, of the members of this field. Chapter Five compares the adequacy of the accounts that we will have seen Kantians, utilitarians, intuitionists, and virtue ethicists offer and argues that, given the different accounts, virtue ethics is in a better position with regard to the central cases than its rivals are.
CHAPTER TWO

BEYOND MAXIMIZING: UTILITARIANISM AND SUPEREROGATION

At first blush, utilitarian moral theories straightforwardly exclude the possibility of supererogation. The problem is that utilitarianism traditionally has a maximizing requirement: individuals must perform the morally best action available to them, which is whatever action will lead to the best consequences, whether the value of potential consequences is determined by preferences satisfied, happiness achieved, or some other criterion. The only occasions on which the agent is not constrained to perform one particular action will be those occasions on which at least two available actions will lead to consequences of equally great worth. But in such cases, the equal worth of the consequences implies the equal worth of the prospective actions. So, in no situation, according to utilitarian theories, is an agent able to perform an action $a_1$ that is morally superior to some other action $a_2$ available to her in that situation, where $a_2$ would not violate (or not entail violating) moral duty in those circumstances. Hence, it appears that utilitarianism does not admit of actions that surpass duty and so cannot allow for supererogation.

The apparent conflict between supererogation and utilitarianism has not stopped utilitarians (or people arguing on its behalf) from offering accounts of supererogation that they see as compatible with some version or versions of utilitarianism. In order to get a
feel for how successful, if at all, utilitarian moral theories have been at capturing supererogation, this chapter will consider three strategies for accommodating supererogation within a utilitarian framework. All of the three are explicitly explained and defended in the literature on supererogation, and they all entail rejecting the maximizing requirement: J.O. Urmson lowers the bar for utilitarian moral obligation to what is required for the existence of civil society; Robin Attfield adheres to the maximizing requirement when great goods and evils are at stake but otherwise eschews duty entirely; and Douglas Portmore permits agents to pay more attention to their own welfare than to that of others. As my analysis will make clear, Urmson’s and Attfield’s accounts, in addition to having theoretical difficulties, are both at once too liberal and too conservative in what they admit into the category of the supererogatory. Portmore’s account provides a more plausible sorting principle for determining whether or not an action is supererogatory, but it departs too far from the heart of utilitarianism to merit classification as a utilitarian moral theory. Reflection on the difficulties encountered by all three strategies shows, I shall argue, that we should not be optimistic about the prospects for developing a plausible account of supererogation within the confines of a utilitarian moral theory.

2.1 Urmson and Basic Duty

In “Saints and Heroes,” J.O. Urmson admits what we have already discussed—that utilitarianism initially seems to crowd out the possibility of morally saintly or heroic
action. Urmson is yet optimistic that utilitarianism can be modified to accommodate it, and Urmson’s strategy for doing so involves rejecting the maximizing requirement. On his plan, people are not obligated to perform actions that lead to the best consequences, the *summum bonum*, but only to act in accordance with rules needed to ensure avoidance of the *summum malum*, the sort of life that Hobbes described as “solitary, poor, nasty, brutish, and short.” All of our moral duties thus turn out to be quite basic—telling the truth, keeping promises, and refraining from murder are examples that Urmson mentions. Urmson enumerates five reasons for relaxing duty in this way, each reason premised on the assumption that “we are aiming at a moral code that will best serve human needs, a code that is ideal in the sense that a world in which such a code is acknowledged will be a better place than a world in which some other sort of moral code is acknowledged.” The five reasons: the relative urgency (for human life) of ensuring that basic duties are fulfilled versus ensuring that more saintly acts are performed; the limited moral capacities of ordinary people; the ease with which requirements of basic duties—but not requirements of saintly and heroic action—can be codified; the limited extent of our right to demand that others behave in particular ways; and the general preferability of encouraging good behavior by means of commendation rather than by


110 Urmson, 327.

111 Urmson, 328, 330.

112 Urmson, 329.
threat of punishment. In sum, to Urmson, the realities of human need and limitation ground a utilitarianism that limits moral duty to basic duty.

While utilitarianism’s standards for obligation change on Urmson’s plan, the standards for evaluating consequences and by extension how acts are valued do not change. How Urmson sees supererogation as possible within such a framework of obligation and value is unsurprising: supererogation occurs when an agent performs an action that brings about a state of affairs that is superior to some other state of affairs that the agent could have brought about without violating the requirements of basic duty.

Problems with Urmson’s five-fold justification for his utilitarian understanding of saintly and heroic action abound, a point that others have been quick to make. Robin Attfield, for instance, has criticized some of Urmson’s “utilitarian” reasons for drawing a distinction between basic duties and what surpasses them. The first of these reasons, recall, is that utilitarianism recognizes the fulfillment of basic duties as having an import that the performance of other good actions does not have. Neglecting to volunteer to watch one’s friends’ colicky child for an evening is one thing; neglecting to refrain from killing the innocent is another. Attfield agrees that fulfilling basic duty can be more important than performing saintly acts, but he points out that this feature of basic duties does not show that the so-called saintly acts are non-obligatory. That some obligations are more important than others is to be expected. Urmson himself would presumably admit that even among what he regards to be basic duties, some (e.g., refraining from

\[113\] Urmson 329-30.

killing the innocent) are far more crucial than others, but this is not a reason to think that the less crucial ones are not actual duties.

Urmson’s second reason for seeing moral duty as limited to basic duty is that human moral capacities are limited, and moral requirements should be compatible with these limits. Urmson may be correct that psychological limitations sometimes render impossible for a person actions that are nonetheless physically possible for her, precluding such actions from being obligatory, but we can adhere to the principle that ought-implies-can—so that no one is required to do what she is incapable, whether for reasons of limited moral development or otherwise, of doing—without limiting moral duty to Urmson’s basic duty. Most people are capable of doing much more than simply keeping their noses clean in the way one must merely to satisfy such basic duties. At least for most people, then, the ought-implies-can principle does not sufficiently motivate limiting moral duty to requirements as sparse as Urmson’s basic duties. In any case, if adherence to the principle that ought-implies-can entails that no one is obligated to do what is beyond her moral capacity, then in any situation in which the ought-implies-can principle limits what is required of the agent, supererogation is impossible for the agent. In such a situation, the agent is incapable of doing what would, for her, be beyond duty. Hence, even if it were true that the principle that ought-implies-can implies that moral duty collapses to basic duty, this collapse would not make room for supererogation.  

115

__________________

115 Urmson may be thinking that moral duty should be what is possible for the least common denominator. Then, humans with very limited moral capacities would be capable only of fulfilling basic duty, but the rest of us would be capable of surpassing it. The problems with this suggestion are twofold. First, the suggestion that what is required of us is limited to what the worst of us are capable of doing is insufficiently motivated. Second, so limiting duty would result in duty’s being far more basic than Urmson’s theory requires. No doubt some compulsive liars are incapable of telling the truth on a regular basis, so regular truth-telling would no longer be a basic duty.
Urmson might respond that the above objection fails to grasp the full significance of his claim that the limits of human moral capacities limit moral requirements. While human limitations render some actions impossible, other possible actions they merely make very difficult. Limiting moral duty to basic duty, Urmson might say, acknowledges this distinction insofar as basic duties are much less difficult than the ones that surpass them. To this line of reasoning we can reply that, in the first place, since basic duty is defined, on Urmson’s plan, not by ease of accomplishment but by import of consequences, there is no guarantee that actions that fulfill basic duties will always be easier than those that surpass them. Indeed, truth-telling and promise-keeping, for instance, are both frequently more difficult than granting small favors. In the second place, we can make a reply similar to the one that Attfield made to Urmson’s point about the import of basic duties as opposed to that of saintly acts: just as we should expect some obligations to be more important than others, we should also expect some obligations to be more difficult to fulfill than others. So, pointing out that some good actions are much more difficult than others does not show that the more difficult ones are optional.

Urmson’s other grounds for relaxing duty in a way that allows for action that morally surpasses it all tend, as M. Clark says, to “confuse the propaganda aspects of morality with morality itself,” concluding that a particular action A is not morally required because saying or insisting that A is morally required would lead to bad consequences.\footnote{M. Clark, “The Meritorious and the Mandatory,” \textit{Proceedings of the Aristotelian Society}, 79 (1978): 27-8.} The third way that Urmson supports his view that only basic duties are
required is to make the point that basic duties are much more easily codified and hence required than are saintly actions. Even if it were true that rules that require the so-called basic duties are more easily codified than are rules that require more saintly action—and it is not obvious that this claim is true—a duty is applicable even if formulating a relatively simple rule that expresses it is difficult. Ease of codifiability is not a necessary condition for duty.\textsuperscript{117} Urmson’s fourth and fifth grounds for the basic duty distinction turn on problems associated with extracting saintly actions as duties (and the preferability of encouraging them via praise instead). As Attfield recognizes, an action can fulfill a duty even if no one can rightly demand fulfillment of that duty.\textsuperscript{118} So, even if praise is a more effective method of encouraging so-called saintly and heroic action than insistence is (and even if, by contrast, insistence is a preferable method in the case of basic duties), it does not follow that the moral worth and obligatory status of actions ever come apart on utilitarianism.

The above problems for Urmson’s five-fold justification are significant, but more problematic for Urmson is the fact that, even if we ignore the problems with his justification of his account, the account itself is implausible as an account of supererogatory action. For, given Urmson’s criterion for basic duty (that criterion, recall, is avoidance of the \textit{summum malum}), basic duty turns out to be even more basic—and supererogation hence much more common—than Urmson may realize. Wealthy, twenty-

\textsuperscript{117} Attfield, “Supererogation and Double Standards,” 483-4. In this context, it is worth pointing out that if ease of codifiability is a necessary condition for duty, we have even more reason to believe that, contra the typical utilitarian stance, moral injunctions against lying, theft, and promise-breaking are absolute.

\textsuperscript{118} Attfield, 485.
first century societies—with their universities, indoor plumbing, vaccines, and multi-room houses—are far from the *summum malum*. Yet these societies are also ones in which people routinely lie and break promises, at least when the promises in question are not backed up by threat of law. Indeed, the typical American admits to lying once a day, on average, and for college students, the rate doubles. In fact, people could probably lie a great deal more than they currently do and still preserve many of the features of modern life that enable them to enjoy lives that involve a great deal of “commodious living” and that, historically speaking, are relatively long, healthy, and comfortable. So, members of contemporary wealthy societies may be obligated to keep most of their promises and to tell the truth most of the time, but a great deal of lying and promise-breaking turn out to be licit. More broadly, if fulfilling duty is merely a matter of doing what is necessary to avoid the *summum malum*, insofar as contemporary wealthy societies are far from the *summum malum*, members of these societies turn out to owe very little to one another.

Given that Urmson’s criterion for basic duty embodies a form of rule utilitarianism, Urmson does have a reply here. He could insist that regular obedience to the rule, “Do not lie!,” is required for avoidance of the *summum malum*: were people to

---


120 Note that I am not here presupposing an overly rigorous understanding of Urmson’s *summum malum*. Urmson himself takes his *summum malum* from Hobbes and his state of nature in which life is “nasty, brutish, and short,” but one might argue that this characterization is an exaggeration of what Urmson really means, that Urmson’s *summum malum* should be understood less drastically as merely life without civil society. What I say here applies regardless of whether we adopt the more or less dramatic interpretation, for my point is that a great deal of lying is compatible not only with a world that surpasses the Hobbesian state of nature but also with a world containing industry, the arts, and developed social institutions.
lie not once or twice a day but, say, in thirty or fifty percent of attempted communications, communication would break down, and so would society. So, his criterion renders refraining from lying a basic duty. Pointing out that regular adherence to less-rigorous rules, e.g., “Lie, at most, thrice per day!,” or “When speaking, tell the truth (to your knowledge) at least ninety-nine percent of the time!,” is likewise required for avoiding the *summum malum* does not show that regular adherence to the more stringent, “Do not lie!,” is not required for avoidance of the *summum malum*. *Regular* adherence to *all* of these rules is required, even if *perfect* adherence to all of them is not. So, Urmson could reason, all of these rules embody basic duties. Of course, if one is to fulfill all of them, one must avoid lying altogether and not simply limit one’s lying to the greatest frequency compatible with the existence of civil society.

The problem with the above reply is that it proves too much. If we take Urmson’s criterion as validating *any* rule—and not merely the least stringent rules—that must not frequently be violated if the *summum malum* is to be avoided, then Urmson’s criterion does indeed support a basic duty of refraining from lying and not merely refraining from lying most of the time. But notice that this simple rule, as we have stated it and as it is supported, does not admit of exceptions even in extreme cases when lying will lead to catastrophic results.121 Since Urmson, as a proponent of a form of rule utilitarianism, is committed to these rules being indicative of actual basic duty and not merely of being

---

121 It will not help to object that an absolute rule against lying is unjustified because the results would be catastrophic if such a rule were always obeyed. In the first place, even if individual catastrophes would ensue on occasion, civil society would not be undone (and the *summum malum* realized) if everyone were to exercise perfect honesty at all times. In the second place, the criterion for legitimate rules that we are considering is (for good reason) not framed in terms of what would ensue given perfect obedience to the proposed rule; it is framed in terms of what rules must be generally obeyed if the *summum malum* is to be avoided.
guidelines that generally get basic duty right,\textsuperscript{122} it follows that, on the interpretation of Urmson’s criterion that we are now exploring, lying is universally morally impermissible, even in extreme cases. This conclusion goes against the heart of utilitarianism, even when in the form of indirect utilitarianism, and seemingly transforms it into a rigid form of deontology.\textsuperscript{123} To avoid this result, we should interpret Urmson’s criterion not as validating any rule whose regular violation would lead to the \textit{summum malum} but as validating only the least stringent rules (or set of rules) whose regular violation would lead to the \textit{summum malum}. This, we might say, should be the real force of saying that regular obedience to these rules is necessary for avoidance of the \textit{summum malum}. But, for reasons we have already seen, it follows that basic duties are quite minimal and that lying an average of once, twice, or thrice per day—and not only to achieve great goods—is, for instance, likely permissible, at least for most people.\textsuperscript{124}

The consequence of duty relaxed to the extent that it is if mere avoidance of the \textit{summum malum} grounds duty—even if it does so by taking a detour through rule utilitarianism—is that fairly trivial actions turn out to be supererogatory: in any of the

\textsuperscript{122} If the rules are mere guidelines, we have an account not of how one can surpass duty but of how one can surpass generally (but not always) correct moral rules of thumb. In such a case, Urmson still owes us a story about what determines actual duty. Regrouping to invoke a form of act utilitarianism that speaks of avoiding the \textit{summum malum} will not help for it will only lead to the problem that we have already seen of setting the bar for duty too low. See David Lyons, \textit{Forms and Limits of Utilitarianism} (London: Oxford University Press, 1965), 119-120.

\textsuperscript{123} Even forms of indirect utilitarianism will allow lying in extreme cases, as when an innocent human life is at stake and the lie is told to a Nazi.

\textsuperscript{124} Note that the same problematic conclusion follows if one interprets Urmson’s criterion as validating only the least stringent rules (or set of rules) whose \textit{general acceptance} is necessary to avoid the \textit{summum malum}. For the \textit{summum malum} could be avoided if people were to accept a code of rules that permitted occasional (an average of twice per day) lying for the sorts of vaguely selfish reasons that most lies today are told. Indeed, it would not be at all surprising to learn that many of the members of our own society do believe that, while rampant lying is ill-advised, a little bit of lying for personal gain is perfectly acceptable.
many situations in which one, given one’s options, has no basic duties, merely performing an action other than the worst available one (ranking the actions in terms of the utility they produce) will turn out to be supererogatory. An action as trivial as eating a sandwich when one could have had a candy bar (or vice versa, depending on one’s palate and state of health) might thus be classified as supererogatory.

Not only is Urmson’s strategy, given its criterion for basic duty, too liberal in what it admits as supererogatory, it is also too conservative, failing to classify as supererogatory some of the cases that, it seems, a utilitarian who wants to acknowledge the possibility of supererogation would most want to capture. Recall Grenade and Kochelleff, both of which were modeled on cases that Urmson himself introduced as involving agents whose actions go “far beyond” what duty requires. Suppose that we stipulate in Grenade that one of the soldier’s colleagues (one lacking an escape) would have jumped on the grenade had our hero not done so. So, had our hero fled to save his own life, the same number of people would have been saved, and the same number of people would have died. If we stipulate that the utility that is lost is the same regardless of which soldier dies (our hero or his sans-escape colleague), our hero’s falling on the grenade turns out to be no better, in terms of utility, than his running from it. So, the only actions open to the agent that yield less utility than his falling on the grenade turn out to be so bad as to be forbidden—they are actions such as not falling on it and ensuring that
no one else does either, the sorts of actions that must be avoided if the *summum malum* is to be avoided.\textsuperscript{125}

Turning to the Kocheleff case, given the severity the AIDS crises that Kocheleff confronted in Burundi and South Africa, doing something to avoid or alleviate such situations ought to count as necessary to avoid the *summum malum*, if anything does. The key question for Urmson’s view is how extensive this “something” is. It is tempting to think that Urmson could say that whatever basic duty requires in this situation, it does not require anything so grand as making alleviating the AIDS crisis one’s life’s work. After all, if everyone of means in the 1970’s had done something to remedy the crisis, the benefit would have been great and the *summum malum* of widespread infection contained. So, the rule regular obedience to which is necessary to avoid the *summum malum* requires (and required) only minimal action. Kocheleff’s work has surpassed this minimum, so Kocheleff life’s work can be said to be supererogatory.

The problem with the above reasoning is that it ignores a salient feature of the long-term AIDS crisis that Kocheleff confronted: its decades-long persistence. The continuance of the crisis served as a breeding ground for new duties for both Kocheleff and others (especially other physicians). If Kocheleff had worked in Burundi for only a year and then returned home to Belgium, many more Africans would have suffered and died early deaths, and the continued suffering would have generated new instances of the obligations that Kocheleff acquired when he first confronted the epidemic. So long as crises persist, they persist in generating obligations. In this case, the persistence of the

\textsuperscript{125} If such an action turns out not to be forbidden on Urmson’s theory, Urmson’s theory loses all plausibility, for then running away from the grenade (without interfering with anyone else’s jumping on it) would count as supererogatory.
AIDS crisis prevented Kocheleff’s work, on Urmson’s criterion, from ever counting as surpassing duty. Of course, if most other capable people had, like Kocheleff, done something in the early years of the crisis, the situation might have been much better. If so, a year of work would likely have been sufficient for Kocheleff to discharge his duty without any new duties emerging. But people did not, in general, behave in this way. The crisis persisted, and duty reemerged. So, on Urmson’s criterion, while Kocheleff was not obligated, upon first confronting the crisis, to make that crisis his life’s work, his life’s work in Burundi and South Africa has nonetheless, given others’ inaction, not been supererogatory. At best, Urmson can attribute to Kocheleff the distinction of having done more than many of his peers have done: while others have continually shirked their duties with respect to the AIDS crisis, Kocheleff has repeatedly fulfilled his, including duties that have emerged in part as a result of others’ negligence.

A final problem with Urmson’s view emerges from the way in which it attempts to combine direct and indirect utilitarianism, and this problem would arise even if Urmson, in his account of moral obligation, endorsed a more stringent, even maximizing form of rule utilitarianism. The problem becomes clear when we realize, given Urmson’s account of saintly and heroic action, that such action can be morally wrong. Consider a case in which an agent has (at least) two possible courses of action open to her. The first of these actions violates no valid moral rule;¹²⁶ the second one does violate a valid moral rule. Despite the fact that the second action violates a valid rule, performing it on this particular occasion will lead to better consequences (in terms of well-being) than

¹²⁶ I am using “valid moral rule” to mean “an actual moral rule,” on the Urmson picture picture, a valid moral rule is one regular adherence to which is necessary for the avoidance of the sumnum malum.
performing the first will.\textsuperscript{127} Given Urmson’s theory, it follows that the second action is wrong (because it violates basic duty) but nonetheless morally supererogatory (because it leads to more well-being than a permissible alternative). This result is unacceptable: one should not be able to surpass moral duty by violating it. Of course, Urmson could do what he has not explicitly done and add a qualifier that no supererogatory action is morally wrong, but this response fails to address the deeper problem with his theory. If basic duties are not simply moral rules of thumb but are rather actual duties that one must not violate, then there should be no fear that any correct way of evaluating actions will evaluate a potential action that violates basic duty as morally superior to an alternative action that does not. This fear is realized in Urmson’s theory—even if we stipulate that an act of supererogation never violates basic duty. Further stipulating that violating basic duty disqualifies an act from being ranked higher (morally) than a licit alternative act would avoid this particular problem but would be \textit{ad hoc}. What Urmson owes us is a way of negotiating the conflicts between direct and indirect forms of utilitarianism that arise when they are combined as they are on his theory. Absent such an explanation, the Urmson account of supererogation is implausible.

The above problem will apply to any utilitarian account of supererogation that trades on blending direct and indirect forms of utilitarianism. Thus, if one tries to make room for utilitarian supererogation by using an enforceability standard, rather than Urmson’s \textit{sumnum malum} standard, to determine which rules are legitimate and hence

\textsuperscript{127} If it is objected that it is not possible for an act that violates a valid moral rule to produce more utility than one that violates no such rule, then we should ask why Urmson frames his account of basic duty in terms of rules, since it would then appear to be extensionally equivalent to an act utilitarianism that takes basic duty to be a matter of whether performing a particular act is necessary to avoidance of the \textit{sumnum malum}. 

59
imply obligation, one will encounter a similar problem. This point is significant because, at least *arguably*, the account of moral wrongness that J.S. Mill endorses in the fifth chapter of *Utilitarianism* involves this standard of enforceability and not the traditional maximizing standard. Thus Mill:

> We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow-creatures; if not by opinion, by the reproaches of his own conscience. This seems the real turning point of the distinction between morality and simple expediency. It is a part of the notion of Duty in every one of its forms, that a person may rightfully be compelled to fulfil it. Duty is a thing which may be exacted from a person, as one exacts a debt. Unless we think that it may be exacted from him, we do not call it his duty. Reasons of prudence, or the interest of other people, may militate against actually exacting it; but the person himself, it is clearly understood, would not be entitled to complain.\(^\text{128}\)

Mill here appears to endorse a (non-maximizing) enforceability standard of obligation, and this standard appears to be embedded in a form of indirect utilitarianism. For Mill is telling us that whether a person has a duty depends on whether it is generally acceptable to extract that duty, not whether it is permissible to extract it in a particular case (note the last full sentence of the quotation). It is tempting to look to this non-maximizing Millian theory for a way of accommodating supererogation, saying that when an act that is non-obligatory (by the enforceability standard) promotes (sufficiently) more good than a licit alternative, the act is supererogatory. To take this approach, however, would leave one vulnerable to the same problems that the Urmson account encounters when it combines direct and indirect forms of utilitarianism: one will be able to surpass duty by doing what is wrong.

2.2 Attfield and Low Stakes Cases

Due to problems he sees in Urmson’s strategy for making room for supererogation within utilitarianism, Robin Attfield has developed his own strategy. Attfield’s approach is similar to Urmson’s in important ways. Like Urmson, Attfield recommends revising the way in which utilitarianism has traditionally determined moral obligation in order to yield a more relaxed standard. On Attfield’s plan, however, obligatory acts are limited not simply to those acts that are necessary for civil society but are rather (assuming act utilitarianism) those acts that are open to the agent and that are both optimific and either productive of important goods or capable of averting important harms.\textsuperscript{129} If rule utilitarianism is correct, obligation is determined by “rules general compliance with which would have these effects [i.e., maximizing benefit and either producing important goods or thwarting important harms].”\textsuperscript{130} Because Attfield, like Urmson, preserves utilitarianism’s traditional criteria for evaluating both the consequences and values of acts, Attfield’s strategy allows some acts to be non-obligatory and yet superior to some morally acceptable alternatives open to the agent. Notably, these supererogatory cases will not involve anything as extraordinary as saving a life; when the stakes are that high, the optimific action will nearly always be required. But one can go beyond duty when stakes are not as high—by weeding a friend’s garden when the friend is out of town, for instance.


\textsuperscript{130} Attfield, “Optimific, Right, But Not Obligatory,” 12.
While Attfield’s criterion for obligation is in some ways more rigorous than Urmson’s, Attfield’s theory still renders supererogation quite commonplace. Whenever a person is in a situation in which neither important harms nor important goods are at stake, any action at all will satisfy duty, which means that, in any such situation, merely doing more than the least one is capable of doing will be supererogatory. Because most of the situations in which we find ourselves in daily living are not so pressing as to involve important benefits or harms but do admit of a range of available actions that vary with respect to how much well-being each promotes, supererogation thus turns out to be quite cheap on Attfield’s account. Again, eating a sandwich rather than a candy bar could count as an instance of supererogation, even in non-extraordinary circumstances.

Attfield’s picture also suffers from a problem that, we have already seen, plagues Urmson’s: when we consider paradigmatic cases of supererogation—e.g., Grenade and Kocheleff—Attfield’s view cannot recognize these cases as involving supererogation. For in both of these cases and in paradigmatic cases of supererogation more generally, the stakes are so high (sometimes human life) that the actions performed must be considered obligatory.\textsuperscript{131} Now, to be fair, Attfield’s primary motive for developing his utilitarian account of supererogation is not to be able to accommodate paradigmatic cases of supererogation, or even supererogation in general. Rather, his goal is to be able to

\textsuperscript{131} One might object that great goods are not at stake in Grenade given that if the soldier does not fall on the grenade, one of his colleagues will, so that, no matter what the soldier does, the casualties will be the same. Alternatively, one might say that while great goods are at stake, falling on the grenade is still not required of the soldier, on Attfield's theory, because such an action is not optimific in that it will not produce more good than all the alternative actions: the alternative of running away and letting a second soldier take the fall produces as much good. Neither of these objections to my reading of the Attfield assessment of Grenade allows Attfield to recognize supererogation in this case. At best, one of these two objections shows that the soldier was not morally required to fall on the grenade; Attfield still has not given us a way of reading this non-required act as morally superior to (and not merely morally equivalent to) some permissible alternative.
classify what he calls “trivial favors”—actions such as mowing the neighbor’s lawn—as both permissible and beyond duty rather than as, at best, mere fulfillments of duty. Attfield’s solution is to limit duty to high stakes cases. The upshot is that only when the action is extraordinary but the consequences more trivial can an extraordinary case count as involving supererogation—a somewhat surprising stance for a consequence-centered moral theory to take.

While the intention of Attfield’s solution may not have been to accommodate the paradigmatic cases of supererogation, this intention is not an excuse for his solution’s difficulties in accommodating them if we are evaluating his account as an account of supererogation. To allow for supererogation in cases of gratuitous lawn-mowing is, to be sure, to allow for supererogation. But, as discussed in chapter one, it is the way in which extraordinary cases capture something important about human life and aspiration that prompts the pressing need for moral theories to make room for supererogation, to develop accounts of moral saints and heroes, or otherwise to provide an error theory for our inclination to bestow the titles “saint” and “hero” on extraordinary persons—and not merely on persons who hand out gratuitous sticks of gum, always hold the door open, and recycle more consistently than even the average resident of San Francisco. Not just any theory of supererogation will satisfy this need. So, if the utilitarian is to respond to some worries about supererogation by endorsing Attfield’s solution, she still owes us a story of why so many high-stakes cases—including the Kocheleff case and, especially, the grenade case—appear to involve supererogation. Absent such an explanation, Attfield’s account proves unsatisfactory.
An additional problem with Attfield’s solution is that it lacks sufficient utilitarian justification. Attfield is one who (in a different article) criticizes Urmson’s justification for relaxing duty (Urmson’s justification for relaxing duty, recall, is grounded in the importance of basic duties and the negative consequences associated with insisting on higher ones). Attfield’s general complaint is that Urmson’s justification merely points to ways in which basic duties differ from other duties; it does not show that these other duties turn out not to be duties at all. Whatever the success of Attfield’s critique of Urmson, Urmson at least can be said to appeal to the right sorts of grounds—utilitarian ones—to justify an account of saintly and heroic action that aspires to be utilitarian. By contrast, Attfield provides no utilitarian grounds for limiting the applicability of duty to high-stakes cases alone; he merely appeals to the common sense intuition that many of the actions that come out as obligatory on traditional forms of utilitarianism are in fact supererogatory. Until we have utilitarian reasons for regarding obligations as emerging in the way that Attfield envisions, Attfield’s account of supererogation either cannot be called a utilitarian account of supererogation or must be viewed as ad hoc, emerging not from the commitments of utilitarian ethical theory, or even from the intuitions that motivate that theory, but solely from that theory’s desire to dodge an objection to which it would otherwise be vulnerable.


133 For Urmson limits duty to his basic duties because so-limiting moral obligation is preferable, in terms of utility, to adhering to a maximizing requirement.

2.3 Portmore and Total Adjusted Wellbeing

Douglas W. Portmore has defended a brand of consequentialism that is modeled on a theory that Sheffler has developed. Portmore accordingly calls his theory “Shefflerian Utilitarianism” or SU. Portmore’s purpose in defending SU is not exclusively to accommodate supererogation in a utilitarian (or even a consequentialist) framework, but SU does accommodate it, purportedly within such a framework. Accommodating supererogation, says Portmore, is one among several strengths of SU, including an ability to allow for agent-centered options and to recognize self-other asymmetry. These virtues in combination are what Portmore identifies as motivating SU.\textsuperscript{135}

SU is a more complicated version of utilitarianism than the ones we have thus far considered. Accordingly, it will take some time to lay the groundwork necessary for understanding what SU is and how it (allegedly) accommodates supererogation. To begin, SU is noteworthy as a form of utilitarianism in that it rejects some standard assumptions of traditional versions of consequentialism. Typically, says Portmore, versions of act-consequentialism provide some way of ranking outcomes and couple this ranking method with a requirement that the agent perform an action that, of the available alternatives, leads to a highest-ranked outcome.\textsuperscript{136} How outcomes (and by extension actions) are ranked is assumed to be an objective matter such that rankings and values of identical outcomes do not vary from agent to agent. Thus, if we consider the state of


\textsuperscript{136} Portmore, 409.
affairs consisting in Sam’s enjoying an ice cream cone (to a particular extent and in a particular way), this single state of affairs does not have one value when Betty brings it about and another when Sam brings it about. Of course, Betty’s *buying* Sam an ice cream cone may have a different (total) outcome than Sam’s buying one for himself, even if the outcomes of both acts would include Sam’s enjoying ice cream. And when outcomes differ, they may have differing values and hence different rankings. When potential outcomes are truly identical, however, standard versions of act-consequentialism keep values and rankings the same regardless of the identity of the agent. To put the point more carefully, a particular outcome $o_1$ cannot, when considered as one among some particular range of outcomes $o_1 \ldots o_n$, have one value and ranking if one agent brings it about and another if a different agent faced with the same alternative outcomes brings it about.\(^{137}\)

As Portmore notes, in theory, a consequentialist could reject the above assumption and say that values and rankings of outcomes are relative to agents. Some versions of ethical egoism might be construed as doing this by making the welfare of the agent the only factor relevant to rankings, so that higher benefits to the agent receive higher rankings (relative to that agent). An opposing way in which the values and rankings of outcomes could be relative to agents would involve outcomes (and by extension actions) being ranked based on the welfare that accrues to everyone *except* the agent.\(^{138}\) On such a theory, actions open to an agent that benefit the agent alone would always be ranked lower (relative to the agent) than actions that would provide a net

\(^{137}\) Of course, if the outcomes differ, their values can differ.

\(^{138}\) Portmore, 410.
benefit to others. So, if Betty were deciding between treating Sam to an ice cream cone and treating herself, then even if Betty would derive more benefit from a cone than Sam would, so long as Sam would receive some positive net benefit from one, the state of affairs consisting in Sam’s enjoying a cone would be ranked higher, relative to Betty, than the state of affairs consisting in Betty’s enjoying one herself. Accordingly, the act of treating Sam would also be ranked higher (relative to Betty) than the act of treating herself. Of course, were Sam the one deciding whether to treat Betty or himself, the rankings would flip, and treating Betty would be ranked higher (relative to Sam) than treating himself.

An additional traditional feature of consequentialism that Portmore identifies and questions is its propensity to rank outcomes on a single scale (e.g., total welfare) that determines obligation. A more “unorthodox” version of consequentialism could, Portmore notes, rank outcomes on multiple, independent scales, and then use all of these rankings to determine obligation. 139 For instance, what Portmore calls “Self/Other Utilitarianism” ranks acts first based on how much well-being they provide for people other than the agent and then based on how much well-being they provide overall. Obligation becomes a particular function of both of these rankings in the following way: people are forbidden from performing an act if and only if some alternative act will be both better for others and better overall. 140 So (roughly), on Self/Other Utilitarianism, an agent is permitted either to promote the good of others exclusively or to maximize overall well-being, which takes into account the agent’s own good. He is not permitted,

139 Portmore, 410.
140 Portmore, 410.
however, to perform an action that serves neither the overall best interests of everyone nor the overall best interests of everyone other than himself. Portmore uses the label “dual ranking act-consequentialism” (DRAC) to denote any such theory “that ranks outcomes in terms of two auxiliary rankings,” whatever those two auxiliary rankings may be.  

Crucially, Portmore holds that a version of DRAC, unlike traditional (single-ranking) forms of act-consequentialism, can accommodate both supererogation and some other important features of common-sense morality, such as agent-centered options and self-other symmetry. Portmore does not say that all forms of DRAC allow for supererogation. Instead, he defends a particular version of DRAC that, he thinks, plausibly captures it. This form of DRAC is what Portmore calls “Schefflerian Utilitarianism” (SU). Notably, SU, in addition to being dual-ranking, is agent-relative with respect to the values of consequences. SU is intended to take seriously the idea that an agent is permitted to treat the welfare of others differently than she treats her own welfare. In particular, she is in entitled both to overweight and underweight her own welfare (against others’) when acting, doing more for herself on some occasions than she does for others and sacrificing her own interests on other occasions for the sake of others. To illustrate, while Sam is not permitted to ignore others’ interests entirely when he acts, he is permitted to take the last cupcake for himself even if Betty wants it more, and he is also permitted to leave it for her even if he wants it more. Portmore wants to appeal to SU’s recognition of this self/other asymmetry in order to be able to say that when the

\[141\] Portmore, 410.

\[142\] SU is similar but not identical to Scheffler’s preferred Distributive Hybrid Theory.
agent chooses not to overweight her own welfare and thereby does more for others than she is obligated to do, she acts in a supererogatory manner.

To understand how SU is supposed to allow for supererogation, we first must understand SU’s own two auxiliary rankings of consequences. These two auxiliary rankings are (i) the total overall well-being of everyone other than the agent and (ii) what Portmore calls “total adjusted well-being.” Portmore assumes that well-being can be quantified numerically in terms of the ever-convenient utile. If he is correct, the total overall well-being of everyone other than the agent will have a straightforward numeric value in terms of utiles.143 “Total adjusted well-being” is also assumed to be numerically quantifiable, but the way in which it is calculated is not quite as straightforward. Total adjusted well-being includes the well-being of both the agent and others, but it is agent-relative in that it reflects a heavier weighting of the well-being of the agent who brings about the consequences in question than it does of the well-being of each of the other individuals involved. Following Scheffler, Portmore quantifies this heavier weighting numerically, weighting the agent’s welfare as ten times as important as any other individual’s welfare.144 Thus, for a particular state of affairs that results from a particular agent’s act, if the well-being of others in that state of affairs combines for a value of 34 utiles, and if the well-being of the agent has a value of 2 utiles, the total number of utiles realized in that state of affairs will be 36, but the total-adjusted well-being of that state of affairs will be 54 (34 + 2 x 10).

__________________

143 Portmore, 411-2.

144 Portmore, 411.
On SU, the above two auxiliary rankings combine to determine obligation in the following way: an act is permissible “if and only if there is no available alternative act whose outcome contains both (1) more total well-being for others and (2) at least as much total adjusted well-being.” So, if an agent’s act has at least one alternative act that is both better for others (in terms of well-being) and as good or better on the scale of total adjusted well-being, the agent’s act is impermissible. If no alternative act satisfies both of these conditions, the agent’s act is permissible. The intuitive idea is that, while one is obligated to take the interests of all affected parties into consideration when one acts, one is not obligated to treat oneself as just one among many; one can weight one’s own interests more heavily. If one decides not to do so, though, or even if one decides to ignore one’s own interests entirely, sacrificing one’s own good for the good of others, that is permissible as well. What one must not do is ignore the interests of others entirely or weight one’s own interests too heavily. One also must not turn down available “win-win” or “something for nothing” opportunities—opportunities to benefit the overall interests of either oneself or others (or both) without hurting the overall interests of either oneself or others.

With this understanding of SU’s account of obligation, we are now in a position to understand how Portmore sees it as providing for supererogation. On SU, says Portmore, supererogation occurs whenever an agent performs an act that is permissible and whose outcome contains more total well-being for others than some permissible

---

145 Portmore, 412.
alternative. In all such situations, the agent will be sacrificing his own good to do
more for others than he is morally obligated to do and thereby can be said to go above
and beyond the call of moral duty.

With respect to its verdicts on particular cases, SU is a marked improvement over
the strategies of both Urmson and Attfield. Significantly, SU does not cheapen
supererogation as, we have seen, the strategies of Urmson and Attfield do. Part of the
reason that supererogation is so cheap on both Urmson’s and Attfield’s accounts is that,
on those accounts, supererogation often requires minimal effort and no self-sacrifice: for
both Urmson and Attfield, merely doing slightly more for oneself than one is obligated to
do is supererogatory. Since the standards for satisfying duty are also very low (in most
everyday situations) on the Urmson and Attfield accounts, supererogation thus becomes
commonplace. On SU, by contrast, supererogation occurs only when self-sacrifice for
the sake of another is involved and hence always costs the agent something. Because
Portmore’s account rarely requires maximizing utility (and then only when doing so
likewise maximizes the agent’s well-being or in the extreme cases when doing so
produces ten times more good for others than any available option produces for the
agent), SU also less frequently runs into the opposite problem of forbidding highly self-
sacrificial actions that nonetheless (barely) fail to maximize utility.  

\[146\] Portmore, 412-3.

\[147\] It runs into this problem in cases aptly described as highly self-sacrificial when the cost to the
agent is objectively high (e.g., her life or livelihood) but low relative to the great goods at stake for others.
If the agent’s options are to live and let 12 die, die peacefully and let 11 die, or die in tremendous pain and
prevent all of the other 12 from dying, SU probably requires her to maximize and choose the third option.
Arguably, however, the second option is (at the very least) acceptable.
More specifically, to its credit, Portmore’s SU, unlike the theories endorsed by Attfield and Urmson, is able to recognize the grenade case as involving supererogation. Indeed, SU seems almost designed to accommodate such a case in which one person sacrifices his own self-interest and saves another.

Still, SU’s record on accommodating paradigmatic cases of supererogation is not perfect: Kocheleff’s work in Burundi and South Africa is another paradigmatic case of supererogation, but it will not qualify as supererogatory, on SU, because in this case the stakes are too high for even a ten-fold overweighting of Kocheleff’s own interests to offset his doing markedly less than he does to aid others. On this point, a defender of SU might insist that the harms Kocheleff was fighting were too great to allow, which is why, in all his years in Africa, Kocheleff was merely fulfilling his duty while others were continually shirking theirs. Even given utilitarian values and the good that Kocheleff was able to achieve, however, the great cost Kocheleff incurred in living the sort of life that he has led makes this reading of his life’s work less than fully satisfying. Now, Portmore says that his selection of the number ten as the overweighting amount for the agent’s own interests is “for the purposes of illustration only,” so I am perhaps making too much of the current version of SU’s inability to accommodate this particular case of supererogation; substituting a sufficiently higher number would solve the problem. Still, a problem remains: on SU, anything may be demanded of a person provided the potential benefit is sufficiently high (how high is sufficient simply depends on the cost to the agent and the overweighting constant selected). But while the demands of morality can be

\[148\] Portmore, 411.
great, there are limits to what morality can require of any particular person (with respect
to sexual self-determination, for instance); no overweighting constant can recognize these
limits without exaggerating the extent to which a person is generally permitted to
overweight her own interests.

Despite the merits of (most of) its verdicts on particular cases, SU has some
serious theoretical shortcomings. One of these problems stems from SU’s having two
ways of ranking potential consequences. Where the consequence of a potential action
stands on each of the rankings determines whether or not the action is permissible. Of
permissible actions, however, only one of the ranking systems—total well-being for
others—tells us whether or not an action is superior to an alternative; only by surpassing
a permissible action on this score can an action be considered “beyond” the call of duty.
We might well ask, though, “why should an action’s rank (relative to other permissible
acts) on this score, and only this score, determine whether or not it is beyond duty?” If
both overall wellbeing for others and total adjusted well-being are legitimate ways of
ranking consequences, we should expect a dual-ranking system that incorporates these
two rankings to yield two ways of surpassing duty and by extension two forms of
supererogation: one form occurs when one performs a permissible action that produces
more well-being for others than required; the other occurs when one performs a
permissible action that produces more total adjusted well-being than required. Until we
have a reason, rooted in utilitarianism, for privileging total well-being for others over
total adjusted well-being when evaluating permissible actions, SU’s doing so appears *ad
hoc.*
A second and potentially more serious objection is that total adjusted well-being does not provide a legitimate way of ranking consequences. SU purports to rank consequences based on their levels of total adjusted well-being, but it cannot do so because total adjusted well-being is a fiction. Accusing utilitarian moral theories of trading in a moral fiction is not a new strategy. Alasdair MacIntyre has argued that utility itself is such a fiction.\textsuperscript{149} And, if MacIntyre is correct, total adjusted well-being also turns out to be a fiction insofar as the very definition of total adjusted well-being essentially appeals to the reality of utility and its base unit, the utile. My objection does not depend on MacIntyre’s, however. Even if we grant the reality of utility and utiles, total adjusted well-being still turns out to be a fiction.

To see why total adjusted well-being turns out to be a fiction, we need only return to my earlier explanation of what it is and how it is calculated:

Total adjusted well-being includes the well-being of both the agent and others, but it is agent-relative in that it reflects a heavier weighting of the well-being of the agent who brings about the consequences in question than it does of the well-being of each of the other individuals involved...[Numerically, it weights] the agent’s welfare as ten times as important as any other individual’s welfare. Thus, for a particular state of affairs that results from a particular agent’s act, if the well-being of others in that state of affairs combines for a value of 34 utiles, and if the well-being of the agent has a value of 2 utiles, the total number of utiles realized in that state of affairs will be 36, but the total-adjusted well-being of that state of affairs will be 54 (34 + 2 x 10).

\textsuperscript{149} The argument that MacIntyre provides needs some filling in, but it proceeds roughly as follows: while pleasures are real, individual pleasures—the pleasure-of-drinking-Guinness and the pleasure-of-laying-on-a-beach, for instance—are such widely disparate creatures that we cannot say that two experiences, each of which involves a very different pleasure, both contain quantities (of whatever amount) of the very same stuff. There may be a resemblance between the two experiences that explains why we characterize both as “pleasurable,” but there is no single thing called “pleasure” or “utility” contained in both experiences. Utility is thus not a thing, and it is a fiction. See After Virtue, second ed., (Notre Dame, IN: University of Notre Dame Press, 1984), 63-4.
Notice that, in discussing the example, I did not say that the total adjusted well-being in the example situation amounts to 54 *utiles*. That is because the example situation does not contain 54 utiles of total adjusted well-being; the whole situation contains only 36 total utiles, so it cannot at the same time contain a greater number of utiles or of some particular class of utiles. But if the example situation does not contain 54 utiles of total adjusted well-being, then we must ask: what is the significance of saying that the total adjusted well-being of the example situation is 54?

The answer is that Portmore is essentially assigning a score to the example situation on a scale that he has developed and that he calls “total adjusted well-being,” while this score may be determined by real, significant features of the situation, namely, the utility distribution of the situation in question and that distribution’s relation to the agent whose action (or potential action) is under consideration, the score does not correspond to some significant property of the situation (relative to a particular agent). Portmore has a formula for assigning the situation a score of 54 (for a particular agent), but it is not the case that all situations receiving scores of 54 have anything of significance in common, that, for instance, there is some significant property of “total adjusted well-being” that they have to the same degree. This is what I mean when I say that total adjusted well-being is a fiction: there is no significant property in the world (or in particular situations) that Portmore’s total adjusted well-being measures or to which it corresponds in the way that “total attendance,” say, corresponds to the number of persons in a room, “total calories” measures energy derivable from a particular food, “average speed” measures the rate at which an object travels a distance, or even “personal commitment” corresponds to the degree to which an individual values something.
Now, I cannot *prove* that total adjusted well-being is a fiction. Portmore has his formula, and nothing prevents him from saying that if his formula assigns a 50 to a particular situation/agent combination, it follows that the situation/agent combination contains 50 total adjusted wellbeing units ("taw units," for short) or exemplifies some property to the degree 50, which is 30 more taw units or a 1.5 times stronger exemplification of that property than is contained in (or exemplified by) any situation/agent combination to which his formula assigns a 20. Nonetheless, belief in these taw units or this property of total adjusted well-being is unmotivated. We have no way of conceiving what *sort* of stuff a taw unit is a unit of or even what type of property is being measured. (The problem is not merely that we aren’t sure how “big” one might be). The only reason we have for believing in total adjusted well-being is that Portmore has introduced a formula for calculating it. By parity of reasoning, however, if I develop a formula for ranking colleges (in terms of graduation rates, acceptance rates, SAT scores of incoming freshmen, or the like), we have good reason to believe that the score my formula assigns to each school corresponds to some common thing or significant property that the schools have to varying degrees. But this suggestion is absurd. My formula might take the average SAT score of incoming freshman, divide this score by two hundred, and subtract average GPA to arrive at a final score. In this way, a university where the average SAT score is 1300 and the average GPA is 3.5 would have the same score (namely, 3) as a university where the average SAT score is 900 and the average GPA 1.5. But even if these two schools were to have the same score, it does not follow that they have anything of substance in common; indeed, learning that the two have the same score on my ranking system tells us nothing about the two schools or any
significant relationships that hold between them. Note that the point is not that my ranking system does not adequately measure, say, the relative academic strength of the schools or of their student bodies; the point is that my ranking system does not measure anything at all. If it did, we could learn something about them and some relationship they bear to other schools by learning that the two schools have the same score of 3, that a third school has a higher score of 3.8, and a fourth a lower one of 2. But, in fact, such information is devoid of significance. And the problem is not (or, at any rate, is not simply) that the inputs of my scale—average SAT scores and grade point averages—are themselves devoid of significance. Even if we were to grant that SAT scores and GPA signify something (as I am granting that utility is not a fiction), my ranking system would still measure nothing of significance. Guaranteeing that the inputs are significant no more guarantees that the outputs are than guaranteeing that each member of a set of words is meaningful guarantees that any combination of these words is likewise meaningful.

If it is objected that total adjusted wellbeing is not a fiction in that utility that accrues to the agent is in fact more valuable relative to the agent than is utility that accrues to others and that total adjusted well-being reflects this fact, it should be pointed out that, while the point about relative utility values may be true, the ranking system under consideration either ranks consequences in terms of the value they have relative to agent or it does not. If it does, then we should ask why utility that accrues to people completely unconnected to the agent should be counted at all. If it does not, then pointing out that utility that accrues to the agent is more valuable relative to the agent no more justifies overweighting the interests of the agent than pointing out that utility that
accrues to others has greater value relative to others would justify overweighting others’ interests.

Any acceptable utilitarian system of ranking consequences must be non-artificial and reflect relative differences in the degrees to which some valuable feature or thing is genuinely present in those consequences. Portmore’s total adjusted well-being ranking system does not do this. As we have seen, it does not rank consequences on the degree to which they realize some actual, significant property or unit of value. Rather, it assesses situations based on a single unit of value—the utile—and then artificially manipulates the base assessments, overweighting utility that accrues to the agent by a factor of ten not because utiles that accrue to the agent are ten times as valuable as utiles that accrue to others but because doing so is more likely to yield the results that Portmore wants with regard to moral obligation. What total adjusted well-being is really capturing is not something or some property prior to moral obligation (and partially determinant of it) that exists in a particular state of affairs, or even in a state of affairs/agent combination, but something about our moral intuitions. Portmore’s total adjusted well-being ranking system thus trades in a moral accounting fiction and thereby departs from the heart of the utilitarian picture.

* * * * *

Significant departure from the utilitarian picture may not be a problem for Portmore, who is explicitly less interested in developing a theory that philosophers will recognize as a form of consequentialism than in defending a plausible moral theory.\textsuperscript{150}

\textsuperscript{150} Portmore, 424-5.
but departure is a problem for the utilitarian who, *qua* utilitarian, wants to adopt Portmore’s strategy to allow for supererogation. A departure problem likewise afflicts the accommodation strategy that we saw in Urmson. Utilitarianism starts with the position that the world has a serious claim on us: we are to contribute to the world’s improvement rather than its destruction; this is the most important thing we could ever do, and we do it by bringing about good states of affairs. The claims the world has on us in Urmson’s picture, however, are so paltry as to undermine the urgency that the prospect of improving the world is, on utilitarianism, supposed to carry. Of the three theories we have considered, then, Attfield’s remains closest to the heart of utilitarianism. Its doing so is, however, little comfort in that the account of supererogation it provides is a feeble one, one on which one’s best bet for achieving moral sainthood is not to follow in the path of a Gandhi or a Mother Theresa but to make a conscientious effort regularly to give up one’s seat on crowded trains. What our investigation of utilitarian attempts to accommodate supererogation thus so far indicates is that a robust theory of supererogation cannot be had without seriously compromising utilitarian commitments. We will revisit this question (and look for a final verdict on it) in chapter five.
CHAPTER THREE

KANTIANISM AND SUPEREROGATION

The prospects for an affirmative answer to the question, “can Kantianism accommodate supererogation?” initially look unpromising. A very general challenge for Kantianism is one that confronts all forms of deontology: acts of supererogation are supposed to be good without being obligatory, which never happens, using “right” in its traditional, obligation-entailing sense, if all good acts are right acts, as is presumably the case if deontology is correct and the right is ultimately prior to the good.\(^\text{151}\)

A somewhat less general case against Kantian supererogation emerges from the fact that the very foundation of Kantian morality is rationality (or, more correctly, freedom, which for Kant implies that rationality has supreme authority). Put baldly, the rationality problem is that Kantianism makes morality purely an issue of practical rationality, but one cannot be “super rational.” To be less crude: for Kant, moral value is a function of rationality, which is in turn a question of conformity to the categorical

\(^{151}\) Thus, Richard McCarty in “The Limits of Kantian Duty, and Beyond,” *American Philosophical Quarterly*, 26 (1989): “Any ethical theory which, following the Kantian turn, derives moral goodness from obligation, instead of the other way around, precludes in its very foundation the possibility of morally good, non-obligatory actions,” 43. McCarty does not, however, conclude that Kantians must reject the possibility of supererogation. Instead, McCarty draws on Kant’s aesthetic theory to develop what McCarty calls “a ‘quasi-moral’ account of the value of supererogation,” 43.
imperative (hereafter the CI). Since conformity to the CI does not admit of degrees, it follows that rationality and hence moral value likewise do not admit of degrees. Thus, if an act does not conform to the CI, it is not supererogatory because it is impermissible. If it does so-conform, it is permissible and has moral value but, because moral value does not admit of degrees, its moral value is no greater than that of any other permissible action. Hence, on Kantianism, supererogation is not possible.

Reinforcing the above conclusion are Kant’s own explicitly stated views on the extent of moral obligation and duty. Although Kant never states plainly, “supererogation is impossible,” some of his discussions of merit and of “moral enthusiasm” for extraordinary actions would do the Protestant Reformers proud. Consider:

By exhortation to actions as noble, sublime, and magnanimous, minds are attuned to nothing but moral enthusiasm and exaggerated self-conceit; by such exhortations they are led into the delusion that it is not duty—that is, respect for the law whose yoke…they must bear, even if reluctantly—which constitutes the determining ground of their actions,…but that it is as if those actions are expected of them, not from duty but as bare merit. For, when they imitate such deeds—namely, from such a principle—not only have they quite failed to fulfill the spirit of the law…; not only do they locate the incentive pathologically (in sympathy or self-love), not morally (in the law); but they produce in this way a frivolous, high-flown, fantastic cast of mind, flattering themselves with a spontaneous goodness of heart that needs neither spur nor bridle and for which not even a command is necessary and thereby forgetting their obligation, which they ought to think of rather than merit.153

152 One might object that conformity with the CI is merely a necessary condition for rationality or for minimal rationality, which leaves open the possibility of degrees of rationality. Such an objection does not readily make room for supererogation, however, because even if nothing Kant says strictly entails that rationality does not come in degrees, Kant gives us no indication that rationality comes in degrees or how it might do so.

153 Critique of Practical Reason, trans. by Mary Gregor (Cambridge: Cambridge University Press, 1997), 5:84-85. Later, Kant makes a similar observation about moral education, saying:

But I do wish that educators would spare their pupils examples of so-called noble (supermeritorious) actions, with which our sentimental writings so abound, and would expose them
Praising certain actions as “noble” and “sublime” is thus, to Kant, highly dangerous: it encourages people to act from emotion (i.e., from such “pathological” impulses as sympathy and self-love) rather than from duty, to underestimate their obligations, and to be self-deceived about their own strength of character.

To be sure, if we continue reading we learn that, for Kant, sometimes praise of extraordinary actions is acceptable:

Actions of others that are done with great sacrifice for the sake of duty alone may indeed be praised by calling them noble or sublime deeds, but only insofar as there are traces suggesting that they were done wholly from respect for duty and not from ebullitions of feeling. But if one wants to represent these to someone as examples to be imitated, respect for duty...must be used as the incentive—this earnest, holy precept that does not leave it to our vain self-love to dally with pathological impulses (as far as they are analogous to morality) and to credit ourselves with meritorious worth. If only we search carefully we shall find for all actions that are praiseworthy a law of duty, which commands and does not leave it to our discretion to choose what may be agreeable to our propensity.¹⁵⁴

Thus, some extraordinary actions can be highly praised, but only insofar as they are motivated by duty, as they all can and should be motivated, since any action that is praiseworthy is non-optional but rather fulfills some duty. If what Kant says is true, it follows that supererogatory action—action that is morally praiseworthy but not optional—is not possible.

¹⁵⁴ *Critique of Practical Reason*, 5:85.
Kantian scholars often agree with the above conclusion that Kant’s moral theory precludes the possibility of supererogation. As with utilitarianism, however, some have tried to make room for supererogation within a Kantian framework. Most famously, Thomas Hill has tried to make room for it by drawing on the distinction between perfect and imperfect duties. Alexandar Jokic locates Kantian merit, but not obligation, in an act’s motive, and develops a way for a meritorious motive to prompt a non-obligatory act. In what follows, I will be discussing these two Kantian accommodation strategies with the purpose of gaining some clarity on the prospects that Kantianism has of accommodating supererogation. My discussion admittedly leaves many possible and even some published Kantian accommodation strategies uninvestigated, a consequence of considerations of economy.

3.1 Hill and Wide Imperfect Duties

Of attempts to reconcile Kantianism with supererogation, Thomas Hill’s is the most well-known. Hill’s Kantian account centers on the distinction between wide (or “paradigm”) imperfect duties—not to be confused with imperfect duties simpliciter—and all other duties, i.e., perfect duties and more narrow imperfect duties. The crucial difference between these two types of duty has to do with how much latitude they allow with respect to their fulfillment. Principles of perfect duty and narrow imperfect duty, unlike principles of wide imperfect duty, are of the “always” or “never” form, which is to

say that when they recommend a particular action (or counsel against it), the action in question is mandatory (or forbidden). By contrast, the principles of wide imperfect duty may recommend a particular action (or counsel against it) even if, strictly speaking, one need not be violating moral duty by refraining from (or performing) the action in a given situation in which it is an option. Thus, a principle of perfect duty forbids lying, a principle of narrow imperfect duty recommends treating others in a respectful manner, and a principle of wide imperfect duty recommends helping others when one can. If one lies or treats another person in a disrespectful manner, it follows that one has violated duty; if one fails to help another person in a given situation in which the opportunity arises, however, it is possible that no violation of duty has occurred.

In Metaphysics of Morals, Kant states that actions can have negative, positive, or neutral moral worth. Kant even denotes these differences in worth using quasi-mathematical symbols: (+a) denotes positive moral worth, (-a) negative, and (0) an absence of moral worth. Significantly, says Hill, according to Kant, when a person has the opportunity to perform an action that a principle of wide imperfect duty recommends, one might think the idea of an absolute imperfect duty is a contradiction in terms, since imperfect duties admit of “latitude” in their fulfillment. Nothing in this discussion turns on this issue, but it is worth noting that Kant’s justification of the perfect/imperfect distinction is not in terms of absolute and non-absolute duties (or even in terms of latitude) but in relation to the CI: the way in which a given violation of the CI violates the CI determines whether the violation is a violation of perfect or imperfect duty. An argument would be required to show that the ways of violating the CI that count as violations of imperfect duty turn out to make all imperfect duties non-absolute in the relevant sense. It should also be remembered that if imperfect duties are considered “non-absolute” in some sense, this should not be a sense that entails the denial of the claim that one’s imperfect duties fall out of the Categorical Imperative, which places an absolute (“Act only …!”) requirement on us.


157 One might think the idea of an absolute imperfect duty is a contradiction in terms, since imperfect duties admit of “latitude” in their fulfillment. Nothing in this discussion turns on this issue, but it is worth noting that Kant’s justification of the perfect/imperfect distinction is not in terms of absolute and non-absolute duties (or even in terms of latitude) but in relation to the CI: the way in which a given violation of the CI violates the CI determines whether the violation is a violation of perfect or imperfect duty. An argument would be required to show that the ways of violating the CI that count as violations of imperfect duty turn out to make all imperfect duties non-absolute in the relevant sense. It should also be remembered that if imperfect duties are considered “non-absolute” in some sense, this should not be a sense that entails the denial of the claim that one’s imperfect duties fall out of the Categorical Imperative, which places an absolute (“Act only …!”) requirement on us.

this omission is generally of neutral—rather than negative—moral worth;\(^\text{159}\) performing the recommended action from a motive of duty, however, is always of positive moral worth provided the action is not forbidden (as it would be if, for instance, it violated a perfect duty). Thus, says Hill, supererogatory acts—acts of moral worth whose omission would not be blameworthy—are possible.\(^\text{160}\) Roughly, such an act occurs when a person, from the right motive, performs an act recommended by a principle of wide imperfect duty in one of the many situations in which that act is neither forbidden nor required.

As he presents them, Hill’s precise criteria for Kantian supererogation are that supererogation occurs when (and presumably only when) an act

(a) is of a sort commended by a principle of wider imperfect duty,
(b) is motivated by a sense of duty (or perhaps, respect for moral reasons),
(c) is neither forbidden nor required by another, more stringent duty…,
(d) is in a context where no alternative is required by more stringent duty and there is at least one alternative that is neither forbidden by more stringent duty nor commended by other principles of wide duty, and
(e) is done by an agent who has adopted the relevant principle of wider imperfect duty and has often and continually acted on that principle.\(^\text{161}\)

In typical cases, then, in which an agent who has already consistently acted on a principle of wide imperfect duty does so (again) out of respect for moral principles, the agent’s act is supererogatory.

Hill goes on to provide an example of an act that would satisfy the above criteria and hence qualify as an act of Kantian supererogation, a meritorious but non-required act.

---

\(^\text{159}\) The exceptions in which refraining from the action is of negative moral worth occur either due to the agent’s lack of past commitment to the relevant principle of wide imperfect duty or due to the requirements that other duties impose in the situation in question. Hill, 70-1.

\(^\text{160}\) Hill, 70-1.

\(^\text{161}\) Hill, 71.
In Hill’s example, a man who has faithfully given to charity all his life gives a treat to a child. In Hill’s imagined situation, no special circumstances render the gift either required by or in violation of some principle of duty, perfect or imperfect. Hill further stipulates that the man, in treating the child, is motivated solely by duty. As Hill himself admits, however, in the circumstances in question, the man’s past actions make the current gift non-required, which raises the question of how the man could be acting from a sense of duty. Hill explains: “Far from rejecting Kant’s principle of beneficence, [the man] is disposed out of a (perhaps misguided) respect for it to do more of the sort of act that it directs than it strictly requires.”

In this way, the principle of beneficence that the man, like everyone else, is required to adopt and whose adoption is consistent with refraining from giving to the child in the current situation can nonetheless motivate the gift. With such a motive, the act is, according to Kant’s reckoning, of positive moral worth (+a); its omission, as we have seen, would have been of neutral moral worth (0); hence, the act is beyond moral duty and (says Hill) can be seen as supererogatory on Kant’s own terms, given typical understandings of what is involved in supererogation.

As the most well-known attempt to reconcile supererogation with Kantian moral theory, Hill’s account has been subject to a wide array of objections. Here, I want to limit my attention to a small set of related objections to Hill’s suggestion for finding a place for supererogation within Kantianism. My own understanding of what is going on in Hill’s work is that in order to find a home for the supererogatory within Kant’s system, Hill has found some passages in the Metaphysics of Morals that suggest that Kant has a

---

162 Hill, 72, emphasis mine.

163 Hill, 71-2.
very specific accounting system for calculating moral worth and that this accounting system, taken together with Kant’s understanding of obligation, implies that morally meritorious but non-required action is possible. Hill’s conclusion is that Kantian morality accommodates—or at least can accommodate—supererogation. Hill’s mistake is that, even if the moral accounting system he presents is to a large degree grounded in the *Metaphysics of Morals*, his particular interpretation of this system is nonetheless not sufficiently so-grounded, and it is at odds with Kant’s moral theory, especially the justification of moral obligation that that theory involves.

We can begin to see that HAS conflicts with Kant’s moral system by considering to what, on HAS, moral merit attaches. On HAS, moral merit (+a) accrues when a person, from a sense of duty, either performs an act that fulfills a duty (perfect or imperfect) or performs an act that does not fulfill or violate a duty but that is nonetheless motivated by a principle of wide imperfect duty. On such a system, moral merit attaches to *discrete, properly motivated, external acts*. We can see that it attaches to (this particular subset of) external acts and not, say, to motive alone by considering two individuals who have adopted the same principles of action and, in particular, have both adopted a principle of beneficence but whose life situations are very different. One of the two is rich, able-bodied, and talented. The other is financially dependent on others and suffers from cerebral palsy to the extent that he is confined to a wheel chair, has little control over his own bodily movements, has limited communication abilities, but is fully cognizant of the people and world around him. On the Kantian picture, both of these two can be equally committed to the principle of beneficence. Yet if both are fully committed, on HAS, almost assuredly the first of the two will accrue more moral merit.
with respect to beneficence than the second. This is because the first will much more frequently act on the principle of beneficence because she will have much more frequent opportunities to do so, and HAS dispenses moral merit only when such discrete external acts (albeit properly motivated) are performed and not merely when the motivating principle is adopted or present.

A problem with HAS, for the Kantian, should now be apparent: HAS makes moral worth (partly) accidental, and the Kantian cannot allow this. Moral merit, for the Kantian, is not an issue of moral luck and hence does not accrue in virtue of an agent’s external acts. Rather, as the *Groundwork* emphatically tells us, moral worth is only to be found in the good will, and the value of this will does not diminish even if it “should be wholly lacking in the power to accomplish its purpose,” a strong claim that should alert us to the extent to which Kant emphasizes will over external act. For Kant, then, having a good will neither is equivalent to nor implies performing a particular discrete external action (or set of actions) from the proper motivation. The performance of these particular external actions is rather the typical result of possession of such a will. Indeed, as Kant is quick to inform us, the good will under discussion is not identical to “a mere wish;” having it implies “the summoning of all the means in our power.”

As Hill himself emphatically notes, Kant makes clear in the *Metaphysics of Morals* that, in the case of imperfect duty, what morality requires of us is not that we

164 For more discussion on why, for Kant, the accrual of moral merit never depends on moral luck, see the discussion of Alexandar Jokic’s Kantian account of supererogation in the second half of this chapter.


166 *Grounding*, 4: 394.
perform particular external actions but that we adopt particular principles, the principles of imperfect duty. Of course, if we adopt such principles, then, given the right opportunities, we will not fail to act in certain ways. No one adopts the principle of beneficence and then, given multiple opportunities to act on it, fails to do so. Nonetheless, speaking strictly, it is the adoption of the principle, or perhaps the sustained adherence to it, not the performance of action, that is required, that is not a matter of moral luck, and that hence should and does bear moral merit on Kantianism.

What Hill is trying to do is to say, in the case of imperfect duties, both that adoption of principle is what is required and that each discrete performance of a properly motivated external act generates merit. Although it is possible for both of these claims to be true, it is counterintuitive that they should both be true and, as we have seen, the second claim conflicts with Kantian concerns about luck. Hill tries to support HAS, insofar as it implies the problematic second claim, by pointing to passages in *Metaphysics of Morals*. If we examine the passages to which Hill appeals, however, we will find that the passages referenced offer less than fully convincing support for HAS. Consider first the following passage:

If someone does more in the way of duty than he can be coerced by the law to do, what he does is meritorious (meritum); if what he does is just exactly what the law requires, he does what is owed (debitum); finally, if what he does is less than the law requires, it is morally culpable (demeritum)\(^\text{167}\)

In this passage, Kant tells us that it is possible to do more than the law can coerce, which is another way of saying that it is possible to do more than bare fulfillment of what Kant calls “juridical duty,” juridical duties being moral duties that others have the right to

enforce. Notably, all juridical duties are absolute, but not all absolute duties are juridical duties. Further, juridical duties require actions but not specific motives. The reason that no juridical duty requires a particular motive is that a requirement that an agent have a particular motive is unenforceable.\footnote{Metaphysics of Morals, 6: 390-1.} Finally, Kant uses the label “ethical duty” to refer to any non-juridical moral duty. If we use Hill’s terminology of “wide” and “narrow” imperfect duties, the class of ethical (non-juridical) duties thus includes the following: all wide imperfect duties; any perfect or narrow imperfect duties that, while absolute, do not imply that others have a right to enforce them; and duties to perform perfect duties from a motive of duty. This last set of duties, Kant says, is a subset of imperfect duties.\footnote{Metaphysics of Morals, 6: 391.}

If, as the passage quoted in the previous paragraph tells us, merit is a matter of “doing more” than what bare juridical duty requires, it is important to ask how one can so surpass juridical duty. Hill considers this question and offers the following response:

Doing more in the way of [juridical] duty can only be doing something which fulfills an ethical duty [i.e., a non-juridical duty], that is, either fulfills a perfect (e.g. juridical) duty from a motive of duty or conscientiously promotes an end prescribed by a principle of imperfect duty.\footnote{Hill, 67, emphasis mine.}

Hill says little to support the above assumption that the only two candidates for surpassing juridical duty are: (i) fulfilling a perfect duty from a motive of duty and (ii) conscientiously promoting an end prescribed by a principle of imperfect duty.\footnote{Likewise, Hill never explains why fulfilling a non-juridical perfect duty without the proper motive should not count as going beyond juridical duty. This lapse is surprising given that Hill does seem to think that some perfect duties are non-juridical.} In particular, Hill excludes without comment the possibility that one could surpass juridical

\footnote{Hill, 67, emphasis mine.}
duty by adopting a principle of imperfect duty. Note that so-adopting a principle of imperfect duty is not equivalent to (ii), conscientiously promoting an end prescribed by a principle of imperfect duty. The latter, unlike the former, strictly entails the performance of an external action. Hill never provides a reason for excluding the former. The absence of such a reason makes the exclusion unjustified, and it is doubly so given Kantian concerns about moral luck and the fact that, as already established, adopting particular principles (and not performing external actions) is what, as Hill admits, imperfect duties strictly require of us.

The unjustified exclusion of adopting a principle of imperfect duty as a way to surpass juridical duty leads Hill to move too quickly to the claim that conscientiously promoting an end prescribed by a principle of imperfect duty (construed as performing a certain sort of action from the right motive) sometimes is in fact meritorious and does count as surpassing juridical duty. The only other evidence Hill offers in support of his claim that conscientiously promoting an end prescribed by a principle of imperfect duty is a way of surpassing juridical duty is (a slightly shorter version of) the following passage in the *Metaphysics of Morals*:

> Imperfect duties are, accordingly, only duties of virtue. Fulfillment of them is *merit* (*meritum* = +a); but failure to fulfill them is not in itself *culpability* (*demeritum* = -a) but rather mere *deficiency in moral worth* (= o), unless the subject should make it his principle not to comply with such duties.  

Here, we see that fulfillment of an imperfect duty is meritorious. Recalling that fulfilling an imperfect duty is a matter not of performing a particular action but of adopting a particular principle (a point Hill explicitly endorses and to which he is committed), we

---

can see that the above passage does not license Hill’s claim that conscientiously promoting an end prescribed by a principle of imperfect duty is meritorious or surpasses juridical duty. Rather, what this passage tells us is that adopting principles of imperfect duty is meritorious and not adopting them implies a mere lack of worth (unless the reason that one fails to adopt them is that one has made it one’s principle not to adopt them, in which case one is culpable).

What Kant says in the next paragraph reaffirms this reading on which merit accrues when one adopts a particular end rather than when one performs a particular properly-motivated action. There, Kant says that an action’s merely being right or in conformity with the law does not make it meritorious; what makes for merit is “the conformity with Right of one’s maxims of [such right actions], as duties.” Hill may be able to agree with this point, since he thinks that, while merit attaches to individual external acts, it only attaches to these acts when they are accompanied by the right maxim. But if we ask why an action’s maxim’s conformity with Right allows for merit, Kant’s answer is that having such a maxim involves “respect for Right” and that in having such a maxim a person “makes the Right of humanity, or also the Right of men, his end.” To give these reasons, however, is not to say that the great features of a particular action/principle combination make for merit; it is to say that the great features of adopting the principle make for moral merit. Finally, the last point that Kant makes in this paragraph provides further confirmation of the point that the merit associated with fulfilling imperfect duties springs from adoption of principles rather than from the

\[^{173}\text{Metaphysics of Morals, 6: 390, emphasis mine.}\]

\[^{174}\text{Metaphysics of Morals, 6: 390-1.}\]
performance of particular external actions: Kant says that “to establish and quicken” a “disposition in oneself” to “act in conformity with duty from duty” is itself meritorious. Because such establishment/ quickening merely requires the adoption of a principle and does not strictly entail the performance of an external action, we thus have evidence that, for Kant, principle adoption is what carries positive moral merit.\footnote{\textit{Metaphysics of Morals}, 6: 391.}

The reason that it is important to get clear that, in the case of imperfect duties, principle adoption, rather than individual act performance, is not only required but also generates moral merit, is that this point ultimately undercuts Hill’s scheme for making room for supererogation in Kantian morality. On Hill’s picture, if two people both adopt the principle of beneficence, and one of these two performs benevolent acts much more frequently than the other, the over-performer’s extra acts of beneficence are supererogatory. Minimal acts of beneficence are required of all, but after some point, one accrues moral extra credit for non-required beneficent acts. And what holds for acts of beneficence also holds for acts motivated by the other principles of wide imperfect duty.\footnote{\textit{Metaphysics of Morals}, 6: 391.} But if Hill is wrong and merit attaches to principle adoption rather than to act performance, no such moral extra credit is to be had. If a person adopts the principle of beneficence then she will act on it with some frequency (assuming special circumstances do not prevent her from so-acting). Each of these individual acts, though, does not yield its own moral merit, and increasing their frequency does not yield increased merit. Beneficent acts, no matter how frequent, are thus not beyond duty. The merit associated
with beneficence (and other imperfect duties) comes only from adopting the principle of beneficence. Because this is something we all have a moral duty to do, even if we do not have a juridical (that is, enforceable) duty to do it, supererogation (i.e., surpassing moral duty) is not to be found in such principle-adoption.

To be sure, there are passages in Kant that suggest that moral worth attaches to actions. Kant frequently speaks of “the moral worth of an action,” so there must be a sense in which individual actions do have moral worth. The sense in which individual external actions have merit, however, must be a secondary sense because all the merit that an action has derives from the principle in the will from which it springs. This is why, as Kant says, that lone “unqualifiedly good” thing, the good will, “always holds first place in estimating the total worth of our actions.”

Thus, if two people both perform an act from a good will, while we can say that their acts are meritorious, this is just shorthand for the claim that the act of each springs from a will that has moral worth. Suppose that, within a two-day period, both individuals have the same good will (recalling what we have already established, that a good will is not some passing fancy but a deep-seated commitment), but one person performs only the one meritorious act and the second person performs two. It does not follow that the second person has accumulated twice as much merit in that two day period, as would be the case if what Nelson Potter calls the “Boyscout view of moral goodness,”—the view that “one must perform as many good deeds (well-motivated deeds) as possible, as if such an accumulation of discrete actions will come to constitute one’s moral goodness”—were

177 *Grounding*, 4: 401. emphasis mine
Rather, because moral worth is ultimately a function of the will, contra HAS, *only a change in the will can result in a change in merit,* and the two individuals in question, having the same good will, thus are equivalent (so long as their wills remain constant) in terms of merit.

So far, my remarks have been limited to the issue of whether HAS is properly grounded in Kant’s writings. An additional (albeit related) problem for Hill’s system emerges when we begin to ask *why* the acts that Hill counts as supererogatory are meritorious at all rather than of neutral value. As I earlier explained, the problem confronting any Kantian attempt to find a home for supererogation is that of explaining how a theory premised on rationality (which, once had, does not admit of degrees) should allow for supererogation. When we ask why fulfilling a moral duty, whether perfect or imperfect, has value, Kant’s theory allows us to find such value in the way such fulfillment conforms with the CI. Hill wants to say that a person can do something better than mere conformity, but although he has developed a system for identifying acts that (he thinks Kant says) so-surpass mere conformity, he has not explained why the acts he has identified as supererogatory do so-surpass mere conformity and do not simply

---

178 Nelson Potter, “Kant and the Moral Worth of Actions,” *The Southern Journal of Philosophy,* 34 (1996): 225-41. Unsurprisingly, Potter argues against the Boy scout view. More generally, he argues against the view that, for Kant, moral worth attaches to actions. Potter holds instead (as I do) that Kantian moral worth attaches to the will, and Potter further equates the will with character. In support of his claim that Kantian moral worth attaches to will rather than actions, Potter points to the fact that, in the *Groundwork,* Kant begins his discussion of moral worth by talking about the moral worth of the good will. It is only because actions are evidence of one’s character (and the only evidence we have of character), says Potter, that Kant discusses actions in the context of moral worth at all (227-9). Potter also points out that Kant’s focus, when he gives his famous four examples, is on what motives the example actions **reveal** (233). This focus on motive rather than act explains why none of Kant’s four examples involve actions that sprang from mixed motives: the point of the examples is to show the value of particular motives; it is not relevant to such value if some of these motives sometimes combine to influence a single act. If, however, Kant were interested in showing us value of the acts, then unless he thought that motives were irrelevant to the moral value of an act (and it is obvious that Kant does not think this), we would expect Kant to talk about mixed-motive acts as well as non-mixed ones (232-4).
constitute alternative ways of merely conforming with the CI. And we do not have a hint in Kant of how such an explanation of surpassing conformity would go. Purportedly supererogatory acts, after all, cannot be said to involve “super-conformity” to the CI, nor can the person who performs them be said to be “super-rational” or “ultra-free.” The Kantian justification for morality as found in the *Groundwork* and reaffirmed elsewhere simply does not extend to non-required acts. Nor is this feature of the *Groundwork* accidental: Kant’s stated goal in providing a justification of morality in the *Groundwork* is to explain how a *categorical* imperative is possible, not to explain what goes on when no categorical imperative applies. The upshot of all of this is that even if Hill is correct that the accounting system for moral merit that is found in the *Metaphysics of Morals* implies that an act can be non-required but morally meritorious, Hill is not yet out of the woods for he still needs to show that that accounting system has a Kantian theoretical justification. This is something that Hill has not yet done. More significantly, the peculiar form that Kant’s justification of morality takes gives us reason to be skeptical that such a Kantian justification of HAS is available.

3.2 Jokic and Moral Luck

One way of objecting to Hill’s account of Kantian supererogation appeals to the fact that Hill’s account locates moral value in supererogatory acts only when they are motivated by duty when in fact such a motive properly applies only to acts actually required by duty, and naturally none of these is supererogatory. More generally, one might say that Kant makes the duty motive a necessary condition of moral value, which renders the category, “morally good but not required,” necessarily empty (if even
intelligible) for the Kantian. Aleksandar Jokic’s view of Kantian moral value is of this sort. Nonetheless, Jokic still holds that a Kantian can make room for supererogation; the key for Jokic is to depart somewhat from the dominant, common sense conception of what supererogation is. That dominant conception says that supererogatory acts are (at the very least) morally good even though not morally required; Jokic’s departure is to deny that a supererogatory act need have moral value at all. Because Kant recognizes non-moral value, once the moral value requirement is abandoned, it becomes an open question whether some non-moral value attaches to some non-required acts and, in so attaching, renders them supererogatory. Jokic admits that not just any non-moral value is of the right sort to make supererogation possible: the acts that are believed to be supererogatory usually have consequentialist value, but the Kantian’s ability to affirm this fact about them does not imply an ability to affirm their supererogatory status, even if we add that the Kantian can see the acts in question as non-obligatory. For Jokic, what the Kantian needs is a way of saying, of supererogatory acts, both that they are non-required and that, while not morally praiseworthy, they are praiseworthy in a sense close enough to moral praiseworthiness.

Jokic finds such value (i.e., value that bears a sufficient affinity to moral value) by making an observation about how what he calls “moral luck” functions with respect to


180 Jokic, 225-6.

181 Jokic, 226.

182 Jokic, 225-6.
the motives present in each of the following types of actions: morally praiseworthy actions (i.e., actions that both fulfill and are motivated by duty), apparently supererogatory actions, actions that conform to the CI but are motivated by direct inclination (as in the case of the loving philanthropist), and actions that conform to the CI but are motivated by either self-love or what Jokic labels (but never defines) “indirect inclination,” which appears to be the sort of inclination that sometimes prompts acts that conform with duty but typically does not do so. According to Kant, says Jokic, a feature of the motive of duty that sets it apart from all other motives is that, without fail, when a person genuinely acts from the motive of duty, that person’s act conforms to the CI. In this way, we can say that the fact that the duty-prompted action conforms to the CI involves no moral luck. If, however, a person is acting from inclination (direct or indirect) or self-love, the issuing action’s conformity with the CI does depend in part on moral luck: in different circumstances, the same motive could have issued in an act that would have violated duty. When we ask about supererogatory acts, Jokic thinks that, given the extraordinary motives that prompt them, these acts do involve moral luck, but the amount of moral luck operative in them falls between that in acts done from duty and that in acts done from inclination. More specifically, Jokic says that supererogatory acts involve “barely more” moral luck than do acts performed from duty and that they involve

\[ \text{183 Jokic, 229-32.} \]

\[ \text{184 Note that the notion of “moral luck” at work here is based on the probability of a given motive issuing in a licit action and not, as is usual in discussions of “moral luck,” on the effect that an action’s contingent consequences have on its moral appraisal, 231.} \]
“infinitely less” moral luck than do acts merely in accord with duty that are done from self-love or indirect inclination.\(^{185}\)

If we accept Jokic’s claims about the levels of moral luck associated with various motives and acts, we find ourselves with a sliding scale of types of licit acts ranked according to the level of moral luck that each type involves. At one end of the scale, we have acts that conform with duty and that are motivated by duty and hence involve no moral luck. Beside these are supererogatory acts, which involve moral luck, but only in very small amounts. Next to supererogatory acts are licit acts performed from direct inclination. At the far end of the scale are licit acts performed from indirect inclination or self-love that involve a good measure of moral luck.

Jokic thinks that the Kantian can see a sliding scale of types of value that is inversely related to the above sliding scale of levels of moral luck, so that actions that involve the least luck (zero) will be associated with the highest type of value (namely, moral value), and licit actions that involve the most moral luck will be associated with a genuine type of value but one that is not nearly as good as moral value. We can begin to see specifically Kantian support for the idea that such a sliding scale of values is genuine by noticing an interesting fact about the acts at the end of the scale at which moral luck is not at all operative and that, for Kant, have moral worth: Kant says that the only thing that is good without qualification is a good will,\(^ {186}\) and the very definition of a good will is one such that its maxims “\emph{necessarily} harmonize with the laws of autonomy,”\(^ {187}\) in

\(^{185}\) Jokic, 231.

\(^{186}\) \textit{Grounding}, 4: 393, emphasis mine.

\(^{187}\) \textit{Grounding}, 4: 439.
other words, the good will is one that requires no moral luck for conformity to the CI. If we use Jokic’s language, we can thus say that, according to Kant, the reason that the acts prompted by a moral motive are morally good is that those motives/acts require/involves no moral luck. And if an absence of moral luck renders an action morally praiseworthy, it may well be the case that the presence of only a small amount of moral luck (“barely more” than none) renders an action praiseworthy in a way that is close to—but yet distinct from—moral praiseworthiness. More support for this idea can be found in the fact that acts from inclination that conform to the CI involve (if Jokic is correct) more moral luck than supererogatory actions do and yet Kant himself tells us that these acts from inclination, while not morally praiseworthy, are praiseworthy nonetheless. \footnote{Jokic also wants to support his view that a motive’s required level of moral luck has a significant bearing on its act’s value by appealing to intuition: seemingly, love is a better motive than self-love, and this corresponds to the (apparent) fact that acts from self-love that accord with duty involve more moral luck than do acts from love. While the intuition in question may be a common one, it will not help the Kantian accommodate supererogation anymore than the intuition that supererogation is a legitimate moral category will help the Kantian to accommodate supererogation. The question, after all, is what the Kantian \textit{quia} Kantian can say about the value and status of (intuitively) supererogatory acts, not what our intuitions tell us about the values of various motives and acts.} So, if Jokic is right about the level of moral luck involved in supererogatory actions relative to the moral luck involved in actions from duty and actions from (various forms of) inclination, then, given that Kant himself makes much of the amount of moral luck involved in actions from duty, it is reasonable to think that a Kantian can see supererogatory actions as, while lacking moral value, still having a type of value that approaches moral value and that comparatively dwarfs the non-moral value found in merely licit acts that are performed from inclination (direct or indirect) or self-love.
At this point, I should make a clarification about Jokic’s position. Jokic often presents his view in the way I have thus far presented it: as providing the Kantian with the resources to evaluate supererogatory acts as, while not morally praiseworthy, having a value that is significant and sufficiently close to moral worth to protect Kantianism from the objection that it places supererogatory acts in the dreaded “zone of indifference” in which morally neutral actions belong. But Jokic also frequently presents his view as stronger than (and strictly incompatible with) the position I have thus far described. On the stronger version of Jokic’s theory, the Kantian can recognize in supererogatory actions not merely a special type of value, but a special type of moral value. When Jokic presents his view in this way, he says that any licit action, however improperly motivated, has moral value of some sort, but there are gradations in types of such moral value. The highest form attaches to acts from duty; the second highest form of moral value attaches to supererogatory acts; and so on down the scale. So-classifying the value of supererogatory acts as one type of moral value is problematic for a number of reasons. In the first place, Jokic provides almost no support for the claim that supererogatory acts have (a type of) moral value, as opposed to the claim that they have some sort of value that bears some affinity to but is distinct from moral value. In the

189 Jokic, 223-4.

190 Jokic, 229-30.

191 The only positive evidence he provides is an argument from authority: Jokic tells us that Jeffrey Murphy argues for the conclusion that there are gradations in moral worth, but we are not given Murphy’s full argument. We are given a passage from Kant that Murphy uses to make a distinction that is essential for his argument, but the passage by itself (with the distinction it makes) is not enough to support Murphy’s conclusion. Moreover, appealing to Murphy’s authority to support his view is a particularly problematic for Jokic because, although Murphy concludes that there are gradations in moral worth, he does so only because he finds evidence that Kant identifies two types of moral worth, and one of these types is better than the other. Jokic wants to say that there are in fact three types of moral worth but the only evidence he
second place, the claim that supererogatory acts do have moral value conflicts with Jokic’s own argument that, for Kant, moral worth attaches only to acts that are duties, and Jokic does not attempt to explain how that argument is defective or even misleading. On the contrary, in the opening pages of his paper, Jokic says that the way for a Kantian to recognize of supererogation is to abandon the traditional moral-value requirement for supererogation and to find some other significant value in supererogatory acts.

The key to understanding how Jokic can say both that moral worth attaches only to acts performed from duty so that at best supererogatory actions have some other value and that supererogatory actions have moral worth may be found in what Jokic says about the value of acts from direct inclination. Even though these acts are not from duty, Jokic says that they have a type of moral worth. Oddly, however, part of Jokic’s support for this point is a passage in which Kant makes clear that such acts are praiseworthy, but not morally so. Moreover, Jokic’s self-devised label for the worth of acts from direct inclination is “so to speak, nearly moral praiseworthiness moral worth” (emphasis mine).¹⁹² Such a label indicates that Jokic is speaking loosely when he says that acts from direct inclination have “moral worth” and that what is actually true is that they have some other value that is close to moral value. If we likewise take Jokic to be speaking loosely in all of the passages in which he attributes moral worth to supererogatory

¹⁹² Jokic, 230. Those who see this label may begin to wonder if the apparent inconsistency on the topic of which actions have moral worth is to be solved by drawing a distinction between moral worth and moral praiseworthiness. This suggestion is of no aid in resolving the apparent conflict in Jokic’s paper, for, far from drawing a distinction between “moral worth” and “moral praiseworthiness,” Jokic uses the two terms interchangeably.
actions, we can reconcile these passages with the ones in which Jokic states that the opposite is true: strictly speaking, for Jokic’s Kantian view, supererogatory acts do not have moral worth, but they have a value that sits in close proximity to that of morally meritorious actions.

While Jokic gives the Kantian a way to value purportedly supererogatory acts, Jokic does not provide an explicit, specifically Kantian account of supererogation: nowhere does he lay out a set of necessary and sufficient conditions for supererogatory acts. Rather, Jokic’s methodology is to assume that acts traditionally recognized as supererogatory are supererogatory and then to ask how a Kantian can see them as valuable (since a Kantian cannot see them as having moral worth). We can nonetheless work backwards from what Jokic says about moral luck and the value of supererogatory acts to develop a Jokic-inspired Kantian account of supererogation. Our account will say that a supererogatory act is one that is neither required by nor forbidden by the CI and that, given its motive, involves very little moral luck. To be more specific with respect to the amount of moral luck involved: whenever the motive of the act in question is present and issues in an action, the probability, given the motive, that the issued action violates the CI is extremely low—almost zero and much lower than even the probability that the motive of love (or some other motive of direct inclination) prompts an act that violates the CI.

In light of the previous discussion of Hill’s theory, I should make one final note of clarification, this one about a way in which Jokic’s understanding of Kantian duty departs from Hill’s understanding of the same. Hill’s understanding of duty, like his account of supererogation, is intended to be firmly rooted in the text of the Metaphysics of Morals,
in which Kant makes clear that imperfect duties are duties to adopt particular principles rather than duties to perform particular external acts. Conflicting with this understanding of imperfect duty is a picture of Kantianism on which all duties—perfect and imperfect—that fall out of the CI are not duties to adopt principles but duties either to perform or to refrain from performing certain types of external actions (e.g., giving to charity or lying). Jokic’s picture of supererogation commits him to this second view. While I think that there are reasons apart from Kant’s own testimony in the *Metaphysics* for thinking that this understanding of the implications of the CI is problematic, I will not here detail those concerns. For now, it will be enough if the reader understands that when Jokic speaks of duties, he is referring to requirements that we perform (or refrain from performing) external actions such as giving to charity, cultivating our talents, and (in the negative case) lying.\textsuperscript{193}

In evaluating the Jokic-inspired account of supererogation, we can say first, in its favor, that it can do what Hill’s account could not do: it can account for the positive (and hence non-neutral) value of supererogatory acts. On this point, Jokic can say that supererogatory acts involve little moral luck, and we have indication that Kant would

\textsuperscript{193} To see why Jokic is committed to this understanding of Kantian duty, consider the particular Kantian duty of beneficence, construed as a duty to adopt the principle of beneficence, and the connection between fulfilling this duty and the possible motives one might have when fulfilling it. Mere inclination—even direct inclination—cannot be a motive for adopting the principle of beneficence. Indeed, although direct inclination can be a motive for performing external acts that benefit others, if direct inclination is one’s only motive for benefitting another, then it follows that one has \textit{not} adopted the principle of beneficence. And if direct inclination cannot be a motive for adopting the principle of beneficence, then it follows that “adopting a principle of beneficence from direct inclination” is not one possible way of fulfilling a duty that adopting the same principle from a superior motive would surpass. Hence, the duty must not be to adopt the principle but to do something else, something that can be motivated by direct inclination—something like performing an (external) action that benefits another. What holds for adopting the principle of beneficence presumably holds for the adoption of any other principle that the CI might be seen as requiring. So, if Jokic’s picture of surpassing duty is correct, fulfilling Kantian duty does not require adopting principles but only the performance or non-performance of certain types of external acts.
take this feature to be a point in their favor. And while Jokic never argues that at least some acts that have this sort of value are not morally required, very plausibly, this is the case. The CI, after all, is typically thought to leave open a wide range of acceptable acts in most situations, and some of these options likely can spring from the type of motive that Jokic thinks is almost as good as the motive of duty.

The fact that supererogatory acts lack distinctively moral value for Jokic may appear to be a drawback of his Kantian view of supererogation, but if Jokic is correct that the value he has found for supererogatory acts is very similar to moral value, the drawback will be a fairly minor one. Crucially, this drawback is what allows Jokic to avoid the problem of Kantian supererogation needing to consist in “super-rationality;” the super-rationality problem only arises if we insist that supererogatory actions are morally above and beyond the call of moral duty.

A more pressing problem for Jokic is whether he is correct when he says that the value that he has found in supererogatory acts is plausibly quite similar to moral value, as he insists that it must be for the Kantian accommodation of supererogation to work (recall that merely finding some score on which the so-called supererogatory acts are valuable—the consequentialist score, for instance—is not enough). On the story we have considered, the value of supererogatory acts is similar to the value of morally praiseworthy acts because the amounts of moral luck involved in both types of acts are sufficiently similarly small. It is not difficult to find ways to criticize this line of argument. To begin, although we have evidence that Kant thought that the total absence of moral luck signifies moral worth, it does not follow that very small amounts of moral luck signify something similar. Indeed, the difference between zero and any positive
number, however small, is, in some contexts, of great significance, and this seems to be the case in the context of Kant’s discussion of chance and moral merit. Thus, what impresses Kant about acts deserving of moral praise is not that they require very little moral luck (specifically, none), but that moral luck is completely absent in them, a characteristic that Jokic’s supererogatory acts do not even “almost” have. Even if the amount of moral luck (and not simply its absolute presence or absence) is what is significant, Jokic’s ideas about the amount of luck operative in apparently supererogatory acts are unsubstantiated. Recall that Jokic says that supererogatory acts involve “barely more” moral luck than morally praiseworthy acts and “infinitely less” luck than acts from indirect inclination or self-love. The “barely more” claim is only true if we are thinking of absolute differences in amounts. The “infinitely less” claim is certainly an exaggeration: even if acts from indirect inclination or self-love that accord with duty always involve a good measure of luck, they do not involve “infinite” luck (if such a concept is even meaningful), for the probability of such motives issuing in licit acts is, however low, greater than zero. So, acts springing from these motives cannot involve

194 Thus, the difference between a God with one fault and a God with no faults is much more significant, in religious contexts, than the difference between a God with one fault and a God with five faults. And the fact that any God that has either none or only two faults can be said to have a very small number of faults does not meant that a two-faulted God has an attribute that is very similar to the attribute of “perfection” that a faultless God has.

195 This discussion bears affinity to Parfit’s discussion of differences that can hold between a property’s “nature” and its “logic” in “Later Selves and Moral Principles,” in Twentieth Century Ethical Theory, Steven M. Cahn and Joram Haber, eds., (Upper Saddle River, NJ: Prentice Hall, 1995), 476-96. For Parfit, a property—such as the property of “being a relative of x”—can admit of degrees in its nature (since one person can be more of a relative of given individual than another person is) but not in its logic (since the question of whether or not a person is a relative of another person can be correctly answered by a simple “yes” or “no”), 477.
How do the different licit act-types vary with regard to mortal luck? To find an answer to this question, we would have to identify the motives at work in various licit actions and compare their respective likelihoods of issuing in licit actions (i.e., actions that do not violate duty). While it is quite plausible that some particularly objectionable motives are much less likely to prompt licit acts than are more admirable motives, among the more admirable motives, we simply do not have sufficient evidence to say that direct inclination motives are much more likely to prompt illicit acts than are supererogatory motives (i.e., the motives most commonly present in so-called supererogatory acts). And without evidence of a great difference, we lack sufficient evidence that supererogatory acts have special value that is much more akin to the value of morally praiseworthy acts than it is to the value of acts from inclination, particularly direct inclination. If such a difference does not exist, then supererogatory acts cannot be said to stand over and above licit acts from direct inclination.

Despite the challenges of justifying the claim that the motive attached to a supererogatory act gives that act special value, the fact that the Jokic account of supererogation pays close attention to the value of motives has its merits. Both Kantianism and our moral intuitions seem to tell us that, if supererogation is possible, motive matters: neither Kant nor the average person wants a view on which, say, giving

196 Recall that on Jokic’s picture all supererogatory acts involve some moral luck.

197 Indeed, the difficulty of identifying the motives that prompt specific actions may make this sort of evidence nearly impossible to acquire. Even if we were able to identify individuals’ motives, making a non-arbitrary determination of what factors are relevant when we are measuring the “probability” that a given motive will issue in a licit act is bound to be difficult.
large amounts to charity is good and supererogatory no matter what motive prompts one to do so. The focus on motive in Jokic’s account thus ensures that motive does matter for supererogation.

Yet some counterintuitive consequences of the Jokic account of supererogation begin to emerge when we start reflecting on just how much motive matters in the account. Let us use the label, “low luck motive,” to refer to any motive that requires some, but very little, luck—of the low quantities Jokic envisions in supererogatory acts—to issue in a licit action. Presumably, if these low luck motives can, as Jokic assumes, sometimes prompt licit acts that are not required and do not fulfill duties, they can also sometimes prompt licit acts that are required or do fulfill duties. Because Jokic determines the worth of a licit action solely on the basis of the value of its motive, three interesting corollaries of Jokic’s schema become apparent. First, no supererogatory action, however extraordinary, that springs from a low luck motive (as all supererogatory actions do) is superior to an obligatory action that springs from the same motive. An instance of tremendous self-sacrifice prompted by love of humanity is thus no better than simple honest dealings with others prompted by the same love. Second, differences in the worth of any two supererogatory acts will always be quite small on an absolute scale. For whether one supererogatory act is superior to any other supererogatory act is simply a matter of the differences in amounts of luck operative in them.198 Since all supererogatory acts involve very little luck, even if one supererogatory act may truly be said to involve much more luck (say, twice as much), comparatively speaking, than

198 At least if we are talking about moral (or very close to moral) superiority.
another supererogatory act and hence be not nearly as good, *comparatively speaking*, it will still be nearly as good in absolute terms. Hence, in absolute terms, giving one's life for the sake of another is at best only slightly superior to giving a great deal to charity for the sake of the beneficiaries. Third, any obligatory action that is motivated by duty is *superior* to any supererogatory action, no matter how extraordinary. Giving a small amount to charity *from duty* is thus always better than giving a very large amount out of any other motive.

The third corollary should alert us to the fact that it is misleading to say that the Jokic account of supererogation is in line with the common definition of supererogation as “going above and beyond the call of moral duty,” even if we assume that surpassing moral duty does not require morally surpassing (i.e., surpassing in terms of moral value) what is morally obligatory. Asking whether Jokic’s supererogatory acts are aptly described as being above and beyond the call of moral duty requires evaluating, in the situations in which they occur, the value of both the acts he classifies as supererogatory and their licit alternatives. In evaluating the value of any act, we should either consider motive or we should not. If we do not consider motive, then Jokic has not given us reason to think that in any situation any act that his theory classifies as “supererogatory” is superior to *any* licit alternative act, in which case supererogatory acts are not beyond duty. The verdict appears more favorable to Jokic if we do consider motive in determining the value of particular acts: for every particular act that Jokic identifies as supererogatory, there will be some other alternative act that is licit and that, because it is prompted by a less impressive (higher luck) motive, is inferior to the supererogatory act. But if the presence of lesser licit alternatives renders Jokic’s supererogatory acts “above
and beyond” the call of moral duty, then it follows that some surprising non-supererogatory acts are likewise “above and beyond” the call of moral duty. To wit: all obligatory acts, no matter how mundane, that are motivated by duty will be beyond duty (since they are superior to a licit alternative, namely, performing the obligatory act from a lesser motive) as will all licit acts, no matter how mundane, that happen to be motivated by a low-luck motive but that the agent could have performed from a somewhat higher luck motive (such as direct inclination). Even most (but not all) licit acts motivated by high luck motives can be said to be superior to some licit alternatives—namely, those licit alternatives that could have been motivated by higher risk motives. But if Jokic can say that supererogatory acts are “above and beyond the call of moral duty” only if he also says that the most mundane instances of telling the truth from duty or brushing one’s teeth from duty are likewise “above and beyond the call of moral duty,” we should question whether Jokie’s “supererogatory” acts are truly above and beyond the call of moral duty and appropriately labeled “supererogatory.”

The prospects, for the Jokic account, of avoiding the above counterintuitive results are not promising. The reason that Jokic’s schema renders instances of duty-promoted teeth-brushing “beyond duty,” is three-fold: the locus of an act’s value in Jokic’s account is motive; the locus of obligation is, by contrast, external act; and motives are accessed as being better or worse based on their similarity to the motive of duty. But these are the same three features of Jokic’s schema that allow (many) acts that have traditionally been thought to be supererogatory to be “beyond duty”/ have a sufficiently

199 In this last case alone, Jokic could say that the superiority in question is not sufficiently great for the superior acts to merit the description “above and beyond the call of moral duty” as opposed to the description “somewhat above the call of moral duty.”
special sort of value without being obligatory. If we deny either that motive is the locus of an act’s value or that similarity to the duty motive determines a motive’s worth, then Jokic has not found a way for (so-called) supererogatory acts to have sufficiently special value. If we deny that obligation attaches to external act, then Jokic has not given us any reason to think that supererogatory acts are beyond duty (as opposed to being illicit, by virtue of involving motives that are less than the duty-motives that the situation requires or as opposed to being simply no better than whatever is required of the agent—e.g., the adoption of or adherence to a particular principle). So, Jokic’s account’s advantage over the Hill account—the ability to ascribe value to those acts that have been traditionally been labeled “supererogatory” without being committed to their being obligatory—while real proves to be insufficient for meaningfully classifying the traditional supererogatory acts (or, indeed, any acts at all) as “above and beyond the call of moral duty.” Again, the advantage is real, but Jokic’s account is still a long way from constituting a satisfactory account of supererogation.

Note that this problem for Jokic’s account of supererogation is likely to infect other Kantian strategies for accommodating supererogation. What prevents Jokic-style supererogation from qualifying as involving going “above and beyond the call of moral duty” will apply to any Kantian accommodation strategy that attempts to make room for supererogation by finding some sort of (presumably special) non-moral value in supererogatory acts. To be true to Kant, any such strategy will have to admit that

200 For more on this point about obligation attaching to external act, see footnote 28 of this chapter.

201 Consider, for example, Richard McCarty’s “Kantian” account of the “quasi-moral” value of supererogation, which finds Kantian aesthetic value in supererogation, McCarty, 43-52.
moral value still trumps the special non-moral value that is under discussion,\textsuperscript{202} which means that no supererogatory act, no matter what special value it may have, will be as good as the act of brushing one’s teeth from duty. Since the moral value that attaches to brushing one’s teeth from duty is not enough to render doing so “above and beyond” the call of moral duty, no non-moral value attaching to a supererogatory act can do so for that supererogatory act. Our investigation of Jokic, like our investigation of Hill, thus should not leave us with high hopes for the prospect of a satisfactory Kantian account of supererogation.

\textsuperscript{202} For a (much) stronger statement of this point, see Jens Timmermann, 14.
CHAPTER FOUR

“VIRTUE ETHICS,” OPTIONAL VIRTUE, AND SUPEREROGATION

The term “virtue ethics” is an ambiguous one. It is commonly taken to refer to a major school of thought within moral philosophy, but when one inquires into the nature of this school, one is confronted with a wide range of views some of which have apparently little in common. To be sure, all theories that are given the label, “virtue ethic,” do involve some emphasis on virtue, and most such theories go a step further and reject deontology and consequentialism. Yet even the rejection of deontology and consequentialism is not a universal position among so-called “virtue ethicists,” leaving “the virtues are important” as apparently the only claim with which all so-called virtue-ethicists agree. Of course, with any philosophical theory, homogeneity among its adherents is not to be expected, and philosophers are gifted at carving distinctions that allow them to adopt one position while remaining uncommitted to the truth (or falsity) of a number of closely related positions. The fact, then, that virtue ethicists have diverse views with few points upon which they all agree is not, by itself, surprising. Yet the extent of diversity within the field of virtue ethics is quite striking. Indeed, if one were

\[203\] I have in mind Julia Driver and Robert Adams, among others, although here it should be noted that (i) the inclusion of Driver and Adams is by no means universal and (ii) in particular, neither Driver nor Adams considers her/his theory an instance of virtue ethics. Given Driver’s and Adams’ own views on this topic, I will be assuming that neither Driver nor Adams is a virtue ethicist in any sense of interest for our purposes.
told that Rosalind Hursthouse, Alasdair MacIntyre, Peter Geach, Michael Slote, and Martha Nussbaum are among the most prominent virtue ethicists and then asked to find the moral theory (however general and however editable) that unites them and separates them from Kantians, utilitarians, emotivists, and intuitionists, one would face quite a challenge.

Yet this challenge of identifying what virtue ethics espouses is one that we must address before we can answer the question of whether or not virtue ethics is compatible with supererogation, for certainly that compatibility depends upon what virtue ethics says and, in particular, what it says about moral obligation and ethical evaluation. The first part of this chapter therefore addresses this question, ultimately arguing that within the field of “virtue ethics,” we can distinguish two groups of virtue ethicists, one “radical,” the other “moderate.” Moderates accept the modern notion of moral obligation and look to the virtues for a way of grounding such obligation; radicals reject the notion of moral obligation entirely but leave a central place in their moral theory for the virtues. Because these two groups differ so dramatically in their positions on moral obligation, and because the notion of moral obligation is so closely related to that of supererogation, the two groups of virtue ethicists face very different obstacles to accommodating supererogation. The second (and larger) part of this chapter (i) argues that at least one of these groups—the moderates—can allow for supererogation and (ii) introduces a way for them to do so. Radical virtue ethics faces a much more daunting challenge to accommodating supererogation, and I will not here claim that it can accommodate supererogation, although I will say more (in chapter five) about what options are open to the radicals.
4.1 Defining “Virtue Ethics”

As I say above, developing a single, informative definition or account of virtue ethics is quite the challenge, but the difficulty of this challenge has not prevented moral philosophers from attempting it. The lack of anything like a univocal understanding of the term “virtue ethics,” even among moral philosophers, becomes apparent when we take a close look at some such recent attempts. Here, I do not refer to characterizations of virtue ethics that are made in passing in, say, undergraduate introductory ethics textbooks and that are not intended to be taken as definitive for the profession. I refer, rather, to characterizations of the field that appear in articles and book chapters that aim to explain the distinctive elements of virtue ethics to a fairly philosophically sophisticated audience.

4.1.1 The Priority Claim

One common characterization of virtue ethics that appears in such pieces says that the key claim that separates virtue ethics from Kantian and utilitarian (or, more broadly, deontological and consequentialist) ethical theories is what I will call the priority claim. According to the priority claim, the virtues are explanatorily prior to right action: the status of certain traits of character as virtues determines which actions are right. Right actions bear some positive relationship to the virtues or, more frequently, to the virtuous agent.

Versions of the thesis that commitment to the priority claim is the (or a) distinguishing mark of virtue ethics show up in the work of Justin Oakley, Gregory
Trianowsky, Daniel Stratman, and Gary Watson.  

Consider Oakley’s, “Varieties of Virtue Ethics,” which is an attempt to identify “the main positive claims made by virtue ethics” and to “explain which of these claims distinguish the approach from both traditional and contemporary versions of Kantianism and Utilitarianism, and also from standard forms of Consequentialism.” The first such claim that Oakley identifies is:

An action is right if and only if it is what an agent with a virtuous character would do in the circumstances.

Although Oakley does not believe that, on its own, commitment to the above claim is enough to separate ethics of virtue from other major ethical systems, Oakley does take the above claim, to be essential to virtue ethics. Moreover, even though the proposition, as it stands, merely states a necessary and sufficient condition for right action, It is clear that Oakley intends more than what he explicitly states. In particular, Oakley takes the above claim to imply the primacy of character in the determination of the rightness of action: an act is right if and only if and because it is what a virtuous person would do if so-circumstanced. According to Oakley, then, an essential commitment of virtue ethics is that the virtues are prior to right action.

---


205 Oakley, 129.

206 Oakley, 129.

207 Oakley, 129-30.
In a similar vein, the first major claim that Gregory Velazco y Trianowsky’s “What is Virtue Ethics All About?” makes is that ethics of virtue differ from “ethics of duty” in that:

an ethics of virtue in its pure form holds that only judgments about virtue are basic in morality, and that the rightness of actions is always somehow derivative from the virtuousness of traits.\(^{208}\)

In elaborating on this point, Trianowsky says that, for the virtue ethicist, at least some judgments about virtue are prior to judgments about rightness of act, and (crucially) what makes an action right is the goodness of certain associated traits.\(^{209}\)

As it stands, the priority claim does not provide sufficient content to delineate a worked-out ethical theory. It states that act appraisal is a function of character (or character trait) appraisal, but not vice-versa. The next question that arises for a theory that endorses the priority claim is: “how is character to be appraised, i.e., what determines which character traits count as virtues and which ones count as vices?” Here, the priority claim literature (i.e., the literature that sees commitment to the priority claim as an essential and defining feature of virtue ethics) responds that different versions of virtue ethics give different answers to this question. The most common of these answers reference flourishing or well-being: virtues are virtues because they contribute to and indeed are at least partially constitutive of human well-being; vices are vices because they detract from human well-being.\(^{210}\) Alternative accounts are available, however, and

\(^{208}\) Trianowsky, 43.

\(^{209}\) Trianowsky, 43.

\(^{210}\) Contrary to initial appearances, this view is not a form of consequentialism (specifically, one that sees human well-being as the consequence that is to be promoted) for two reasons: (i) the view does not see the virtues as mere instruments for achieving human well-being for the virtues are at least partially
these alternative accounts can differ from one another and from the well-being view in dramatic ways: for Slote, the virtues are virtues because they are essentially admirable, making virtue dependent on the views of outsiders rather than the good of the individual or human kind; a neo-Humean ethic of virtue might emphasize the way in which virtues are either useful or pleasant; and the focus of Swanton’s Nietzsche-inspired virtue ethics is “excellences,” understood neither in terms of human-wellbeing nor in terms of what humans admire.

Given the wide diversity in moral theory that adherence to the priority claim thus allows, to characterize virtue ethics merely as an approach to ethics that endorses the priority claim is not to give a terribly robust picture of virtue ethics. (To be sure, it may not be possible to offer such a detailed picture of “virtue ethics”). But there is another danger associated with characterizing all ethics of virtue as endorsing the priority claim. We can see this additional danger by attending to the guarded way in which some philosophers have appealed to versions of the priority claim when explaining virtue ethics. In the introduction of his Virtue Ethics: A Critical Reader, Daniel Stratman first says that “the meaning of virtue ethics” has finally “become more or less fixed. It now refers to a rather new (or renewed) approach to ethics, according to which the basic judgments in ethics are judgments about character.” Stratman goes on to explain that this commitment entails what I have labeled “the priority claim,” that is, that character

\[\textit{constitutive}\] of human well-being and (ii) there is no commitment either to the claim that human well-being is exhaustive of good consequences or that good consequences (however they are defined) are always to be promoted.

\[211\] Statman, 7.
assessments are explanatorily prior to the rightness of actions. Stratman also follows the scripted course in adding that most virtue ethicists say that what justifies or makes a trait a virtue is that trait’s contribution to (in a constitutive way) flourishing.

So far, Stratman’s explanation of virtue ethics is fairly pedestrian. Yet further into Stratman’s discussion, we find that the ways in which virtue ethicists endorse the claim that the virtues are prior to the right vary widely: at one end of the spectrum are those who think that, while the moral valence of traits plays a role in determining the moral valence of acts, the valence of traits is not the only factor that plays such role; consequences, for instance, may also be relevant to rightness of action. At the other end of the spectrum are those who, following Elizabeth Anscombe, want entirely to replace talk of moral obligation with talk of the virtues. On the Anscombe view, actions can be cowardly, wise, just, benevolent, greedy, and so forth, but they cannot be morally obligatory. For virtue ethicists of this latter sort, then, either (i) the priority claim should not be interpreted as implying that any acts actually are morally right or wrong, (ii) an act’s being right does not entail that it is morally obligatory, or (iii) the priority claim is false. Given these options, and given the number of virtue ethicists who are firmly in the Anscombe camp on this point, the statement that all virtue ethicists endorse the priority claim is either false or very misleading.

---

212 Statman, 7.
213 Statman, 8.
214 Statman, 9.
215 Gary Watson’s “On the Primacy of Character” makes moves similar to Statman’s. Thus, Watson at first says that in using the phrase, “ethics of virtue,” he is referring to “a set of abstract theses about how certain concepts are best fitted together for the purposes of understanding morality,” 57. Specifically, “[o]n an ethics of virtue, how it is best or right or proper to conduct oneself is explained in
To put the point somewhat differently, philosophers frequently stress the major component of virtue ethics as being a commitment to the thesis that the good is prior to the right. On this front, virtue ethics is seen as having something important in common with consequentialism. It is seen as nonetheless differing from consequentialism in that (i) virtue ethics considers virtue in some way constitutive of goodness and/or (ii) virtue ethics is not committed to the claim that the promotion of good things or states of affairs wholly determines the right. But this way of characterizing the (or an) essential difference between virtue ethics and its rivals ignores an important commitment of a significant subset of virtue ethicists: the belief that “the right” or, at least, “the morally right” is an empty category, that moral obligation is a fiction. Thus, one of the three major theses of Anscombe’s “Modern Moral Philosophy,” a paper that is often seen as both an instance of and a catalyst for contemporary virtue ethics, is that the “moral” sense of things ought to be jettisoned. Anscombe is referring to the sense in which some actions are thought to be “morally” obligatory and others not so, and she believes that this sense ought to be jettisoned because it is intelligible only within a framework (namely, a divine law conception of ethics) that western society has long since abandoned. Whereas Urmson’s 1958 article says that modern moral philosophy focuses too much on the morally obligatory and needs to branch out and pay attention to saintly and heroic action terms of how it is best for a human being to be,” 58. Note that this statement, which makes no mention of the virtues or character assessment, is not equivalent to the priority claim that we have been discussing. Nonetheless, Watson knows that one common way of cashing out “how it is best for a human being to be” is in terms of the virtues.

Watson also recognizes, with Statman, that there are replacement virtue ethics views that want to replace concepts of moral right and wrong with virtue concepts (e.g., Anscombe). Watson sees this as just an extreme way of endorsing his “claim of explanatory priority,” 58-9.
as well, Anscombe’s article from the same year sees a similar but much deeper problem, and her solution is accordingly much more radical. For Anscombe, modern moral philosophy focuses too much on moral obligation because it accepts moral obligation at all. To Anscombe, moral obligation is a hopelessly confused concept. The concept should be abandoned entirely, and we ought instead make use of thick ethical concepts, including concepts of virtues, and develop a more robust philosophical psychology. Presumably, the virtues would be grounded in this philosophical psychology.216

Anscombe’s view of moral obligation as a confused concept is not one that gained a large following initially, yet it did have a receptive audience among a significant portion of those who are now considered the most prominent virtue ethicists. Consider, for instance, Anscombe’s own student, Phillippa Foot. While Foot does not claim outright that moral obligation is a fiction, she does say that the imperatives of morality have no better claim to being categorical than do rules of etiquette. This is not to say that moral rules or the “ought” of morality have no force, but their force is hypothetical, not

216 I do not mean to imply that Anscombe’s view is that we can or ought to articulate the philosophical psychology independent of all ethical commitments and then either (i) discover (happily) that the generally recognized virtues are grounded in this psychology or (ii) develop a list of virtues grounded in this psychology, being little surprised if few of the generally recognized virtues show up on the list. It is consistent with what Anscombe says that we cannot engage in the appropriate philosophical psychology independent of all ethical commitments. Thus, Anscombe’s prescribed project may turn out to be similar to Plato and Aristotle’s project as Bernard Williams characterizes it:

“[The outlook of Plato and Aristotle] is formally egoistic, in the sense that they suppose that they have to show to each person that he has good reason to live ethically; and that reason has to appeal to that person in terms of something about himself, how and what he will be if he is a person with that sort of character. But their outlook is not egoistic in the sense that they try to show that the ethical life serves some set of individual satisfactions which is well defined before ethical considerations appear. Their aim is not, given an account of the self and its satisfactions, to show how the ethical life (luckily) fits them. It is to give an account of the self into which that life fits.”

categorical. Thus, if Foot can be said to accept moral obligation, her notion of moral obligation is atypical. Given that the modern notion of moral obligation is categorical, it follows that for Foot, like Anscombe, the modern moral ought is confused.

MacIntyre, too, is firmly in the Anscombe/Foot camp when it comes to the status of the moral ought. Thus, MacIntyre explains the purpose of his paper, "Some More About Ought," as follows:

The Prichardian or distinctively moral 'ought' was a ghost and it is a ghost that still walks in certain quarters, although more and more obviously, like other ghosts, a survival. Yet so long as it survives, morality involves a degree of bluff and deception that can only have the effect of engendering cynicism wherever it is once more expressed. This paper is therefore not only an attempt at analysis; it is also hopefully an exorcism.  

For MacIntyre, like Anscombe, the distinctively moral ought is deceptive, and should be exposed as such and abandoned. Also like Anscombe, MacIntyre believes that the reason the concept is deceptive is that it presupposes a framework that has been abandoned. And like Foot, his views imply that morality, at least in the form available to us today, cannot avail itself of categorical imperatives but only of hypothetical ones.

---


218 In Against the Self-Images of the Age: Essays on Ideology and Philosophy, (Notre Dame, IN: University of Notre Dame Press, 1989),171-2, emphasis mine

219 See also After Virtue: A Study in Moral Theory, second edition, (Notre Dame, IN: University of Notre Dame Press, 1984), esp. chapter five, in which MacIntyre's discussion of the meaning of evaluative claims implies that a statement about moral obligation is either nonsense, a statement about a hypothetical obligation, or a statement about divine commands. Note that the point is not that moral obligation, for MacIntyre, must thus be rooted in divine commands. The point is stronger. What MacIntyre says about evaluative statements implies that "you ought to do y," if intelligible, either means "if you want x, you ought to do y" or "God commands y." It follows that it is nonsensical to say that there is an absolute moral obligation to obey God.
4.1.2 Two Types of Virtue Ethics

Despite the fact that some prominent virtue ethicists reject the modern notion of (categorical) moral obligation and hence cannot endorse any theory that sees the virtues as the source of such obligation, it cannot be ignored that the group of virtue ethicists whose view the priority claim was developed to capture does exist, that is, a number of virtue ethicists disagree with Anscombe and MacIntyre and accept both modern moral obligation and its source in the virtues. Bearing in mind this split of views, we can divide virtue ethicists into two groups: (i) “moderate” virtue ethicists, who accept moral obligation (in its modern form) and, in accordance with the priority claim, see such obligation as somehow rooted in virtue and (ii) “radical” virtue ethicists, who reject moral obligation (in its modern form) entirely but still can be said to make virtue fundamental to their ethical theory. 220

__________________

220 I am not the first person to note a distinction between radical and moderate virtue ethics. I take it that many have noted or assumed it in conversation, some have gestured at it in print, and W. David Solomon has explicitly carved up the field of virtue ethics in a similar, albeit non-identical, way in “Virtue Ethics: Radical or Routine?” in Intellectual Virtue: Perspectives from Ethics and Epistemology, Michael DePaul and Linda Zabzebski, eds., (New York: Oxford University Press, 2003), 57-81. Solomon’s own distinction is in service of a larger point that there are two major schools of thought that operate under the label, “virtue ethics.” Virtue ethics that he calls “radical” are radical in their response to modern moral thought generally; by contrast, the virtue ethics he classifies as “routine” are much more conventional, 59, 67-9. Solomon lists nine substantial commitments that his radical virtue ethics tend to share, only one of which is suspicion of the moral (although another is suspicion of rules or principles), 68-9. My radical/moderate distinction, in focusing exclusively on moral obligation as the point of separation, is stricter than Solomon’s distinction because my purpose differs from Solomon’s. I am trying to separate ethics of virtue based on factors that present challenges for a virtue ethics account of supererogation. Since not all ethics of virtue face the same obstacles to supererogation, I want to be able to separate ethics of virtue, by and large, by the obstacles that they face; as it happens, the most significant such obstacles are related to conceptions and accounts of moral obligation.

What Solomon’s classification system teaches us, though, is that even if I separate ethics of virtue based solely on the single factor of most relevance to the question of whether supererogation is possible (i.e., the question of whether and how moral obligation is possible), the actual theories (and theorists) that fall under the “radical” side of my classification system, as a matter of fact, have much more in common with one another than a simple rejection of the moral ought. For (almost) every virtue ethicist who rejects the moral ought shares a strong family resemblance, captured in Solomon’s nine point classification schema, with the other radical virtue ethicists who reject the moral ought. Theories on the “moderate” side
This way of carving up the field of virtue ethics, i.e., according to members’ stances on the coherence of the modern concept of moral obligation, reflects a genuine and substantial difference that we see within the field of virtue ethics. The difference is genuine in that it simply is the case that some virtue ethicists admit of moral obligation while others do not. The difference can is substantial in that, given the central role that the notion of moral obligation has played in modern moral thought, any rejection of this concept ought to be seen as significant.

Even though the classification system for ethics of virtue that I am suggesting is genuine and substantive, I do not assume that it is the only or ‘best’ taxonomy for virtue ethics. The reason that I carve up the field as I do is not for the mere sake of classifying forms of virtue ethics but rather because the particular dividing line I am employing is connected to the topic that concerns us, that is the question of whether or not an ethics of virtue can accommodate supererogation. For the challenges associated with such an accommodation will of course depend upon whether or not the virtue ethic in question accepts the possibility of moral obligation. For moderate virtue ethics, the issue will be that of explaining how a theory that grounds both obligation and moral goodness in the virtues, which are presumably ideals, can find a way for surpassing obligation. For of my classification system, on the other hand, rarely share the family resemblances to the radicals that Solomon identifies, although it should be admitted that theories on each side of my classification schema can be more or less radical/moderate. The reason for the family resemblance among (my) radicals is that those who are radical tend, as a matter of fact, to be radical not merely in rejecting moral obligation but also (as Solomon shows us) in rejecting much of the modern ethical problematic. And they tend to reject that problematic for similar reasons and in similar ways. The moderates, by contrast, are united in accepting not only the modern concept of obligation but also modern approaches to ethics that see an articulation of the application standards of this concept as the central ethical question. Qua virtue ethicists, the moderates answer this question by appeal to virtue. The major differences among the moderate views, then, tend to be limited to the particular ways in which virtue can be said to ground obligation and, more substantively, the grounds of virtue itself.
radical virtue ethics, the challenge will be the much more daunting one of explaining how anything like the surpassing of moral duty is possible given that there is no such thing as moral duty. The question of whether virtue ethics can accommodate supererogation can then be seen as the question of whether either of these groups, with their different obstacles, can accommodate supererogation. The remainder of this chapter argues for an affirmative answer in this respect: moderate virtue ethicists, at least, have access to a plausible account of supererogation. The much more difficult question of what radicals can and should say about supererogation is one I will address briefly in the final chapter.

To be clear, I am not merely interested in arguing that, in theory, virtue ethics, at least in its moderate form, can accommodate supererogation. Rather, I am interested in whether actual virtue ethicists can—without giving up some of their most deeply-held commitments—accommodate supererogation. For this reason, as I argue for a particular moderate virtue ethics account of supererogation I will be concerned with what actual moderates tend to say about goodness, obligation, and morality, the reasons they offer in support of their positions, and what relevance these commitments have to the possibility of supererogation. Since we cannot expect homogeneity among even all moderates on all commitments relevant to the question of whether supererogation is possible, I will not attempt to find a single account of supererogation that will satisfy all actual moderate virtue ethicists. My goal is less ambitious than a universal moderate account and more ambitious than an account that a moderate could theoretically endorse. My goal is to find an account of supererogation that a large, significant subset of actual moderates can endorse without giving up their foundational moral commitments. If I do this, I will have, in turn, provided an account of supererogation that a significant subset of actual
virtue ethicists can endorse, which, recall, is the primary goal of this dissertation. To the goal of finding an account of supererogation for moderates, I now turn.

4.2 Supererogation for the Moderate Virtue Ethicist?

As I hinted above, the moderate virtue ethicist who wants to make room for supererogation does not face the daunting task that the radical virtue ethicist does, that is, the task of explaining how one can surpass moral duty in a non-trivial way when moral duty is a fiction. Nonetheless, the moderate faces her own obstacle. We can appreciate this obstacle by turning to the sparse literature on the question of whether virtue ethics can accommodate supererogation, for this literature is nearly exclusively concerned with the question of whether the moderate virtue ethicist can accommodate supererogation, and the conclusion proffered is typically negative. Moderates are seen as unable to accommodate supererogation due to the way in which they cash out moral obligation.

221 The only possible exception of whom I am aware is David Heyd, who may be thinking of radical virtue ethicists when he remarks:

Now a completely non-deontological theory—that is to say, a theory based on a concept other than duty (virtue, sympathy, etc.)—cannot accommodate supererogation because if there is no duty then \textit{a fortiori} there cannot be action which transcends duty.


Slote is also something of an exception: at one point in \textit{From Morality to Virtue}, (New York: Oxford University Press, 1992), he indicates that \textit{rational} supererogation is compatible with virtue ethics and indeed implied by the form of virtue ethics that he develops. Although he does not further develop an account of supererogation—rational or otherwise—for his form of virtue ethics, he seems to assume that the bare possibility of degrees of virtue makes possible supererogation (in the case of high degrees of virtue), 128-9.
Their problem is comparable to utilitarianism’s maximizing problem, and we see it articulated in the following comment from Julia Annas:

In an ethics of virtue, there is no obvious room for supererogation. There is no ‘floor’ of minimal moral obligation for the agent to rise above; being a fully virtuous agent is an ideal for everyone. The development of virtue is a process that everyone starts and continues to go along; there are no levels that only moral heroes are supposed to reach.223

In other words, on a moderate ethic of virtue, all people are obligated to act in accordance with full virtue; full virtue is the sole and highest moral ideal; so, it is not possible morally to surpass what one is obligated to do; supererogation is accordingly an empty category.224

4.2.1 Jason Kawall’s Ideal Observer Account

I know of one explicit attempt to develop an account of supererogation for the type of virtue ethics that I am calling moderate, and that is Jason Kawall’s virtuous ideal observer account. Kawall avoids the above problem by altering the virtue ethicist’s standards for obligation. Instead of identifying moral obligation with what the virtuous agent would do (as is usual for moderates), Kawall identifies moral obligation with what


224 See Michael Byron, “Could Aristotle Satisfice?” in *Satisficing and Maximizing: Moral Theorists on Practical Reason*, Michael Byron, ed., (New York: Cambridge University Press, 2004), 190-212. Byron does not actually say that virtue ethics cannot accommodate supererogation. What he offers is, rather, an argument against satisficing—doing only what is minimally required—as a part of the virtuous life. Byron’s argument is that if it turns out (for whatever reason) that satisficing is a virtue satisficing is accordingly simply obligatory as it is part of the perfect life. Byron’s reasoning thus presupposes that the virtuous life is one of perfection. If the virtuous life is obligatory, we thus have no opportunities for supererogation, 206-8.

See also Nancy Sherman, “Common Sense and Uncommon Virtue,” *Midwest Studies in Philosophy*, 13 (1988): 97-114. Sherman sees the Aristotelian account as not admitting of supererogation. On Sherman’s picture of Aristotelian ethics, virtues can be had to various degrees; right action is defined in terms of virtue; so, as long as one does not have a virtue to the highest degree, one cannot be said to have done more than what ought be done, 97-105, esp 101 and 105.
virtuous agents would *judge* to be obligatory—or, more correctly, what all virtuous agents would judge to be obligatory were they to have all relevant information.

Specifically, for Kawall:

> An action is morally *required* for an agent in a given set of circumstances iff all fully-informed, unimpaired, virtuous observers would deem the action to be so.\(^{225}\)

From this understanding of moral obligation, it is a short step to use the same base—the hypothetical judgments of ideally situated virtuous agents—to accommodate supererogation. Kawall’s account of supererogatory action thus says:

> An action is *supererogatory* for an agent in a given set of circumstances to the extent that fully-informed, unimpaired, virtuous observers would deem the action to be supererogatory.\(^{226}\)

Note that an important shift occurs from the role of ideal virtuous observers in Kawall’s account of moral obligation to their role in his account of supererogation. Universal agreement among these idealized observers is necessary for the action in question to be obligatory; if one such virtuous observer would judge differently, the action is not obligatory. For supererogation, the opposite is true: one idealized observer’s judgment that the action is supererogatory is sufficient to make the action supererogatory. This leads to a threshold account of supererogation: an action is supererogatory to the extent that idealized virtuous observers would agree that it is. In Kawall’s clarificatory remarks, we later learn that one other factor shapes the extent to which an action is supererogatory, and that factor is the strength of the observers’ judgments on the supererogatory nature of

---

\(^{225}\) Kawall, 187.

\(^{226}\) Kawall, 187.
the action. We also learn that the full-information requirement entails that the idealized observers grasp the concept of supererogation, so we needn’t worry about observers failing to label an act, “supererogatory,” for that reason. Of course, there is no guarantee that all virtuous agents familiar with the concept will believe that it denotes a non-empty category, and indeed it is likely that at least some such observers will not take any actions to qualify as supererogatory. So, it is unlikely that any action, no matter how great, will qualify as maximally supererogatory, but that will not stand in the way of many an action’s qualifying as supererogatory, at least to some extent.

The problem with Kawall’s account, as an account of supererogation for the (moderate) virtue ethicist, is that the claim it makes about supererogation—that supererogation coincides with the judgments of idealized virtuous judgers—is likely false, on moderate virtue ethics. Virtue plausibly guarantees fulfillment of obligation, at least when the virtuous person has access to all relevant knowledge, but full virtue, even when combined with access to all relevant knowledge, does not plausibly guarantee error-free moral judgments about the moral status of another’s act.

One reason that full-virtue does not guarantee correct assessment of others’ acts is that while “the thing for me to do” may always be clear to the fully virtuous person under idealized epistemic conditions in any situation in which she might find herself, it does not follow that “the thing for A to do” will likewise be clear to her, when A is some other

---

227 Kawall, 187.

228 This last point is not one that Kawall discusses.

229 It may be that the knowledge qualifier is irrelevant, that what is obligatory is relative to what one knows, so that if full virtue guarantees right action in situations of full knowledge it likewise guarantees right action in situations of ignorance, although what is right (for the agent) in these two situations may differ.
person, even if our hypothetical virtuous observer knows as much about A’s circumstances as is theoretically possible. Some of the factors relevant to deciding what to do in a given situation are relative to who one is. And full knowledge of some of these factors is accessible only from the first-person perspective. Only the agent can, as it were, know her own heart, certain experiential features of her past, and the depths of certain commitments she has. Since such factors can be relevant to what the agent morally ought to do—some of them shape what fidelity, for the agent, would amount to, for instance—a virtuous ideal observer cannot be guaranteed to know what another person morally ought to do and hence when another person is surpassing what she ought to do.

A second reason that an idealized virtuous observer may fail accurately to evaluate the moral status of another’s act is that a virtuous person’s ability to make wise moral decisions is compatible with her not knowing which of her options are morally obligatory, which are forbidden, and which are supererogatory. We can see why this might be the case, at least for some virtuous agents, by considering what John McDowell says about how virtuous agents perceive their options and reasons for action. McDowell sees the virtues as certain perceptual capacities that prompt one to see a particular action

__________________

230 Anyone who doubts this fact should try to imagine how knowledge of the depths of a given agent’s commitments could be communicated to an ideal observer. An ideal observer can know that “agent A cares very much about the prevention of cruelty to animals,” and she can also know that “A likewise cares very much about her friendship with B.” The ideal observer can even know many facts about A’s history that are relevant to A’s concerns—e.g., that A worked on a factory farm one summer in college and subsequently developed depression and that A and B have known each other for twelve years. But not having actually lived A’s life, that is, not actually having held the commitments herself and experienced the events and actions that shaped these commitments over time, the observer’s perspective on A’s life will always be incomplete. She will never have full knowledge of A’s background; if such externally inaccessible facts are ever relevant to A’s obligations, our ideal observer’s judgment may accordingly be in error.
as “the thing to do” in the situation precisely for the reasons that that action is the thing to do in such circumstances. “Reasons” that might incline others not to act in the appropriate way will have no pull for the virtuous agent, and they will not be seen by him as reasons in the situation in question. Thus, if the virtuous agent is confronted with a situation in which saving a child requires putting himself at great risk, he will see taking the (risky) steps to save the child as “the thing to do;” he will see it as the thing to do precisely because it will save the child; and this reason (i.e., saving the child) will “silence” any reasons he has for not doing what will save the child, reasons associated with self-preservation, for instance. So, for the virtuous agent, it is not that the good associated with saving the child outweighs the risk to self; rather, the risk to self does not gain the attention of the virtuous agent in the face of the larger concern of saving the child.231

On McDowell’s picture, then, when a situation requires a particular action, the virtuous agent, by virtue of being virtuous, is sensitive to this requirement, but in a very particular way: not as “what is morally required” but as the only option under consideration.232 If McDowell is correct, though, we have reason to think that many a virtuous agent will approach acts of supererogation in the same way: supererogatory options will be seen as the only options worth pursuing, not as “supererogatory.” We can even imagine that, for some ideally virtuous persons, the question of where obligation leaves off and supererogation begins will be uninteresting. Such a person will be interested in doing what is good, not in keeping her nose clean, and this interest may lead

232 McDowell, 334-5.
to her, in some situations, simply not knowing where the line between a clean nose and
supererogation is to be found. And here I do not just mean “not knowing” in the sense
that the agent is not certain about the exact place to draw the line between the obligatory
and the supererogatory; I mean that she has little idea of where to draw the line. She
would no doubt recognize, if asked of a particular case in which an agent faces a range of
options, that certain base options are immoral, but for a still wide-range of the options,
she would not know whether to draw the obligation line somewhere in the low, middle,
or high-end range of the options (if she is ranking the options according to some standard
of moral goodness). If asked, she might well say “I don’t know if middle-range action X
is supererogatory, a bare fulfillment of duty, or less than what morality requires. And
frankly I’m not all that interested. If it were me, I would opt for something like (high-
end) action Y, for Y would mean so much to B (the beneficiary). I suppose I would advise
others in my shoes to do Y as well, but if such another person were to opt for X instead of
Y, I really couldn’t say what the ‘moral status’ of his action would be, whether he would
be violating, barely fulfilling, or surpassing duty.”

What the above reflection suggests is that full virtue does not guarantee anything
like perfect application of moral concepts, even if familiarity with the concepts is
presupposed. And if full virtue does not guarantee such perfect application, then
Kawall’s account of supererogation is likely mistaken; the hypothetical judgments of the
virtuous, like the hypothetical judgments of the rest of us, fail to map on to the
supererogatory in a number of cases.

Kawall could object that I am misunderstanding his accounts of obligation and
supererogation, that his point is not that virtuous agents know what is obligatory and
supererogatory (and hence judge appropriately) but rather that it is precisely their hypothetical judgments that make certain actions obligatory and others supererogatory (to varying degrees), which guarantees a perfect correspondence between their hypothetical judgments and supererogation. If such is the case, my objections so far can be seen as begging the question against Kawall insofar as they assume that the obligatory/supererogatory status of acts exists prior to the relevant hypothetical judgments. Further reinforcing this reading of Kawall (i.e., the reading on which what makes an act supererogatory is the hypothetical judgments of idealized virtuous agents) is the fact that it is what Kawall must say if his account of supererogation is to count as an account of supererogation and not just be an affirmation of some (perhaps epistemically useful) connection that holds or tends to hold between acts of supererogation and virtuous agents.

If the hypothetical judgments of idealized virtuous agents are what make actions forbidden, obligatory, and supererogatory, though, what we have is not a form of virtue ethics but a form of subjectivism, for ethics is now grounded not in virtue but in the (hypothetical) judgments of a (hypothetical) panel of evaluators. Kawall’s picture does differ from standard subjectivisms in that Kawall allows only virtuous persons on his panel, whereas typical forms of subjectivism are more democratic in whom they allow. Limiting the panel members in the way that Kawall does, though, is not enough to qualify his view as a form of virtue ethics; for that, the tie between supererogation and the virtues would need to be tighter. On the Kawall account, nothing prevents some panelists from
believing that the bar for moral obligation is very low and perhaps even non-existent. So long as one idealized virtuous panelist is like this (and very likely at least one will be), it will turn out that the bar for moral obligation is very low indeed (since obligation occurs only with the agreement of all panelists) and likely even non-existent. Accordingly, some decidedly vicious actions will likely turn out to be permissible and possibly even to a small degree supererogatory. I take it that any view on which decidedly vicious actions can be supererogatory, even to a small degree, is incompatible with moderate virtue ethics. Kawall’s account, then, ought to be rejected by any serious moderate virtue ethicist as an account of supererogation.

Kawall’s strategy thus does not enable the moderate virtue ethicist to accommodate supererogation. The moderate who wants to allow for it accordingly must address the problem we considered above, the problem of explaining how one can surpass duty when duty is grounded in the virtues, which are themselves moral ideals. To accommodate supererogation, the moderate has two options: she can, as it were, lower the bar for moral obligation, saying that people are not always obliged to act in accordance with full virtue or she can identify a way of surpassing full virtue. Of course, the panelists in question will never use the low bar as an excuse for mediocrity, so the panelists in question will never believe in a low bar and then regularly meet this bar but do no better. But that does not change the fact that their judgments about obligation embody a low bar, and it is their judgments, not their actions, that Kawall’s account uses to ground supererogation.

If only one fully virtuous person, perhaps following the radical virtue ethicist, would understand but reject the concept of moral obligation, the bar will be so low as to be non-existent. Kawall’s approach can be seen as embodying both of these approaches (albeit in ways that will be implausible to the moderate virtue ethicist): Kawall lowers the bar in that, for him, moral obligation exists only when all ideally situated virtuous observers would agree that it does; he allows for the surpassing of full virtue in that impressing hypothetical virtuous observers is a good (or so Kawall seems to think) that is not guaranteed by the possession of full virtue.
course, whichever route she takes will have to be properly motivated. It is no help to stipulate that the bar for moral obligation is appreciably lower than the bar for full virtue without explaining why such is the case, and it is likewise of no use to create a story on which there is some way to surpass full virtue unless that story has been properly motivated.

4.2.2 Better than Virtuous?

Let us consider, first, how one might argue that surpassing full virtue is possible. One difficulty of this strategy emerges when we consider virtue, following Aristotle, as a mean between vices. Virtues are, by definition, the good, the target we are to achieve. One can miss the target in a number of ways, but any of these ways manifests vice, whether it is a common enough vice to be named (e.g., “cowardice”) or a much less common (and hence unnamed, albeit genuine) vice. Overshooting the virtues, what to use Gary Watson’s language we might call, “virtue in excess,” is, in the end, vice.\textsuperscript{236} Hence, it is wrong, and the only way to avoid this wrong and all others is to hit the mean of virtue beyond which there is nothing better.\textsuperscript{237}

The above case against the possibility of surpassing virtue is by no means water-tight; it may well be that some virtues are not means between vices, and if so, this fact

\textsuperscript{236} Gary Watson, “Virtues in Excess,” \textit{Philosophical Studies} 46 (1984): 57-74, esp. 57-8. It should be noted that when he uses this phrase, Watson is not discussing the issue of supererogation as it relates to virtue ethics. Watson is instead looking at the question of whether a virtue could ever prompt wrong behavior and how the answer to this question is related to the thesis of the unity of the virtues.

\textsuperscript{237} See Heyd, 36-7.
may allow for the surpassing of virtue. Even if the above argument is not decisive, I take it that it, coupled with the already mentioned point that virtues are ideals, makes a strong enough prima facie case that the burden of proof is on the person who believes that surpassing virtue is possible to find a non-arbitrary reason for thinking that such is the case.

Aristotle’s discussion of magnanimity is a natural place to look for such a reason. Aristotle, not explicitly concerned with making room for supererogation, talks of magnanimity not as a virtue but as “an adornment of the virtues,” which suggests that it is something of value that goes beyond ordinary virtue. Specifically, Aristotelian magnanimity is a proper response to great honor on the part of the person of great worth, i.e., the virtuous person. The magnanimous person is worthy of great honors, recognizes this fact, and acts accordingly. Magnanimity is opposed not to two other inappropriate attitudes toward honor but to three: vanity (thinking one is great when one is not); temperance (recognizing one’s own lack of worth); and pusillanimity (thinking one is lower than one actually is). What makes magnanimity an adornment of the virtues is,  

---

238 Here, it is worth pointing out that objections to Aristotle’s doctrine of the mean (with respect to virtue) tend to focus on the explanatory value of the doctrine, not its truth.

239 Notably, at least one person, Nancy Sherman, has considered Aristotelian magnanimity in the context of a discussion of supererogation, and Sherman does not think that magnanimity makes supererogation a live possibility on Aristotelian ethic (which is the stated theory of concern to Sherman, whose understanding of “Aristotelian ethics” appears to make it a form of what I am calling “moderate virtue ethics”). For Sherman, like Annas, right action is defined in terms of virtue, and as long as one does not have a virtue to the highest degree, one cannot be said to have done more than what ought to be done, 97-105, esp. 101 and 105. Sherman, far from seeing Aristotelian magnanimity as a solution that makes supererogation possible, sees it as a problem in that it functions as an excuse for sustained neglect of ordinary duties, 107.


241 NE, 1123a-1123b.
first, that magnanimity, by definition, is possessed only by the virtuous person and, second, that it adds to the greatness of such a person: to be virtuous but lack magnanimity is to suffer from pusillanimity.\textsuperscript{242}

In addressing the question of whether Aristotelian magnanimity provides the moderate with a way of capturing supererogation, an obvious issue for us to consider is whether Aristotelian magnanimity ought to be valued, and if so, for what reasons. Modern sensibilities react strongly against certain aspects of Aristotelian magnanimity. Thus, Aristotle’s claim that the magnanimous man is ashamed to receive benefits because the role of beneficiary is reserved for inferior people\textsuperscript{243} strikes many as elitist and silly, seemingly implying that there is something unseemly in accepting an offered umbrella, dining at a friend’s house, or receiving a Pell grant. In response, others will no doubt complain that these examples betray an unfair reading of Aristotle, that one can be magnanimous without being a self-satisfied snob, and that Aristotelian magnanimity is worth a closer, more charitable look.

Even if we set worries about elitism aside and ignore some of Aristotle’s more radical-sounding claims about magnanimity, we still need to ask what, exactly, about magnanimity ought to be seen as of value. When we ask this question, the answer is at most that the magnanimous person has a correct understanding of self. For while knowing oneself to be virtuous (and acting accordingly) is preferable to knowing oneself to be vicious (and acting accordingly), that is only because being virtuous is preferable to being non-virtuous. And even if magnanimity is preferable to virtue without

\textsuperscript{242} NE,1123b-1124a.  
\textsuperscript{243} NE, 1124b.
magnanimity (and some, citing the value of humility, will dispute even this claim), the way in which magnanimity improves upon virtue is no more significant than the way in which self-knowledge improves upon vice.\footnote{244}

But the reason that self-knowledge improves upon a person’s state of character, whether that state be full virtue or virtue mixed with vice, is that self-knowledge is itself a virtue and self-deception a vice, a point that finds support in Socrates’ dictum “know thyself.” Our social practices of moral education likewise indicate that self-knowledge is a virtue, for these practices depend upon humans knowing where they stand \textit{vis-à-vis} one another and specific virtues and vices. Consider, for instance, the moral education that occurs within the parent-child relationship. The success of this education depends upon both children and parents recognizing, at least generally, where they stand relative to one-another \textit{vis-à-vis} virtue broadly and, in the case of the parents, where they stands \textit{vis-à-vis} particular virtues and vices. Parents will be ineffective moral educators if they do not view themselves to be more morally developed than their children, and the more knowledge they have about their own state the better position in which they will be to instill virtue in their children. Self-knowledge is thus indispensable to moral education.

If self-knowledge is a virtue, however, magnanimity, insofar as it is nothing more than self-knowledge in a case in which the possessor of this knowledge has all other virtues, cannot be said to be an adornment of the virtues in the sense that might make the

\footnote{244 One could object by saying that, in the case of magnanimity, unlike the case of temperance (roughly, self-knowledge with vice), the agent is able to receive great honor, the “prize of virtue,” and this is why self-knowledge improves upon virtue in a greater way than it improves upon vice. \textit{NE}, 1123b. I concede that the self-knowledge of magnanimity improves upon possession of the other virtues in a way relevant to honor, but temperance has an analogous way of improving upon partial vice for it saves us from the shamefulness of partaking of honors of which we are not worthy.}
magnanimous person surpass virtue, i.e., be better than (merely) virtuous. Rather, the magnanimous person simply has all virtues. To be sure, this is no easy feat, but it is not the case that the person in question has surpassed virtue; she could not have all of the virtues and fail to be magnanimous (in the Aristotelian sense). Aristotelian magnanimity, then, appears to offer the moderate no solution to the problem of supererogation. Of course, even if magnanimity is not appropriately construed as an adornment of the virtues (in the sense that it is possible to have all virtues and yet lack this additional positive feature), it may be that some other state or achievement can rightly be construed as such an adornment or in any case as some way of surpassing virtue.\(^{245}\) I know of no other plausible suggestions for what state or achievement might qualify for doing so. Until one is suggested, it behooves us to move on to explore the second way a moderate virtue ethicist might try to accommodate supererogation: lowering the bar for moral obligation so that one can satisfy duty without full virtue.

4.2.3 Lowering the Obligation Bar

The literature on the topic of supererogation and virtue ethics involves discussions that at least suggest that a moderate virtue ethicist can place the moral obligation bar below the requirements of full virtue. Christine Swanton, in particular, has stressed the need for a somewhat relaxed picture of morality.\(^{246}\) Swanton argues that virtue ethics can

\(^{245}\) Magnanimity might be said to be an adornment in the following very loose sense: magnanimity makes one properly disposed to receive those honors that rightly adorn the virtuous person. Here, though, it is the honors that, strictly speaking, are the adornment. Note also that the adornment in question is not a way of surpassing virtue but is rather a way of receiving one’s due for virtue.

\(^{246}\) Swanton discusses this thought in two pieces: chapter nine of her *Virtue Ethics: A Pluralistic View*, (New York: Oxford University Press, 2003), and her later “Satisficing and Perfectionism in Virtue
succeed in being “relatively undemanding” by employing either of two strategies. The first strategy is to emphasize the agent-relative nature of the virtue of perfectionism, so that weaker, less-developed agents needn’t be required to imitate the fully virtuous in order to have this virtue. For Swanton, their doing so would be ill-advised and frustrate their own moral development given that it would likely lead to these weak agents’ “overreaching” their abilities. The second strategy is to separate rightness from moral blame in a non-traditional way such that an agent can fail to do what is morally right without being blameworthy.

Swanton’s first strategy is of little obvious aid to the virtue ethicist who wants to make room for supererogation (and here it should be noted that, although Swanton herself is open to the possibility of supererogation, making room for supererogation is not her goal for either of the two strategies). In the first place, even if what perfectionism amounts to (and hence requires) is agent-relative, that gives us no obvious way of holding that, for any particular agent, there is a way for him to do what is better for him than what morality requires of him; morality can be relatively undemanding—because agent-

---

247 I should note that Swanton mentions three strategies, but she only endorses and defends two of them. In “Satisficing and Perfectionism in Virtue Ethics,” these defended strategies are her “second” and “third” strategies; I am calling them her “first” and “second” strategies, respectively, because I am not discussing her paper’s first strategy, the one that she does not endorse, 177-8.

With respect to the two strategies that Swanton does endorse, what is unclear from Swanton’s work is whether these two strategies are mutually exclusive or complementary. Swanton initially presents them as opposed, but later comments suggest both that the first depends on the second and that the second is justified only by appeal to the first, 177-8.


249 Indeed, Swanton specifically states, after arguing for a virtue ethic that is undemanding in certain respects, that she wants to leave open the possibility of valuing supererogatory acts. Unfortunately, she does not explain how this is to be done, Virtue Ethics: A Pluralistic View, 210-11.
relative—without it being a good idea for any particular agent to “exceed” the relaxed demand. If Swanton is right that the reason the demands are lower is that it is ill-advised for the weak to imitate the strong, then so-called supererogatory acts will be required of the strong but foolish for the week; in neither case will they surpass duty in an admirable way.

A second problem for Swanton’s first strategy is that, as Swanton herself admits, it only lowers the standards of virtuous perfectionism; the standards of the other non-agent-relative virtues remain intact, and these other virtues are arguably sufficiently demanding to make any relaxation of duty with respect to the virtue of perfectionism insufficient to relax the standards of overall duty.\textsuperscript{250} The only response that Swanton has for this problem relies on appeal to her second strategy: one can say that falling short of these difficult, non-agent relative virtues constitutes failure to do what is right without committing oneself to the view that such failure is blameworthy.\textsuperscript{251}

Swanton’s first strategy thus depends on her second strategy, but her second strategy also is of little aid in accommodating supererogation. For Swanton must do more than \textit{declare} that an act can be right without being blameworthy; she must justify this statement, and she does not do this. She attempts to justify it by returning to her point about the agent-relative nature of perfectionism, stating that we should not blame a person who does what perfectionism \textit{for him} requires. When Swanton says we should not blame such a person, she has in mind, of course, a case in which what perfectionism asks of the agent is less than what the non-relative virtues ask of him. As it stands,

\textsuperscript{250} Swanton, “Satisficing and Perfectionism,” 177-8.

\textsuperscript{251} Swanton, “Satisficing and Perfectionism,” 187.
Swanton’s statement about not blaming such an agent, while it has a measure of appeal, is false: one does not gain complete immunity from moral blame simply because one is acting in a way consistent with the agent-relative requirements of virtuous perfectionism. If one is sufficiently vicious, what perfectionism requires may be very minimal, but it does not follow that such a vicious person is not liable to any sort of moral blame. To be sure, no person ought to be blamed for striving for perfectionism, but he can justifiably be blamed for any number of other vicious ways in which he behaves while so-striving.

An example will help to clarify the above point. Suppose Sarah is a weak, uncreative, somewhat lazy, dishonest person. She frequently lies when it is convenient for her to do so, and in the case of white lies she rarely stops to consider that this sort of thing ought not to be done or even that what she is doing is, in fact, lying. Sarah is not completely amoral: she does have moral qualms about her “big” lies, and she sometimes reflects on her white lies in retrospect, recognizing them for what they are and evaluating them as wrong. Sarah also wants to improve herself morally with regard to dishonesty. Having (happily) the agent-relative virtue of perfectionism, she realizes that she cannot and ought not immediately try to imitate the fully virtuous agent on this front; she will have to take smaller steps. So, she aspires to do two things with respect to dishonesty: stop telling large lies almost entirely and start attending to her small lies, recognizing them for what they are before she utters them. She need not be so legalistic as to attempt to cut herself off from these white lies entirely, however. Indeed, at this point in her moral development, she would not even know how to give up the white lies without experiencing dramatic changes in her personal life, for she would not know how to relate to friends, family, co-workers, and acquaintances if she were to give up all lying cold-
turkey. For her, white lies are a tool she uses to relate to others, and she is so accustomed to deploying this particular tool that she does not yet know how to maintain normal relationships without it.

In the situation we have imagined, agent-relative perfectionism plausibly advises Sarah to monitor her lying—in hopes that she will, over time, come to see new ways of interacting that she is at present too uncreative, unreflective, and vicious to recognize—without recommending that she immediately cease lying altogether. Suddenly giving up the white lies (or even giving up all of the unjustified white lies) is a task that is “beyond her strength,” to use one of Swanton’s preferred expressions. Nonetheless, Sarah is still blameworthy for her dishonest behavior in the mean time. Suppose she flatters a colleague because, although there were plenty of ways she could have behaved that would have involved neither flattery nor rudeness nor coldness, she can think of none of these options in her current moral state. In such a case, her behavior and lack of moral imagination betray her vice(s), and the flattery is accordingly blameworthy, even if in accord with the agent-relative standards of virtuous perfectionism.

The example of Sarah shows us that Swanton’s appeal to the agent-relative nature of the virtue of perfectionism does not grant a person immunity from moral blame in a way that separates what is morally right (i.e., in accord with all of the virtues, including the ones that are not agent-relative) from that which one is morally blameworthy for failing to do. Because this is the only way in which Swanton supports her second strategy, that second strategy is unmotivated. We have already seen that her first strategy

depends on her second and, moreover, has an additional problem of its own if one tries to appeal to it to justify the possibility of supererogation. More work must be done, then, if a bar-lowering strategy is to help the moderate accommodate supererogation.

4.2.4 Magnificence and Optional Virtues

One way, not explored in the contemporary literature, that the moderate might consider lowering the obligation bar so as to make room for supererogation would be to draw a distinction between obligatory and non-obligatory virtues, saying that while all virtues are ideals, some virtues are not required. Supererogation would then be possible when a person manifests these optional virtues. The suggestion that some virtues are optional may at first sound contrived, but it arguably has a long-standing precedent, and we can find this precedent if we return to Book IV of the *Nicomachean Ethics*, looking this time not at magnanimity but at a virtue often associated with it, magnificence. An argument can be made that this virtue, for Aristotle, is not required but is nonetheless an ideal (at least for some people) that allows for the surpassing of duty and, hence, supererogation, even if Aristotle did not himself see the implication.\(^{253}\)

For Aristotle, magnificence is a virtue that operates with respect to wealth and large-scale expenditures.\(^{254}\) As such, it is available only to the wealthy; a poor person who attempts to act magnificently is foolish for so-attempting.\(^{255}\) Magnificence is not merely a matter, however, of spending large amounts of money when one can afford to

\(^{253}\) It is worth noting here that the thesis that Aristotle recognizes categorical moral requirements is controversial.

\(^{254}\) *NE*, 1122a.

\(^{255}\) *NE*, 1122b.
do so; a person who has this virtue is concerned with the fine, and the manner of expenditure accordingly reflects this concern.\textsuperscript{256} Moreover, the virtue involves a success condition on this concern with the fine: a person who spends large sums of money putting on a tacky show, even if she would describe it as “fine,” has not the virtue of magnificence but the vice of vulgarity.\textsuperscript{257} The other vice that stands in the way of magnificence is stinginess.\textsuperscript{258} What is perhaps most notable about Aristotle’s discussion of magnificence, for our purposes, is that, unlike the vices that flank other virtues, neither of magnificence’s opposing vices brings reproach because, says Aristotle, they do not hurt others and are not “too disgraceful.”\textsuperscript{259} So, not only are the poor excused from being magnificent (because they are incapable of it) but also the wealthy who are capable of it are not reproached for falling short of it. This feature of magnificence gives it a claim to being an optional virtue, at least on Aristotle’s picture. One is good when one has it but violates no obligations if one does not.

An initial objection to the theory that magnificence, \textit{qua} optional virtue, makes supererogation possible is that, far from being an ideal, magnificence is a mere vice masquerading as virtue. In this vein, Nancy Sherman complains of the problem that occurs when people use large scale expenditures as an excuse for systematic neglect of less dramatic, everyday duties.\textsuperscript{260} One cannot buy one’s way out of these ordinary duties

\textsuperscript{256} \textit{NE}, 1122a.
\textsuperscript{257} \textit{NE}, 1123a.
\textsuperscript{258} \textit{NE}, 1122a.
\textsuperscript{259} \textit{NE}, 1123a.
\textsuperscript{260} Sherman is discussing magnanimity rather than magnificence, but the objection she raises for the one looms for the other.
by engaging in dramatic projects, and this is why one might reject magnificence as a virtue entirely.\textsuperscript{261} To this objection, the defender of magnificence \textit{qua} virtue has a response: magnificence does not imply the neglect of everyday duty, as the example of Babette, who faithfully executed her daily duties for fourteen years prior to her magnificent feast and, we have reason to believe, for many years thereafter, shows us.\textsuperscript{262} It may be true that a common failing of people attempting magnificence is neglect of ordinary duty, but that does not negate the value of magnificence in cases in which the agent does not so-neglect ordinary duty.

Moreover, if a person does neglect everyday duty because she wants to perform acts that many would believe to be magnificent, we need not say that the expenditure “makes up” for the neglect of ordinary duty, so that the agent has successfully bought her way out of ordinary obligation. Indeed, given the plurality of ordinary (non-optional) virtues, each of which the moderate virtue ethicist sees as grounding what can be very different sorts of moral obligations, it is hard to see how a great expenditure of wealth, even if done for the sake of the fine, could count as making up for most—and perhaps even any—moral obligations. Building an opera house does not negate any lack of courage, sensitivity, or honesty from which the financer may suffer; insofar as distinct duties emerge from each of these virtues, we cannot assume that magnificence, even if it is an optional virtue, can compensate for their lack.

\textsuperscript{261} This is how Sherman rejects magnanimity as a virtue, 106-109, esp. 107.

\textsuperscript{262} The film implies that Babette returns to the sisters and plans to live with them the rest of their lives.
At best, Sherman’s objection shows not that if magnificence is a virtue, it is one that enables people to buy their way out of morality but that, as a matter of fact, people are often neglectful when they do apparently magnificent things. At the end of the day, though, this point is of no greater significance than Aristotle’s own recognition that apparently magnificent acts can be foolish (when the gift exceeds the giver’s means) or vulgar (when the gift is not in good taste). A virtue remains a virtue even if people often confuse it with what is in fact vice, and it likewise remains a virtue if some, confused about its nature, act negligently in trying to embody it. The suggestion that magnificence is an optional virtue that allows for supererogation thus survives the Sherman objection.

A stronger objection to the suggestion that magnificence is an optional virtue is not that magnificence turns out not to be a virtue but that it turns out not to be optional, at least not in a way that allows for supererogation. Recall that one reason for considering magnificence optional is that only the wealthy are capable of manifesting it. But even if magnificence is not required of most people, it does not follow that it is optional; it could be required whenever it is a possibility for an agent, in which case no person who manifests magnificence is manifesting a virtue that she is not required to have. Of

---

263 If one shares Heyd’s view that the value of supererogatory acts is a continuation of the value of obligatory acts, Sherman’s objection likely goes through. Heyd’s view suggests a picture of life as a sort of moral test on which some minimal score is required. One’s actual score is a function of the points one achieves either by fulfilling obligations or performing supererogatory actions, since both trade in the same sort of points. One’s minimal required score is identical to the number of points one would receive were one to fulfill all of the duties one encounters over the course of one’s life. The extra credit available for supererogatory actions can thus “make up” for missing ordinary duties in that one can achieve the minimal required score while neglecting ordinary duties by virtue of performing enough supererogatory actions. Of course, we need not think that a moderate virtue ethicist is committed to this picture or the idea that the merit of supererogatory acts is commensurable with the demerit one acquires when one fails to fulfill a duty, whatever the virtue source of that duty may be. For more on the “extra credit” metaphor of supererogation, see the last chapter of this thesis.
course, as we have seen, Aristotle arguably does not think that magnificence is required, even if we restrict our attention to cases involving persons who have the means to be magnificent, for magnificence’s opposing vices, namely stinginess and vulgarity, do not, Aristotle says, make a person liable to criticism. How might one try to counter this line of thinking? One reason for holding that magnificence, contra Aristotle, is required of those with means has to do with magnificence’s connection to generosity. Generosity is not typically classified as an optional virtue, but given that what it takes to be generous is relative to one’s means, it might be thought that the wealthy person fails to be magnificent only if she likewise fails to be generous. For such a person, the requirements of generosity entail large, well-considered expenditures and hence coincide with the requirements of magnificence. So, just as generosity is required when possible, magnificence is as well.

One could quibble with the previous argument by noting that the requirements for generosity and magnificence do not perfectly coincide, even in the case of the wealthy: a wealthy person, via a number of small-scale expenditures, can exhibit generosity without ever exhibiting magnificence.264 This quibble is of no help to those who want to use magnificence to make room for supererogation, however, because even if the standards for magnificence are slightly different (and more narrow) than the standards for generosity, it is hard to see why a magnificent, generous, wealthy person should be considered morally better than a merely generous wealthy person when the only difference between the two is that the former’s gifts are always in large lump sums rather

__________________________

264 NE, 1122a.
than several smaller ones. So, the fact that the requirements for magnificence are
different and narrower than the requirements for generosity does not show that
magnificence is superior to (mere) generosity, as opposed to a species of it. Generosity
and magnificence appear to be equally strenuous virtues, so if generosity is required (as
most moderates say that it is), magnificence is likewise required, or at any rate amounts
to nothing more than one way of fulfilling the broader moral requirement to be generous.

One might here try to identify some special feature of magnificence that makes it
better than mere generosity. Large-scale giving might be thought to reflect a depth of
commitment not present in small-scale giving. Such a thought, however, is at best
tentative: there seems to be no lack of commitment on the part of the widow in the New
Testament parable who gives her last mite,265 small as the objective gift may be, but the
generous widow’s gift is clearly not an example of magnificence. Even if we exclude
discussion of the widow by explicitly restricting our attention to the wealthy, looking at
how a (merely) generous wealthy person differs from a magnificent wealthy person, we
will find that while magnificence has a place—extraordinary gifts produce goods that
would otherwise not be produced—(mere) generosity on the part of the wealthy likewise
has a place. Both large-scale and sustained small-scale gifts make important
contributions to human well-being. The same can be said of expenditures that serve the
fine versus those that, say, prevent or alleviate poverty. So, neither the object of
magnificence (the fine) nor the particular way in which it is implemented (through large
gifts) ought to be construed as making magnificence superior to mere generosity.

__________________

265 Mark 12:41-44, NRSV.
A final way of valuing magnificence over mere generosity does so on the grounds that the requirements for the former are more difficult, not because acts of magnificence are one-time affairs, but due to the sheer extent of a magnificent person’s giving. On this reading, the magnificent wealthy person gives more in total than the generous person and accordingly surpasses the merely generous person. If the magnificent person were to give more (in total) than a similarly situated (merely) generous person, we might be able to see why magnificence could properly be thought to be optional without the same being true of generosity. Here, however, it needs to be remembered that ‘mere generosity,’ as we are using the term, is not minimal generosity. The ‘mere’ in our ‘mere generosity’ is intended to contrast simple generosity with the more specific virtue of magnificence.

One can be merely generous, in the sense in which we are using the term, and give away as much as the magnificent person; indeed it may even be the case that a supremely magnificent person does not give away as much as the supremely generous person. (Bishop Moriel is supremely generous, but he is not magnificent, and it is doubtful that he could be magnificent, at least in Aristotle’s sense, without increasing his personal standard of living, which would mean acting less generously\(^{266}\).) Having the virtue of magnificence to the highest degree is thus at least as difficult as having the virtue of generosity to the highest degree. In speaking of what is required of us with respect to virtue, moderate virtue ethicists typically assume that it is not minimal but full virtue that is required; all are required to have full virtue, and an individual’s specific obligations in a given situation are a function of what the virtuous agent, construed as the agent with

\(^{266}\) Thus, Aristotle: “It is also proper to the magnificent person to build a house befitting his riches, since this is also a suitable adornment,” *NE*, 1123a.
full virtue, would do. But if the standard for obligation is full virtue, then we cannot say that the requirements of magnificence are more strenuous than those of generosity.  

We are left with the conclusion that magnificence has no better claim to being an optional virtue than generosity has. With respect to the suggestion that optional virtues are the way to supererogation for the moderate virtue ethicist, we thus have three possibilities: (i) that all virtues are required, rendering the optional virtue suggestion as a path to supererogation a dead-end; (ii) that there is at least one optional virtue, but magnificence is not it; or (iii) that magnificence and generosity are both optional virtues (possibly along with some others). Absent a positive suggestion for which virtue(s), other than magnificence, could be optional, option (ii) is the least plausible. Magnificence, both due to Aristotle’s treatment of it and due to our own attitudes toward it, has as high a prima facie claim to being optional as any virtue does; if magnificence is non-optional, it is hard to see which virtue(s) could be optional. Option (i) has a measure of appeal, but drawing this conclusion would be premature until we have considered option (iii). In the next section of this chapter, I will be exploring and defending a form of this third option. Specifically, I will argue that not only is generosity (along with magnificence) optional but also all other virtues are (in a sense) optional.  

\footnote{To those who worry that, if such is the case, the moderate’s notion of obligation is an impossibly high standard, requiring at once that an individual give a large sum to an arts project (magnificence) and at the same time that she give many smaller sums to a number of other projects, leaving her nothing to live on, the moderate has a straightforward reply. Moral obligation is a function of what the virtuous agent, i.e., a person who fully possesses all of the virtues, would do in a given situation. For the wealthy person with still finite means, a number of options will be both possible and consistent with possession of all of the virtues.}

\footnote{For those who question magnificence’s status not as optional but as a virtue, I recommend considering, again, the example of Babette, and the way in which she manifests magnificence. Recall, too, that one can recognize magnificence as a virtue without committing oneself to the claim that Aristotle’s understanding of that virtue is unobjectionable.}
will make this argument by appealing to the vices, rather than the virtues, as the (indirect) source of moral obligation.

4.3 Virtue as Optional

My suggestion is that moderate virtue ethicists who want to accommodate supererogation can do so by adopting the position that, strictly speaking, it is vices rather than virtues that ground moral obligation. Failure to act virtuously is not, in and of itself, morally blameworthy. Rather, acting viciously is always morally blameworthy, and so-acting is the only way to be liable to moral blame. If one avoids acting viciously, one has fully satisfied one’s moral obligations, even if one frequently fails to manifest positive virtue and rarely if ever manifests full virtue. As a matter of practical fact, most people who manage to avoid acting viciously will succeed in manifesting virtue as well, at least to some degree, but the important point for my purposes is that even if virtue often accompanies fulfillment of moral duty, the vices rather than the virtues are what ground this duty. In bumper-sticker form, my view is that while avoiding vice is obligatory, virtue is optional.

On my view, moderate virtue ethicists ought to adopt the following formal account of moral blame:

For any intentional activity $A$ performed by any agent $S$ in any circumstances $C$, $S$ is morally blameworthy for $A$-ing in $C$ iff $S$’s $A$-ing in $C$ manifests some vice.

A few comments are necessary in order to clarify the above account of moral blame. First, I use the term “intentional activity” rather than “act” (the typical term used in discussions of this sort) in order to avoid some of the more narrow ways in which “act” and “an action” tend to be interpreted. The broader term, “intentional activity,” should be
understood to include refraining from action in cases in which the refraining is voluntary. Further, something can count as a person’s “intentional activity” even if it is not an isolated action (or non-action); thus, “waving to Glenda,” “writing the rent check,” and “taking out the trash today” will all count as intentional activities, but so will “making sushi,” “growing tomatoes (over the course of a summer),” and “taking out the trash regularly.” Thus, an intentional activity can be a collection of determinate actions that occur over a relatively long period of time. Accordingly, the circumstances in which an intentional activity occurs can be narrow or broad and span a short period of time or a much longer one. Not just any collection of an agent’s actions will count as “an intentional activity;” the actions in question must be unified by the intentions of the agent, which is to say that the agent’s reasons for performing the various actions must be connected to some overarching reason, way of life, or end to which the agent is committed, even if the agent is not conscious of the commitment. This notion of unity in intention is admittedly vague, and a full explication of it would take us too far from the topic at hand. I trust, however, that the idea of an agent’s actions being unified in that agent’s intentions is sufficiently intuitive that I may reference it in an explanation of “intentional activity.”

A second point I should make about the above account of moral blame is that it presupposes an independent account of vice. Such an account of vice is explicit or implicit in almost every defended version of moderate virtue ethics, but the various vice accounts do differ significantly from one another. For this reason, moderate virtue ethicists can disagree with one another substantially on the question of which character traits are vices. What turns out to be morally blameworthy, on my view, will of course
depend on which of the competing lists of vices turns out to be correct. For the sake of
convenience, I will assume that the character traits that appear on standard lists of the
vices—traits such as cowardice, cruelty, arrogance, selfishness, pettiness, callousness,
and malice—are genuine vices. I recognize, however, that discussion of how any
particular moderate virtue ethicist could adopt my account of moral blame would have to
proceed from the list of the vices recognized by that particular moderate’s theory.

A third clarification I should make concerns what it means for “S’s A-ing” to
“manifest” vice. To say that an agent’s intentional activity manifests a vice is to say that
there is some vice that the agent has, at least to some degree, and the possessed vice in
some way explains and describes the agent’s behavior. So, if the activity manifests
cowardice, we can truly say of the agent both that she is (to some degree) cowardly and
that, in this particular situation, she acted cowardly, that she acted as she did (at least in
part) because she is cowardly. The reason that vices can partly explain intentional
activity is that they are character traits, so their possession presupposes certain facts about
the possessor’s psychology. Calling a person cowardly is thus not simple shorthand for
the claim that the agent often fails to take risks she ought to take; it is a claim about the
agent’s psychological state, a state that can at least partially explain some of her actions.

Precisely because vices are (often complex) character traits that map onto certain
psychological facts about a person, whether or not a particular intentional activity
manifests a particular vice (or virtue) will depend upon a number of factors beyond the
ones that tend to concern deontologists and consequentialists. Deontologists and
consequentialists tend to be concerned with the circumstances in which an action occurs,
the action’s actual and/or expected consequences, and/or whether or not the action falls
under a description seen as “morally significant” (e.g., “lying,” “benefitting,” “cheating”). On the virtue account defended here, we may know all of the previous information and yet still not know whether or not a given activity, as performed by a particular agent, is cruel, petty, or callous, for instance. To make that evaluation, we may still require information about the agent’s intention (the end for which she acts) and affective state (before, during, and after the activity). Taking all of these factors into consideration allows my account to honor the virtue ethics mantra that it is not just what is done but how it is done that is relevant to moral assessment (including, for the moderate, blame assessment).  

As it stands, the view I have put forward tells us when agents are morally blameworthy but not what they are obligated to do. It is often assumed that the latter explains the former. On my view, however, the explanation goes the other way. To determine moral obligation in a given situation (whether the situation be broad or narrow), we begin by assessing all of the various ways in which a person might perform all of the intentional activities open to her. For some of the activities open to her, any intelligible participation in them on her part would make her liable to blame for it would involve manifesting vice; for other activities some ways of performing them would involve vice (making her liable to moral blame), and other ways would not (making those non-vicious ways of participating in those activities morally licit). Our agent is obligated

\[\text{\footnote{I consider this feature of my account a mark in its favor. Notably, standard moderate accounts that appeal to what the virtuous agent would do don’t honor the mantra.}}\]

\[\text{\footnote{Speaking of agents as morally blameworthy is somewhat misleading. People are morally blameworthy, but not only \textit{qua} agents. We rightly blame people not only for acts but also for more interior features of their ethical lives, for instance, their attitudes and affections.}}\]
to avoid manifesting vice. Hence, she is obligated to avoid all intentional activities that she can only intelligibly perform if she manifests vice. Hence, she has a disjunctive duty to perform one of the other activities that she can perform non-viciously, and she must perform whichever of these activities she selects in a non-vicious way.\textsuperscript{271} Provided she does this, she fully satisfies her moral obligations and is not subject to moral blame.

With the above understandings of moral blame and obligation in mind, we are now in a position to consider my preferred moderate virtue ethics account of supererogation, which I will call the Optional Virtue Account of Supererogation (OV):

For any intentional activity $A$ performed by any agent $S$ in any circumstances $C$, $S$’s $A$-ing is supererogatory in $C$ iff (i) $S$’s $A$-ing in $C$, manifests no vice; (ii) $S$’s $A$-ing in $C$ manifests some virtue $V$ to some degree $D$; and (iii) $S$’s acting intelligibly and manifesting no vice in $C$ does not imply $S$’s acting in $C$ in a way that manifests at least one virtue to degree $D$.\textsuperscript{272}

\textsuperscript{271} Only when only one intelligible option can be done non-viciously will she have a positive duty to engage in a particular intentional activity. If we think of “circumstances” narrowly as particular states of affairs that obtain only for an instant or a short period of time, then positive duties will be rare—for at almost every moment we face a wide range of ways of acting non-viciously. If we think of “circumstances” more broadly, though, as states of affairs that can extend over long periods of time, positive duties will be more common. If I promise to repay someone a certain sum of borrowed money within the week, provided I pay him before the very end of the week, I will never face a particular point in time at which not manifesting the vice of injustice or infidelity requires that I *now* repay the borrowed money. Nonetheless, I can know, at the outset of the week that I have a positive duty to repay the money at some point during the week. For, in all likelihood, if the week goes by and I fail to repay the money, I will, in keeping the money (whether intentionally or unintentionally), be manifesting at least one vice.

\textsuperscript{272} Jason Kawall investigates some suggestions for lowering the virtue ethics obligation bar (from what the supremely virtuous agent would do) in order to accommodate supererogation. Kawall complains that such attempts will yield implausible implications for what is “right.” If “right” is equivalent to “obligatory,” then the claim that an action is right iff a minimally virtuous agent would characteristically perform it implies that mediocre actions are obligatory. If “right” is equivalent to “morally acceptable,” then the claim that an action is right iff a minimally virtuous agent would characteristically perform it still permits the morally wrong—because minimally virtuous agents sometimes do what is wrong, and they do so characteristically, as the Milgram experiments show us. Notably, OV, with its associated account of moral obligation, does not fall to the problems Kawall raises because OV does not appeal to what virtuous agents (or non-vicious agents) would do. Instead, it discusses whether or not vice is manifested. The behavior we see in the Milgram experiments manifests vice and can accordingly be seen as wrong, 180-7.
The central idea of OV is that one satisfies moral obligation by avoiding vice, and one’s intentional activity is supererogatory when, in addition to not manifesting vice, it manifests virtue in a way that cannot be fully attributed to the avoidance of vice. In stipulating that manifestation of the virtue cannot be fully attributed to the avoidance of vice (iii), I am trying to avoid crediting a person with supererogation in cases in which that person will avoid a particular vice (e.g., dishonesty) only if he manifests its opposing virtue (e.g., honesty), at least to some degree. If there are situations like this (and I am not sure that there are), manifesting the virtue to the degree required to avoid vice (in the situation) will not count as supererogatory, on OV, although manifesting the virtue to a higher degree than is required for the avoidance of vice will count as supererogatory.

OV easily lends itself to a threshold understanding of supererogation. In all cases of supererogation, the agent manifests a higher degree of virtue than he needs to manifest in order to avoid vice. An act is supererogatory to the extent that it manifests a higher degree of virtue than what (if any) is needed to avoid vice. Both increasing the number of virtues that one manifests and increasing the extent to which one manifests each will increase the extent to which one’s activity is supererogatory. If we set aside, then, the cases in which some degree of virtue is required for avoidance of vice, we can say that, in a case of supererogatory activity, the activity is supererogatory to the extent that it is overall virtuous.

4.4 Defending the Optional Virtue Account

The most controversial aspect of OV is likely its associated accounts of obligation and blame. Moderates have long seen right action as grounded in the virtues or the
actions of virtuous agents, and switching to the vices as the grounds of blame and (by extension) moral obligation is a significant change. Even Swanton, who wants to defend a less-rigorous virtue ethic than the standard moderate picture of full-virtue-as-obligatory, never considers drawing the line for moral obligation below virtue at the avoidance of vice. My switch to the vices might thus be seen as ad hoc, insofar as it departs from the tradition, and too lenient, insofar as avoidance of vice is a standard too easily met (and, for that matter, surpassed).

While I admit that no moderate, as far as I know, grounds moral obligation in the avoidance of vice rather than in the manifestation of virtue or the actions of the virtuous agent),\(^{273}\) I do not think my view constitutes as great a departure from the moderate virtue ethics tradition as may at first appear. In the first place, there is reason to doubt that the “full-virtue as obligatory” picture of moderate virtue ethics has as much support as is generally assumed. A number of moderates identify right action with (roughly) an action that a similarly-situated virtuous agent would perform. But notably, although most of these moderates appear to understand “right action” in the usual sense in which right action corresponds to what an agent is obligated to do, at least one prominent moderate, Rosalind Hursthouse, does not understand “right action” in this way. Hursthouse develops her account of “right action” to respond to the objection that virtue ethics cannot guide action. Her use of the term “right” in “right action” is, she says, the same use of “right” that occurs in a statement such as “he knew the right answer.” We thus cannot

\(^{273}\) Some of Trianowsky’s comments can be seen in this vein, although he never gives us an actual account of supererogation, and his views on less-than-full virtue are incompatible with OV. See “Supererogation, Wrongdoing and Vice: On the Autonomy of the Ethics of Virtue,” *Journal of Philosophy*, 83 (1986): 26-32.
assume that “right action,” Hursthouse, is obligatory action.274 The virtuous agent can be a guide to behavior, “right” behavior in the sense of “best choice,” without serving as the standard one must achieve in order to avoid moral blame. If Hursthouse is using the term “right” in this ideal, and action-guiding but not necessarily action-requiring sense, then my view is not far from hers.

Even if my view is far from Hursthouse’s, and even if I cannot point to any moderate virtue ethicist who endorses my account of moral blame and obligation, my account of obligation should still not, I shall argue, be seen as ad hoc. As I have already mentioned, seeking a relaxed view of moral obligation is not uncharted territory for moderates, and the thought that morality is somewhat relaxed is an intuitive one. The difficulty for the moderate who wants to give credence to this intuition is to that of providing a non-arbitrary line of moral obligation (or blame, if one were to follow Swanton’s suggestion of distinguishing between the right and that which must be done to avoid blame) that is located below the line we normally associate with full virtue. Drawing the line between virtue and vice, as I do, solves this problem, enabling the moderate to say that, because the virtues are always (positively) good character traits, one is liable to moral criticism, and hence moral blame, if and only if one manifests vice.

We can further appreciate the non-arbitrary nature of setting the standard for moral obligation at the avoidance of vice, rather than at the manifestation of positive

274 Indeed, the very fact that Hursthouse takes time to specify her notion of “right” suggests that she is not using the term this way, On Virtue Ethics (New York: Oxford University Press, 1999), 108.

275 There is a sense in which we can criticize failure to achieve full virtue, but this is not the sort of criticism associated with accusing a person of vice. Telling a person she is vicious is fundamentally different from telling a person that she has room for growth.
virtue, by attending to some specific virtues the very logic of which appears to imply their being optional. To return to two examples we have already discussed, both generosity and magnificence seem, contra the standard moderate picture, non-required. Generosity concerns giving: if a person acts generously, she is giving more than what she owes, if anything, in the situation in question.\textsuperscript{276} Bearing this feature of generous action in mind, it is hard to see how acting generously could ever be morally required. Since generosity is a virtue, we thus have at least one virtue that ought not be seen as generating moral obligations with respect to actions: the fact that an act is such that the generous person would do it, or such that a person with all virtues, including generosity, would do it, or even such that it manifests some degree of generosity and is not otherwise wrong should not be seen as sufficient reason for the act’s being positively morally required. Generous acts, by their very nature, are not morally owed.

What is true of the logic of generosity is also true of the logic of a number of other virtues. Specifically, magnificence, magnanimity (in the common, non-Aristotelian sense that is associated with forgiveness and does not presuppose full virtue), thoughtfulness, kindness, sensitivity, and gentleness are all virtues such that, the mere fact that a particular act would manifest these virtues does not give us reason to think that the act in question is \textit{obligatory}. Nor is the absence of obligation fully attributable to potential conflicts with competing virtues: even if we know, of a given situation, both that act \(X\) would manifest one of the virtues under discussion and that act \(X\) would not violate moral obligation in the situation, we still do not know that \(X\) is obligatory,

\textsuperscript{276} This is not to say that giving more than what one owes is a sufficient condition for acting generously.
whether strictly or disjunctively (so that \( X \) constitutes one way in which the agent could fulfill her duty in the situation in question). Action \( X \) may be a good act, a virtuous act, without being an obligatory act simply because failure to manifest the virtue it manifests when one has the opportunity to manifest it is not the sort of failure that makes one liable to moral blame.

Note that one can hold that the virtues now under discussion are optional—so that in no situation is a person obligated to manifest them—without holding that the vices opposed to these virtues are morally acceptable. Pettiness is opposed to magnanimity, stinginess and prodigality to generosity, callousness to sensitivity, and so forth. Actions that manifest these vices are blameworthy precisely because they manifest these vices, but it does not follow that manifesting the opposing positive virtues is required. A person who is obsessed with securing what is rightfully hers is stingy and is accordingly liable to moral blame when she manifests stinginess, but it does not follow that the person is ever required to act generously; it is enough if the person abandons her obsession with getting her share. If, despite abandoning the obsession, the person gives to others only what is their due and never gives any more while manifesting the mental states associated with generosity (a surprising feat, but a possible one), we have an agent who has managed to satisfy moral obligation (in one sphere of life) without ever acting generously. Similar patterns will hold for the other virtues and vices that are now under discussion: there will often be a neutral state between vice and virtue such that the agent can avoid vice without acting from virtue. If the agent does so, the agent has discharged relevant moral obligations.
For a number of virtues, then, the claim that manifesting them is optional rather than required is not only not ad hoc, it is quite intuitive. The question we must now ask is whether all virtues are of this sort. For the claim that honesty, justice, and even gratitude are non-required is much less intuitive than the claim that gentleness and generosity are not required. Since my view is that no virtue grounds an obligation, this is an objection that I must take seriously. The first thing to say in reply is that, as I have already stressed, the fact that virtues are optional does not imply that their corresponding vices are acceptable. So, even if humans are not required, on pain of moral blame, to have the virtue of honesty, manifesting dishonesty will remain reprehensible. Given this implication of my view, it may turn out that manifesting virtue is sometimes actually required, despite the fact that the vices rather than the virtues (indirectly) ground moral obligation. For there may be some situations—most plausibly those involving honesty and justice—in which neutrality with respect to the manifestation of virtues and vices is not an option; that is, there may be situations in which one will avoid manifesting a particular vice, such as dishonesty or injustice, only if one manifests its opposing virtue, at least to some degree. In such situations, acting virtuously will be obligatory, but crucially the optional virtue account of supererogation presented here and its associated account of obligation allow us to recognize such virtuous action as obligatory but, in a sense, accidentally so, the source of the obligation being not a virtue that, qua virtue, must be manifested but the vice that, qua vice, must (on pain of moral blame) be avoided.

One problem with the above response is that the vices I cited, dishonesty and injustice, that render low-level honesty and justice required are arguably not actual vices. Let us take the example of dishonesty; what I will say about it can be adapted for
discussions about injustice as well. In order for dishonesty to be a vice, it has to map on to some feature in our psychology that explains behavior. For this reason, the vice of dishonesty, if it is an actual vice, is *not* “a disposition to lie,” for few people, if any, have such a disposition, even if they do, as a matter of fact, frequently lie. What explains the frequency with which individuals lie is not some “disposition to lie” but certain other psychological traits combined with a lack of respect or concern for the truth or their listeners. It is at least arguable that we cannot intelligibly speak of such a lack as a psychological state. If such is the case, we cannot intelligibly speak of ‘dishonesty’ as a vice. Saying a person is dishonest, then, does not attribute a vice to that person; it simply (i) reports that the person lacks the virtue of honesty and/or (ii) reports that the person frequently engages in deception. Similar comments can be made about injustice.

I am not sure about what to make of the above discussion and whether or not we should consider dishonesty and injustice (and possibly also ingratitude) genuine vices. For this reason, I will not assume that they are and will develop a second reply to the concern that certain virtues—namely justice, honesty, and gratitude—are non-optional; this second reply will not presuppose that injustice and dishonesty are vices that, *qua* vices, must be avoided. This second reply begins by noting that it is not so much that we

---

277 This point closely parallels Swanton’s view of (some) virtues as involving an appropriate response to the values in its field, *Virtue Ethics: A Pluralistic View*, 19-24. We thus might say that honesty involves an appropriate response to the value of truth. Such a characterization of honesty is non-circular. Truth has a certain value, a point we can assume just as we can assume that pain has a certain negative value. Because truth has this value, it is not a stretch to think that respect for this value constitutes one way in which human beings flourish and that this explains why honesty is a virtue. When we say that honesty is a virtue because it is or involves or makes possible an appropriate respect for truth, crucially, we are not defining “appropriate” intuitively based on how much and in what context we think certain responses to truth are good or bad. What makes the respect “appropriate” is that it involves a successful recognition of a prior value that exists in our world. Thus, virtues, at least some of them, may involve appropriate ways of seeing the world, where propriety is not determined by intuitive admirability.
think that we are required always to manifest the virtues of honesty, justice, and gratitude—sometimes no action open to one will positively manifest justice, for instance, but that does not mean that all of one’s options are immoral—but rather that morality prohibits certain behaviors such as lying, cheating, and the forgetting or ignoring of one’s benefactors qua one’s benefactors. The challenge for my view is to explain how such behavior could be ruled out, even generally, on a virtue-ethics account of obligation if (i) positive virtues do not ground moral obligation and (ii) the ‘vices’ that would (almost) automatically rule out the behaviors in question, specifically the ‘vices’ of dishonesty, injustice, and ingratitude, are not assumed to be actual vices. 

The challenge can be met by realizing that, even without the inclusion of dishonesty (and the other questionable vices), the list of vices is extensive enough to rule out most instances of the problematic behavior. Thus, in most situations, the requirement to avoid manifesting other vices such as greed, vanity, wrath, self-centeredness, and jealousy will render forbidden all intelligible forms of deception that are open to the agent. (Similar points can be made about activities normally described as “unjust” or “ungrateful.”) To see why this is the case, we need to attend to both the concept of intelligibility of action and the actual reasons that people lie (or cheat, or bite the hand that feeds). An action is intelligible only if performed by an agent for intelligible reasons. For most (but not all) situations in which an agent is able to lie, the only ways in which she can lie for intelligible reasons are such that, given her reasons, the lie will

\[\text{\textsuperscript{278}}\] I explain below why it is the case that the vices of dishonesty and injustice (assuming these are actual vices) only almost rule out the problematic behaviors.
manifest at least one of the vices that I have mentioned. So, we are typically obligated to refrain from lying.

The previous argument holds because people lie for reasons, and a consideration of these actual reasons reveals certain morally significant patterns; thus, they do not (typically) lie because they find uttering falsehoods to be an intrinsically pleasant activity. I am not the first to note this feature of lying, i.e., that it is associated with certain patterns of reasons. Kant at least implicitly recognizes it and relies on it in generating a universal prohibition against lying. Kant’s categorical imperative rules out lying because the maxim involved when a person lies specifies that person’s reason for action, and as it turns out, people’s actual maxims with their associated actual reasons are not universalizable, a point we appreciate when we apply the categorical imperative to particular cases of lying, as Kant does. Consider, however, the following possible non-standard case of lying: a person utters words that are false simply because he enjoys uttering false sentences aloud; he does not care at all whether or not his listener believes what he says, but he fully realizes that his listener will in fact believe him and form a false belief. The categorical imperative cannot condemn such an action: it is possible for us to live in a world in which, whenever one feels like uttering a false sentence, one does so, and it is also possible for a person who acts in such a way to will that such a world be actual. So long as the liar’s achieving his intended end does not instrumentally require successful deception of others, the categorical imperative cannot prohibit the lie. For

---

279 One might object that if one does not intend for the false statement to deceive (either as an end or a means), then the false statement does not count as a lie, even if one knows that the false statement will deceive and is unconcerned with this feature of it. I am willing to concede this point, in which case Kant’s categorical imperative does rule out all forms of lying. Note, however, that my larger point remains true: the reasons that people engage in lying are tightly connected to why lying, for Kant, is forbidden.
both Kant and me, then, the reason that lying is usually immoral has something to do with
the actual reasons that people lie. Kant and I differ, however, in what use we make of
these reasons in constructing our rules against lying. He looks at their connection to
universalizable maxims; I look at their connection to particular vices.

My view differs from Kant’s, too, in being more tolerant of certain instances of
lying. Although my vice-check on right action is stringent, it does not force us to say that
any intelligible instance of lying in an effort to deceive violates moral duty in virtue of
manifesting some vice, so that even lying to the Nazis about the presence of Jews in
one’s house would be morally wrong. The vices, as I have stressed, are character traits,
not principles or dispositions to break particular rules. To tell a Nazi that there are no
Jews in one’s house when one knows this not to be the case is to lie, but not all instances
of lying manifest the vice of selfishness, vanity, ruthlessness, or any other. Indeed, even
if dishonesty is a genuine vice, not all instances of lying manifest it. To draw upon a
point that Hursthouse makes, we can say that because vices are character traits, whether
or not one’s lie manifests any particular vice depends upon certain features of one’s
psychology and the connections those features have to one’s lying. The dishonest (or
cowardly, etc.) person’s attitude toward a lie that manifests dishonesty (or cowardice, or
any other vice) will be quite different from the attitude of the person who lacks this vice
and utters the same words when the proposition expressed has the same (false) truth-
value. (Thus, often the non-dishonest person will experience a type of regret not present
in the dishonest person). And if a person who lies on a particular occasion does so
despite not possessing the vice of dishonesty (as happens in usual readings of the Nazi
case), it is not correct to say that the person manifests the vice of dishonesty; one cannot

166
manifest a vice one does not have. Unless our imagined agent’s lie manifests some other vice, then, the agent is not morally blameworthy for lying to the Nazis. We thus can, by appealing to the vices, generate a fairly rigorous morality—rigorous given how many vices must be avoided—without making the morality so stringent as to forbid even lying to the Nazis to protect the innocent.

There is one type of case in which supererogation can be said to be rather easy, on OV: when vice is easily avoided and a low degree of virtue easily manifested, small acts of virtue will count as supererogatory. Note, however, that the degree of supererogation in such a case will be low. Considering there is a literature, independent of my account of supererogation, that defends the supererogatory status of opening the door for a stranger, this implication of my account should not be seen as a great liability. Supererogation is a threshold notion, so we should not be surprised if some activities count as supererogatory in a technical sense while involving such a low degree of virtue as to render their inclusion in the category almost—but not quite—overly liberal.

A more serious problem for my view is not that the standards I set for obligation and supererogation are too low but that they are too high. Attempting to catalogue all of the vices reveals a long list of character defects. While some of these defects, e.g., cruelty, are not terribly difficult to avoid, at least for most of us, a number of them, e.g., selfishness, vanity, and thoughtlessness, are common character traits that infect much of what we do. On my account, one cannot be credited with supererogation for a particular activity, no matter how extraordinary that activity may be, if it in some way manifests rashness, fickleness, insensitivity, shallowness, or any other vice. Avoiding all manifestation of vice in all of one’s activities is a tall order and not one that any of us
meets. Even though OV can say that all of one’s intentional activities need not be free from vice in order for some of them to count as supererogatory, given that any manifestation of a vice “poisons” the activity, one might argue that “large” or long-term activities will almost always be infected with vice at some level. Thus, when we look at Paul Kocheleff’s work in Africa, given that Kocheleff no doubt has some vices, we almost assuredly will find evidence of his vices multiple times in the various ways in which he performed his work in his decades-long stint in Pietermaritzburg. These vices, and their manifestation, though, should not disqualify his extraordinary work as supererogatory, even if they make that work as a whole less-than-perfect. So, OV’s standard for supererogation, it may be argued, is too high.

In responding to the objection that OV’s standard for supererogation is too high, I want first to admit outright that OV’s standard for moral obligation (and by extension supererogation) is quite difficult to meet: most people violate it on a daily basis, for most people struggle with various vices that they regularly manifest in their activities. Whenever they so-act out of their vices, they make themselves liable to moral blame. Note, however, that acting viciously—i.e., out of vice—ought to be seen, on any plausible virtue ethics account of moral obligation, as wrong, even if avoiding this sort of behavior is very difficult. The fact that my account has this implication does indeed mean that its standard for moral obligation is high, but this is not a problem for my account; on my view, we simply must recognize, as most moral theories do, that morality is difficult. Virtue ethics may push the standards, in one respect, higher than other theories, given that it is not merely one’s outward actions but one’s inward life that must
measure up, but this again is simply a fall-out of the standard moderate virtue ethics picture of the moral life as concerned with more than action-performance.

Morality is not so difficult as to make supererogation practically impossible, even in the cases of “large,” long-term activities, however, and OV can recognize this fact, the objection introduced above notwithstanding. Let us return to the example of Paul Kocheleff’s work in Africa. Insofar as Kocheleff’s work (or, perhaps better, his “project”) can be unified under one overarching goal—and I take it that it can—it can be seen as a single activity and hence as a candidate for supererogation. The two checks on whether or not Kocheleff’s work counts as supererogatory, and to what degree, are (i) the question of whether or not his work manifested vice and (ii) the question of whether or not it manifested virtue, and to what degree. The argument we have seen so far for why Kocheleff’s work does not count as supererogatory states that Kocheleff had vices, and he likely manifested some of his vices on multiple occasions during his tenure in Africa (even while working on his ‘project’ in Africa, not merely in his recreational time). Thus, if he has the vice of irascibility, say, no doubt there were many times that he lost his temper while working; perhaps this happened fairly often when he was seeing patients, giving instructions to nurses, or arguing with political officials.

For the sake of argument, let us assume that Kocheleff did indeed sometimes behave in the ways indicated at the end of the previous paragraph. The question for us is whether or not such behavior implies that the intentional activity that we have been calling “Kocheleff’s work in Africa” manifested the vice of irascibility. Answering this

\[280\] If this project is to count as an intentional activity, it must be unified under a single intention. I suggest that we think of that intention as something along the lines of “preventing and alleviating some of the most devastating consequences of world poverty.”
question requires getting clear on how to understand the intentional activity under consideration. Recall that intentional activities are individuated by intentions: when actions are unified in an agent’s intention, all of the actions that are so-united are considered part of a single intentional activity. Let us stipulate that the intention that unites Kocheleff’s work and enables it to count as a single intentional activity is something along the lines of “preventing and alleviating the worst effects of global poverty.” All actions that Kocheleff performed that were intended to serve this goal (directly or indirectly) will thus count as part of the single activity we are trying to evaluate. So, if Kocheleff lost his temper while discussing a patient’s condition with a nurse, then given that Kocheleff engaged in the discussion as a part of his larger project, Kocheleff’s losing his temper will be included in the activity, “Kocheleff’s work.”

Even if this is true, and even if we can say that, in losing his temper, Kocheleff was acting out of the vice of irascibility, we cannot conclude that his work manifested the vice of irascibility, and there are two reasons for our inability to draw that conclusion.

First, determining whether or not Kocheleff’s work manifested vice is not equivalent to determining whether, on some occasion while Kocheleff was doing his work, Kocheleff lost his temper. The question is whether the activity as a whole manifests vice, not whether some constitutive part of it manifests a vice. Proving that the part manifests a vice simply proves that the part qua distinct intentional activity violates duty; we still have the question of whether the admittedly imperfect whole, as a whole, manifests vice. The answer will be negative, even if Kocheleff lost his temper multiple times.

This will be the case even though Kocheleff did not lose his temper in order to pursue his goal, for how Kocheleff performed the various sub-actions that were performed in pursuit of his overall goal is relevant to how he performed his overall activity/project, which is in turn of concern to the virtue ethicist.
times over his decades-long tenure, provided that the frequency of such outbursts relative to more controlled interactions is sufficiently low.

If Kocheleff’s outbursts were so frequent and strong during his work in Africa as to make it the case that his work, *on the whole*, manifested the vice of irascibility, even to a low degree, then OV is forced to say that Kocheleff’s otherwise extraordinary work fails to qualify as supererogatory. In making this admission, it may appear that the objection under consideration—that OV’s standards for supererogation are too high—has succeeded, but I think such a conclusion is premature. If Kocheleff’s temper is so bad that the *way* in which he went about alleviating poverty over a period of multiple decades on the whole manifests his irascibility, then it is not at all obvious that we should say that what he did was supererogatory; those who disagree should consider, alongside the goods Kocheleff secured for his patients, the frustration so many of his colleagues and others would have faced in working with a person like him. Note, too, that even if we cannot say that Kocheleff’s work, as he performed it, was supererogatory, we can still find ways of praising it, on the moderate virtue ethics picture. For the moderate can still talk about how Kocheleff’s activity manifested some virtues he was not required to manifest and that he manifested them to a very high degree, one that very few achieve. So, the work Kocheleff did can be seen as non-supererogatory and even as making Kocheleff liable to moral blame *in certain respects* even as it is seen as extraordinary and, in many ways, quite virtuous.

* * * * * * *

As we have seen, OV is an account of supererogation that emerges quite naturally from the central commitments of moderate virtue ethicists. For on OV, states of
character ground moral obligation; vicious ways of acting are universally off-limits and blameworthy; and the line demarcating moral obligation is relaxed enough to fit many of our intuitions about the rigorous nature of morality without being \textit{ad hoc}. OV’s standards for supererogation are high, but not so high as to make supererogation practically impossible. There is also a sense in which OV’s standards for supererogation are low, for when one can avoid manifesting vice without manifesting virtue, any intentional activity that manifests any degree of virtue will count as supererogatory. Yet the way in which OV’s standards are low is not problematic because the degree of supererogation involved in barely surpassing a low standard will be low, which is acceptable given that threshold understandings of supererogation more generally are acceptable. Given the way in which OV balances the commitments of moderate virtue ethics and the rigors and freedoms of morality, we have in OV a candidate account of supererogation for the moderate virtue ethicist that can be evaluated alongside the Kantian and utilitarian accounts that we considered in the previous two chapters. We will now turn our attention that side-by-side evaluation.
5.1 Paring Down the Accounts

We have now considered a number of utilitarian, Kantian, and virtue ethical strategies for accommodating supererogation. Given the limitations of space, if we are to ask the question of how well, relative to one another, the moral theories of utilitarianism, Kantianism, and moderate virtue ethics can accommodate supererogation, we will have to pare down the accounts we have thus far considered to a more manageable number. Fortunately—or unfortunately, depending on one’s perspective—some of the accounts we have seen are so problem-riddled that we can eliminate them from consideration altogether: for some of the proffered accounts are not just weak but fail outright as accounts of supererogation rooted in the moral theory they are intended to represent. They either depart too radically from the home moral theory to claim membership in that moral theory, or they fail adequately to explain why what they say is supererogatory is supererogatory, or their picture of supererogation is so unusual that it cannot be said to be a picture of the moral phenomena that we (in chapter one) set out to accommodate.

The first section of this chapter identifies such unworkable accounts, thus eliminating them from further consideration. The accounts that remain—ultimately, one utilitarian, one Kantian, and one virtue ethics account—are evaluated alongside one
another in the second section of the chapter. The third section argues that this side-by-side evaluation indicates that virtue ethics, at least in its moderate form, is better positioned than Kantianism or utilitarianism to make sense of the apparent moral phenomenon of supererogation. I then say something about what sorts of prospects a radical ethic of virtue has of making sense of this phenomenon. The chapter closes with some thoughts on one way in which the contemporary literature on supererogation, as a result of the dominance of deontological and especially consequentialist patterns of thinking in contemporary moral philosophy, fails adequately to appreciate an entire species of supererogation: what I call “extravagant love.”

5.1.1 The Utilitarian Accounts

Let us quickly review the utilitarian accounts of supererogation that we have so far considered. We looked at three: Urmson’s basic duty account, Attfield’s low-stakes view, and Portmore’s Schefflerian Utilitarianism (SU) and its associated form of supererogation. Of the three, Attfield’s low-stakes account, which states that we are obligated to maximize utility when much is at stake but are free to act as we please in low-stakes cases, making supererogation possible when in such low-stakes situations one promotes more utility than the minimum possible, is so far from capturing the phenomenon that concerns us that we will here leave Attfield behind. Precisely because the account is a low-stakes account, it cannot view any of the (extraordinary) central cases (or any cases like them) as supererogatory and so does not qualify as the sort of
view that we set out to find. Even if we can admit that supererogation is a threshold notion and that low-stakes cases can involve low-level supererogation, the paradigm cases of supererogation remain, for the most part, high-stakes cases, and any account of ‘supererogation’ that rules out all such cases is inadequate.

Of the two remaining utilitarian accounts, Urmson’s and Portmore’s, Urmson’s is the only one that fits sufficiently well with utilitarian commitments to qualify as a genuinely utilitarian account of supererogation. Portmore’s SU, if it is to recognize apparent instances of supererogation as supererogatory and see them as valuable, is so far from the heart of utilitarianism that its status as a utilitarian theory is, at best, formal. SU, recall, distinguishes two types of utility (utility that accrues to the agent and utility that accrues to others), institutes a weighting function to calculate a third type of value (‘total adjusted well-being’) by using the first two types of utility as inputs, and stipulates that moral obligation is a function of the so-calculated (expected) total-adjusted well-being and the (expected) utility that accrues to others. To allow for supererogation, SU further stipulates that when a person has multiple acceptable options, the person’s act is supererogatory if and only if the act yields more utility for others than some other morally acceptable option.

The reasons that the above SU account of supererogation is too far from the heart of utilitarianism to qualify as utilitarian in substance are multiple, but we will restrict our attention to one such decisive reason. This reason turns on the insight that at the heart of utilitarian moral theory is the claim that morality and moral obligation supervene on

---

282 One of the central cases, Torture, it should be noted, is not a high-stakes case; it’s a no-stakes case. Precisely because it is no-stakes, though, Torture also will not come out as supererogatory on Attfield’s account.
significant, pre-moral values of states of affairs that actions produce. Total adjusted well-being, however, a primary determinant of obligation on SU, does not correspond to any such significant, pre-moral value that inheres in states of affairs, or even in states-of-affairs relative to agents. As we saw in chapter two, learning that a particular state of affairs has a particular total-adjusted well-being score of 90, say, relative to a particular agent and that another state of affairs has a score of 30 relative to the same agent does not tell us anything of pre-moral significance about the two states of affairs (even relative to the agent) or any pre-moral relationship that holds between the two states of affairs. We do not know, for instance, that the first holds three times the pleasure or even that it is three times as good for the agent in some particular non-moral way as the second. Insofar as SU bases moral obligation on total-adjusted well-being, it is thus not calculating and using exclusively pre-moral value to determine moral obligation. On SU, the moral is not supervening on non-moral value. Rather, SU is making certain assumptions about the extent to which people ought to benefit one another (relative to themselves) and stipulating that states of affairs have values that correspond to such obligations. But because the assumed obligation is the basis for the value and no further value is the basis for the (assumed) obligation, the good is not prior to the right, and the heart of utilitarian moral theory has been compromised.

With Attfield’s account rejected for its inadequacy in recognizing supererogation and Portmore’s account rejected as insufficiently utilitarian, we are thus left with Urmson’s view, or perhaps something like it, as the utilitarian solution to supererogation. According to Urmson, we are obligated to obey only those rules regular adherence to which is required if the summum malum is to be avoided. We fulfill our duty—“basic
duty” as Urmson calls it—just in case we obey these rules. We act supererogatorily just in case, on a particular occasion, we promote more utility than we could have promoted merely by acting in conformity with any applicable rules. Urmson’s view’s fit with traditional utilitarian views is certainly far from perfect: his basic-duty distinction entails the rejection of the maximizing requirement, a standard commitment of utilitarian moral theories, and, as we saw in chapter two, the particular reasons Urmson offers to justify his basic duty distinction are often seen as poorly motivated. But, despite all of the justificatory problems, Urmson’s general way of reducing the moral requirement—by blending direct and indirect utilitarianisms and appealing to the propaganda aspects of morality to justify the indirect—while a departure from the direction that utilitarianism has taken in the past century, stays true utilitarianism’s traditional value theory. Moreover, the extent to which Urmson’s basic duty view of obligation counts as ad hoc is mitigated by a particular relationship that view of obligation arguably bears to the picture of moral requirements (and their justification) that we see in Mill’s *Utilitarianism*. Mill there arguably endorses a similar basic standard for moral obligation that he justifies in similar ways (i.e., by appealing to the utility of publicly adopting the standard).\(^{283}\) Whatever theoretical problems Mill’s standard and its justification may have, Mill cannot be accused of inventing his theory merely to protect utilitarianism from the charge of being unable to accommodate supererogation. The similar view of obligation that emerges on Urmsonian assumptions thus has its difficulties, but it at least can be seen as genuinely utilitarian. And Urmson’s basic duty distinction enables him, unlike Attfield, to view at least some extraordinary, high-stakes acts as supererogatory. For these

\(^{283}\) See J.S. Mill, *Utilitarianism*, (Indianapolis, IN: Hacket, 1979), chapter V.
reasons, Urmson’s view will be the utilitarian representative that we will consider when asking the question of which of the three major moral theories is best able to accommodate supererogation.

5.1.2 The Kantian Accounts

Turning—or returning, depending on one’s perspective—to the Kantian accounts of supererogation, recall that we considered two: Hill’s imperfect-duty account and Jokic’s less-well-known moral luck account. On Hill’s view, the Categorical Imperative (CI) requires that we act/not act in particular ways (perfect duties) and adopt particular principles (imperfect duties). For some of these principles—the principles that Hill calls principles of narrow imperfect duty—principle adoption is consistent with not performing certain external acts every time the possibility presents itself. Thus, a person can adopt the required principle of narrow imperfect duty associated with self-improvement or beneficence without capitalizing on every opportunity for improving oneself or assisting others. Given this feature of principles of narrow imperfect duty and given some passages in the *Metaphysics of Morals* about moral value that Hill takes as implying that a person does what is morally meritorious whenever she performs an (external) act recommended by a principle of imperfect duty, Hill concludes that a person performs an act of supererogation iff she (i) is in a situation in which no perfect or imperfect duty strictly requires that she perform any particular action; (ii) performs an act that a principle of imperfect duty recommends and does so from the right motive, i.e., the motive of duty; and (iii) has performed acts that conform to the relevant principle of imperfect duty on sufficiently many previous occasions. Presumably, what counts as
“sufficient” will depend on what patterns of previous behavior, in the particular individual’s life, are consistent with adoption of the relevant principle.

We have already, in chapter three, discussed some of the problems confronting Hill’s account of supererogation. The question of interest to us at present is whether these problems are, like the ones facing Urmson’s account, mere weaknesses, or whether they parallel the problems facing the Attfield and Portmore accounts in rendering Hill’s account unacceptable as a Kantian account of supererogation. Here, I think the answer is that Hill’s account ultimately fails, but the reasons for the failure of Hill’s account are much more complicated than the reasons for the failure of Attfield’s or even of Portmore’s.

If Portmore’s account fails as a utilitarian account of supererogation and Attfield’s fails as an account of supererogation, the short problem for Hill’s account is that it fails as a Kantian account of supererogation. An account must be capable of explaining why supererogatory acts are so extraordinary as to be above and beyond the call of moral duty. Hill wants to claim that acts not entailed by conformity to the Cl that nonetheless promote the ends of principles of narrow imperfect duty are likewise of moral worth and hence supererogatory. But the value—and hence supererogatory status—of these acts is simply stipulated; Hill never explains how they acquire their value. To be sure, Hill does do something to support his claim; he points to some passages in the Metaphysics of Morals that, on his reading, imply that the relevant acts have moral worth. But even if we grant Hill’s interpretation of the passages (and such a concession is far from trivial, as explained in chapter three), we still need to know how, on a value theory compatible with Kantianism, the acts in question acquire their value.
This is a devastating problem if left unanswered: an account of supererogation that simply states that certain acts that are not required by duty are of positive moral worth and does not explain why they are of positive rather than neutral moral worth is no account at all. And showing that Kant once said something that implies they are of positive moral worth does not provide the needed justification.

Although Hill does not himself explain why the relevant acts that he sees as supererogatory are of value, on a Kantian framework, it would be uncharitable to stop here and conclude that the account is unworkable. We need instead to explore what non-explicit Kantian reasons Hill might have for valuing acts that promote the ends of principles of imperfect duty, at least under the circumstances that Hill stipulates. Hill may be seen as gesturing at either of two sources of value for the acts he considers supererogatory. He requires both that an act of supererogation be “motivated by a sense of duty (or, perhaps, respect for moral reasons)” \(^{284}\) and that it promote an end promoted by a principle of imperfect duty. \(^{285}\) Perhaps Hill sees one of these two features of the relevant acts as giving them positive moral value. The problem with the latter feature is that, however plausible it may be as a utilitarian or Rossian source of moral value, it is implausible as a Kantian source of moral value. Kant wants to separate moral worth from the achievement of or the concern with the achievement of particular ends such as happiness, so the fact that the acts promote the ends that principles of imperfect duty promote will not give the acts positive moral value.


\(^{285}\) Hill, 67.
A more plausible suggestion is that Hill’s requirement that the acts be motivated by duty or “respect for moral reasons”—assuming the latter phrase is cashed out in a way that entails the presence of a good will—is what ensures that the relevant acts are of moral worth. After all, Kant says that the only thing that is good without qualification is a good will, so quite plausibly a good will is the only thing that could account for an action’s having moral value. One problem with this view is that it may be the case, precisely because the good will is the only thing that is good without qualification, that no action can, properly speaking, have moral value in the sense required for the act itself to count as supererogatory (having a moral value over and above what is morally required). This problem was explored in chapter three, but we will sidestep it for now and assume that particular acts can bear the sort of moral worth needed to make it possible for a single act to count as being above and beyond the call of moral duty.

The reason we can sidestep the previous problem is that there is another, more decisive problem facing the present suggestion that the acts that fit Hill’s criteria for supererogation have moral value because they are prompted by a good will. If we take this suggestion and say that the acts in question are prompted by a good will, then given that a good will is a respect for duty, any actions so-prompted will simply be the result of the agent attempting to do what duty requires. So-attempting is not supererogatory and is in fact required of us all, so citing this feature of an act cannot provide the basis for considering it to have moral value beyond that which is required. Given that the cited

---

286 Recall that Hill tells us that for every duty to perform some action, there is a corresponding duty to perform it from the right motive. Remember, too, that the CI has implications for the maxims of our actions, that is our wills, not merely for the external actions that we perform.
value is no more than the value of doing or committing to one’s duty, the cited value can only extend as far as duty extends.

The fact that Hill’s account has difficulty explaining why the acts he considers to be supererogatory are, on a Kantian view of moral value, in fact supererogatory should not be surprising. On Kant’s view, the good will is both the only thing that is good without qualification and something that we all must have on pain of irrationality. Moreover, what separates a good will from a bad will simply is its connection to rationality; a particular will is good iff rational. And by all indications, rationality is for Kant an absolute rather than a threshold notion; hence the fact that rationality can be captured by conformity to the CI, which does not admit of degrees. This leads to the rationality problem that we identified in chapter three: for Kant, morality is a matter of rationality; it appears that a person cannot be “super-rational” so it appears that surpassing moral duty is impossible. Any Kantian account of supererogation must be capable of diffusing this apparent problem. More specifically, any Kantian view that tries to ground the value of supererogation in the value of the good will is going to have to explain why the apparent “supererogation” is not an outcome of the (simple) rationality that is required of all. And any Kantian view that does not ground the value of supererogation in the good will is going to have to explain how, despite Kant’s statements about the exclusive absolute value of the good will, some other appropriate source of value is available. Hill does not attempt to fulfill either of these two tasks; he never addresses, even latently, the rationality problem. It is thus not surprising that he is

---

The question of whether or not a person is acting in such a way that she can at the same time will that the maxim of her action be a universal law of nature has either an affirmative or negative answer and does not in either case admit of degrees.
unable to account for the supererogatory status of the acts he wants to call—based on the
criteria he identifies—supererogatory and that he has not done enough to develop a
legitimate Kantian account of supererogation.

Notably, Jokic’s account, though it has significant shortcomings, does not have
the fatal one that Hill’s does. Jokic has a straightforward response to the rationality
problem: he admits that supererogatory actions are not morally praiseworthy, so he does
not need to address the question of how supererogation can be, in some sense, “super-
rational,” or explain how moral praiseworthiness, for Kant, could fail to be a function of
rationality. The challenge for Jokic becomes that of explaining the non-moral sense in
which purported acts of supererogation are praiseworthy and therefore qualify as
supererogatory rather than merely “morally acceptable.” Jokic’s solution, recall, is to say
that supererogatory acts are praiseworthy in a sense that is similar to but distinct from
moral praiseworthiness. An act is morally praiseworthy iff it is prompted by a motive
whose presence guarantees the resultant act’s conformity with the CI: no act could be
prompted by such a motive and fail to conform to the CI. Noting this feature of morally
praiseworthy acts (and the fact that Kant himself at one point appears to cite this feature
of such acts as the basis for their moral worth), Jokic reasons that a similar value—nearly
moral praiseworthiness—is had by an act iff the act’s motive makes highly likely, but
does not guarantee, the act’s conformity with the CI. This is the sort of value that
supererogatory acts have. Jokic thus has a view on which he can account, in arguably
Kantian terms, for the value of supererogatory acts.

Given that Jokic’s account has the previous mark in its favor, the question is
whether the account faces any other devastating objections. We saw in chapter three that
its probability condition on motive may not give it the verdicts on particular cases that Jokic wants it to have, but we have not yet seen any reason to think that it is so problem-riddled on this front that, like Attfield’s account, it cannot at all be said to accommodate the moral phenomenon that we are interested in accommodating. We also saw in chapter three that Jokic’s account departs in important ways from the traditional picture of supererogation and supererogatory actions: it has strange implications for the value of supererogatory acts relative to the value of morally praiseworthy acts, and it cannot recognize moral value in supererogation at all. On top of this, Jokic’s understanding of moral obligation arguably involves a departure from Kant’s own understanding of obligation (a point that will be addressed later in this chapter). The upshot of all of these problems, though, is only that Jokic’s account may have some strange things to say about Kantianism and supererogation; it is not that Jokic’s account fails to qualify as a Kantian account of supererogation. We will thus consider Jokic’s account as our representative Kantian account of supererogation, exploring the difficulties previously mentioned to see how strong it is as a Kantian account of supererogation.

5.1.3 The Virtue Ethics Accounts

Turning finally to moderate virtue ethics accounts of supererogation, it will not come as a surprise that the only account I wish to investigate in further detail and consider alongside the Urmson and Jokic accounts is my own. The only other virtue ethics account of supererogation that has been defended in the literature is Kawall’s, and we saw that that account either fails as a virtue ethics account of supererogation or else fails as an account of supererogation. Kawall’s account says that an act is supererogatory
iff an epistemically idealized virtuous observer would judge that it is. If the ontological (not epistemic) basis of the value of the acts in question is the judgment of the virtuous observer, then the account in question is not a form of virtue ethics but subjectivism, for judgments rather than virtues are its ultimate bearers of value. If the judgments of the hypothetical virtuous observers are not the ontological basis of the value of the acts in question, then we do not have subjectivism, but we have not been told why the acts in question are of positive moral worth (or any other sort of worth). Kawall has told us a characteristic that all supererogatory acts share (virtuous observers consider them supererogatory), but he has not told us what makes them supererogatory, which is of course a requirement of any satisfactory account of supererogation. Kawall’s account is thus unacceptable as a virtue ethics account of supererogation, and we need not concern ourselves with it any further.

We explored other ways, rooted in virtue ethics, of trying to accommodate supererogation, but the only way that yielded fruit is the one that led to my Optional Virtue Account of Supererogation (OV). OV is both a genuine account of supererogation and one that is sufficiently rooted in standard virtue ethics commitments: blame and obligation are grounded by the vices; the positive moral worth of supererogatory acts is grounded in the virtues; and the account as a whole takes seriously the virtue ethics mantra that it is not just what is done but how it is done that is relevant to moral assessment. While there are questions about how well the account fares when applied to some particular cases (which we will consider below), we can say now that the account is positioned to see many morally extraordinary ways of acting—precisely because these ways of acting involve extraordinary virtue—as supererogatory. So, OV can be seen as
positioned to account for the supererogatory status of at least some extraordinary acts and to do so while remaining true to central moderate virtue ethics commitments. It is thus at least a viable virtue ethics account of supererogation, and we will accordingly continue to consider it.

5.2 A Side-by-Side Comparison

The three accounts that we have found worthy of comparison are Urmson’s basic duty account (hereafter BD), Jokic’s near-moral-worth account (hereafter NMW), and my optional virtue account (OV, as before). For ease of reference, it may help to lay out, in one place, the necessary and sufficient conditions of supererogation, according to each view. On Urmson’s BD, an act $a$ will be supererogatory for an agent $s$ in circumstances $c$ iff

$$a \text{ brings about a state of affairs } x \text{ that is superior to some other state of affairs } y \text{ that } s \text{ could have brought about in } c \text{ without violating any rule of basic duty } r.$$ 

In turn, a rule $r$ will be a rule of basic duty iff

(i) $r$’s regular violation would lead to the sumnum malum and

(ii) there is no alternative\(^{288}\) rule $r^*$ such that $r^*$ is less stringent than $r$ and $r^*$’s regular violation would lead to the sumnum malum.

For Jokic’s NMW, an act $a$ is supererogatory for an agent $s$ in circumstances $c$ iff

(i) performing $a$ in $c$ is neither required by nor forbidden by the CI and

\(^{288}\) By “alternative rule” I mean a rule that is similar in scope (so that if, for instance, a given rule concerns lying, its alternative rules do as well). I am assuming that this idea is sufficiently intuitive and clear to make this reference to it appropriate; if it isn’t, then Urmson’s theory has a new difficulty that it must address.
(ii) \( s \)'s motive (for \( a \) in \( c \)) \( m \) is such that, whenever \( m \) is present and issues in an action \( a^* \), the probability, given the presence of \( m \), that \( a^* \) violates the CI is extremely low.

In the above context, to say that the relevant probability is “extremely low” is to say that it is almost zero and lower than even the probability that the motive of love (or some other motive of direct inclination), if present, prompts an act that violates the CI. Finally, my OV says that a particular intentional activity \( a \) performed by an agent \( s \) in circumstances \( c \) is supererogatory iff

(i) \( s \)'s \( a \)-ing in \( c \), manifests no vice;

(ii) \( s \)'s \( a \)-ing in \( c \) manifests some virtue \( v \) to some degree \( d \); and

(iii) \( s \)'s acting intelligibly and manifesting no vice in \( c \) does not imply \( s \)'s acting in \( c \) in a way that manifests at least one virtue to degree \( d \)

In evaluating the above three accounts, we will be considering how they fare relative to one another on the following criteria: fit with home moral theory (e.g., how utilitarian is BD?); sensitivity (is the account able to accommodate the central cases)\(^{289} \); specificity (is the account too easily satisfied, classifying wrong or neutral acts as supererogatory?); and any remaining advantages or problems, theoretical or practical.

5.2.1 Fit with Moral Theory

Let us consider, first, how well the accounts fit with their respective home moral theories. Jokic’s NMW involves the greatest departure from its home moral theory, for Jokic’s account involves the invention of an entirely new category of value, never hinted

\(^{289} \) As noted in chapter one, some allowances will be made for the different value systems at the hearts of the different normative theories.
at in Kant’s ethical theory or recognized by later Kantians, that makes certain non-required, permissible acts supererogatory. Jokic’s way of tying his “near moral value” to Kantian commitments is ingenious, to be sure, but it is precisely because the tie requires such ingenuity that the resultant picture of value involves such a significant departure from traditional Kantianism, even though NMW manages to preserve a legitimate claim to qualifying as Kantian.\textsuperscript{290} Urmson’s BD and my OV, meanwhile, adhere to the traditional value theories of their home moral theories. They do depart from their home moral theories on the issue of moral obligation by endorsing more relaxed standards of moral obligation, but the new standards, though significantly lower than what is typical for their home theories, can still largely be seen as relaxed versions of their home theories’ typical standards rather than as completely new ways of conceiving of duty. In this respect, we might say that the relaxed standards differ from the home theories’ more

\textsuperscript{290} Jokic’s view of moral obligation also involves a departure from Kant’s stated position on obligation, although this departure is not as significant as the value departure. Jokic sees obligation as always attaching to acts: a particular external act is either required, permitted, or forbidden by the CI. Acts that are permitted are ones that, when prompted by what I previously called a low-luck motive, are supererogatory. If we look at Kant’s CI, it too appears to apply to acts, for it tells us to “act only” in particular ways. Upon closer inspection, however, we find that the limits on our actions are entirely rooted in the principles or maxims of our actions and not at all in some other external feature of them. As explained in chapter four, it is not anything about the action “uttering a known falsehood” that makes lying always fail Kant’s CI test; it is a fact about the maxims involved when people lie. So, Kantian obligation, strictly speaking, applies to maxims rather than external acts. The limit on maxims it provides, however, strongly influences the actions that are possible.

One may wonder why I press Hill so much for departing from Kant on the issue of act versus principle-adoption when I do not do the same to Jokic. The reason has to do with the goals and prior commitments of Hill’s and Jokic’s respective accounts. Hill is committed to principle adoption being what imperfect duties require but act performance being what generates merit, in the case of imperfect duty. Jokic is not committed to the former so his endorsement of the latter is less problematic. Jokic’s focus on act rather than principle does constitute a departure from Kant, in my view, but even here Jokic can cite precedent for the departure, not in the work of Kant, but in popularized “Kantianism.” Hill cannot appeal to this precedent, however, in supporting his theory because Hill wants a view rooted in Kant’s work (specifically, the Metaphysics of Morals), and indeed this is the only way in which he finds support for crucial aspects of HAS. If Hill thus abandons the Kant of the Metaphysics of Morals for the popularized Kant, he loses essential textual support for HAS.
typical standards in extent but not in kind. Moreover, Urmson can point to Mill and I to Hursthouse (and, to a less extent, to Slote and Swanton) for evidence that these more relaxed pictures of utilitarian morality and virtue ethics morality are not without precedent within the respective home moral theories.

Although Jokic’s NMW is thus further from Kantianism than Urmson’s BD is from utilitarianism or OV is from moderate virtue ethics, BD and OV initially look to be equidistant from their home theories. For the most significant way each departs is the same in each case; both BD and OV relax their home theory’s traditional standard of moral obligation. In fact, however, the departure for BD is greater than the departure for OV. This is because the relaxed moral standard that is latent in OV is more natural an outgrowth of central virtue ethics commitments than is the standard of basic duty espoused in BD. OV’s standard—that vice be avoided rather than virtue positively manifested—is non-arbitrary insofar as nearly every ethic of virtue recognizes a morally significant, definite distinction between virtue and vice. The same cannot be said for BD’s distinction between the *summum malum* and all states of affairs preferable to it.

And given the fact that the virtues are positive character traits and the vices negative, it is not at all surprising for a moderate virtue ethics to ground guilt in the vices and moral praiseworthiness in the virtues. Urmson’s standard for basic duty is not so natural an outgrowth of utilitarianism, which is why he has such difficulty defending his standard, as we saw in chapter two.

An additional point in OV’s favor on the “fit with moral theory” criterion is that there is a way in which OV’s relaxed moral standard, though different from the most commonly articulated virtue ethics standard for moral obligation, is actually closer to the
heart of virtue ethics than is the more commonly articulated standard. The most commonly articulated virtue ethics standard makes moral obligation a simple matter of performing whatever action a similarly-situated virtuous agent would perform; for OV, by contrast, moral obligation is a matter of character trait manifestation. The latter standard better reflects the traditional virtue ethics emphasis on character over act type, on how one acts and not merely what action one performs. So, although OV’s relaxed standard pulls it away, in some ways, from mainstream moderate virtue ethics, in another way OV’s new standard better embodies core virtue ethics commitments than do the mainstream versions of the theory! (And, as we saw in chapter four, OV’s standard better fits the logic of some particular virtues, e.g., generosity and magnanimity, than the usual ideal virtuous agent standard of obligation does). BD has no such mitigating feature of its way of relaxing duty; there is no way in which BD’s relaxed moral standard is more in line with traditional utilitarian concerns (e.g., the import of utility) than the traditional maximizing requirement is. The upshot of all of these considerations is that OV is more in line with traditional virtue ethics and its core commitments than BD is with traditional utilitarianism, and both are more in line with their respective home theories than Jokic’s NMW is with traditional Kantianism.

5.2.2 Sensitivity to the Central Cases

When it comes to accommodating the central cases, the sorts of cases for which we set out to account, OV also has the advantage over BD and NMW. The five central cases, recall, are: Grenade, Torture, Kocheleff, Babette, and Candlesticks. As we saw in chapter two, Grenade and Kocheleff are the two that a utilitarian would most likely want
to accommodate, but BD does not easily accommodate either. It cannot accommodate *Grenade* if we stipulate that one of the beneficiaries of the soldier’s action would have jumped on the grenade had the soldier run away, for BD makes no distinction between the moral worth of a self-sacrificial action and that of an other-sacrificial action. BD cannot see Kocheleff’s work in Africa as supererogatory because, given the circumstances and Kocheleff’s skills, working to alleviate the AIDS crisis there is precisely the sort of thing that people such as Kocheleff must do if the *summum malum* is to be avoided. BD is also unable to accommodate *Torture*, since the hero’s refusal to denounce his country yields no more utility than any alternative action yields. Precisely for this reason, however, *Torture* is plausibly the sort of case that utilitarians will not be interested in accommodating, so we should not penalize BD too harshly—and perhaps not at all—for not classifying *Torture* as supererogatory.

The two central cases that plausibly count as supererogatory, on BD, are *Babette* and *Candlesticks*. Adhering to the rules needed to avoid the *summum malum* probably has few implications for what one does if one, in circumstances like Babette’s, suddenly acquires a large sum of money, and the Bishop’s adhering to the rules needed to avoid the *summum malum* likewise does not imply that he forgive Jean Valjean in the way that he does. Thanks to this latitude that basic duty allows Babette and the Bishop, each has the opportunity to act in a morally acceptable way that yields less utility than the path chosen. (Babette, to be sure, may also have a wide array of options that will yield much more utility, for she could, say, give a very nice but less extravagant dinner for the entire

\[\text{291} \] We will ignore the possibility that Babette could have used the money to save a number of people from dire poverty and assume that, although Babette could have given the money to charity, the situation of any potential beneficiaries was not so dire as to make such a gift obligatory.
village). So, what each does is arguably supererogatory on BD. The problem for BD with respect to *Candlesticks* and *Babette* is that it cannot account for the supererogatory status of either case in the right sort of way. What is extraordinary about both Babette’s and Myriel’s actions is a kind of extravagant, redemptive love and mercy that each involves, albeit in very different ways, not the fact that the extravagance promotes a measure of utility. These two cases would have been tragic, but still involved supererogation (albeit perhaps to a lesser extent), had, in the case of the feast, the guests carried out their decided game of stoicism through the full evening and had, in the case of the Bishop’s gift, Valjean’s character not subsequently changed as it did.\(^{292}\)

Assessing what NMW has to say about the central cases is somewhat more difficult. *Torture* will not involve supererogation, on NMW, because refraining from renouncing one’s country is required no matter what the circumstances, on Kant’s moral theory. *Candlesticks* and *Grenade* at least both obviously will not come out either as obligatory or wrong, on Kantianism, making them eligible for supererogatory status on NMW if the proper motive is present. Given how Kant argues for duties to help one another, it is less clear in both *Babette* and *Kocheleff* that the agent is not obliged to do something like what she/he does, but we can at least say that the popularized understanding of imperfect duty would allow Jokic to insist that neither Babette nor Kocheleff is obligated to act as she/he does. So, for now, we will assume that *Babette* and *Kocheleff* pass the Kantian obligation test; the question then becomes whether they (and *Candlesticks* and *Grenade*) pass the NMW (near moral) value test.

\(^{292}\) And neither such result would have been unlikely. Such a result would have happened in *Babette* had the unexpected visitor not shown up. And plenty of people shown the sort of grace that Valjean is shown are not transformed as a result.
We have, then, at most four candidate cases for NMW. What determines which ones (if any) are supererogatory will be the agents’ motives. In Babette, I would argue, the motive is demonstrably not the sort that would allow for Jokic-style supererogation. Babette chooses to give the feast primarily out of her particular love for and gratitude toward the sisters. Although honoring the sisters’ father is the occasion for the feast, Babette is giving it primarily for the sake of the sisters and not primarily for the sake of anyone else, including the deceased father, whom she has never met. We can thus say that Babette’s motive is highly “particularized” with regard to her intended beneficiaries, meaning that the actual motive she has is not the sort that would prompt her to do this sort of thing for anyone other than the sisters. The very fact that Babette’s motive is so particularized is what makes it implausible that it is the sort of low-luck motive that it needs to be if giving the feast is to count as supererogatory, on Jokic’s account. For if inclination requires too much luck for it to prompt a supererogatory act, certainly particularized love or gratitude will have a problem that is at least as serious. Babette, far from acting out of anything like universal respect for humanity, is acting out of a love that she has for only two people. A love this restricted is almost as likely to prompt acts that violate duty as is the only slightly more restrictive self-love. Babette’s motive in giving the feast thus should not be seen as the sort of low-luck motive that is needed on NMW.

______________

293 This is not to say that either type of love has a high likelihood of prompting immoral actions; the point is that we do not have a sufficiently low-luck motive, not that we have a high-luck motive.

294 Despite the fact that Babette’s primary motive in giving the feast is particularized, I will grant that some of the things Babette does as she gives the feast do not involve particularized motives. Specifically, when one of the guests arrives with her driver, Babette invites the driver into the kitchen to eat the same fare that the other guests are eating; this is arguably not done for the sake of the sisters, and it
For the remaining three cases, what the agent’s motive is—and whether it is a sufficiently low-luck motive—will depend at least in part on how we choose to cash out the cases. We could stipulate that Myriel, the soldier, and Kocheleff are motivated by compassion, particularized love, or some form of indirect inclination, in which case the action, per Jokic, will not be supererogatory. Alternatively, we could say that each agent’s motive is remarkably similar to, but distinct from, the motive of duty and requires very little luck to generate a morally acceptable action. Perhaps each agent is motivated by some sort of extreme conscientious-like concern with putting other humans’ needs ahead of his own, (a concern that needn’t imply, of course, that the agent ignores his own needs). If the latter sort of motive were present (which is consistent with *Grenade*, *Kocheleff*, and perhaps even *Candlesticks*), the action would be supererogatory, on the Jokic view. Since *Grenade*, *Kocheleff*, and *Candlesticks* all count as central cases even when we specify that the agent’s primary motive is love, what this means for Jokic’s account is that, at best, it accommodates three of the central cases (*Grenade*, *Kocheleff*, and *Candlesticks*), but the accommodation is only partial, since it leaves out versions of the cases in which the agent is motivated by love or some particularized regard. NMW may thus accommodate some of the central cases, but the accommodation is limited.

If Jokic’s view can only accommodate some forms of *Grenade*, *Kocheleff*, and *Candlesticks*, and Urmson’s can only accommodate *Babette* (for the wrong reasons),

---

295 The way in which the narrator describes Myriel and his actions make it difficult to see Myriel’s primary motivation as being anything other than compassion (or something like it).
Candlesticks (for the wrong reasons), and some forms of Grenade, what can we say about OV’s verdicts on the central cases? We have already shown that OV will recognize most forms of Kocheleff as supererogatory, the exceptions will be the unusual cases in which Kocheleff has some vice that so infects his project in South Africa that it is accurate to say that his engagement in the project as a whole manifests some vice, even if it also manifests a number of virtues. It is hard to see how a similar problem could apply to the other central cases; whatever vices the agents in Torture, Grenade, Candlesticks, or Babette may have, these vices are not manifested in the intentional activities we see in the cases.

OV will clearly accommodate Torture, the one case that neither BD nor NMW could accommodate, for one can, under extreme torture, renounce one’s country and pledge allegiance to terrorists without manifesting any positive vice. (Recall that vices are character states, not principles, so the mere fact that an agent renounces her country and pledges allegiance to terrorists does not entail that she has and manifests some vice, such as infidelity; if the torture is sufficiently severe, renouncing will be a possible action of a person who lacks the relevant vice, even though it may be correct to say that the action violates the principle of fidelity). In addition, the agent in Torture does manifest a

---

296 That is, assuming that Kocheleff satisfies condition (iii). This should be obvious, though; clearly Kocheleff could have permanently moved to the Netherlands from Burundi and worked as a physician there without manifesting vice and without manifesting virtue to the degree he manifests it in his work in South Africa.

297 The soldier, for instance, even if he has the vice of pride, is not manifesting that vice, and if he were, his jumping on the grenade would cease to be supererogatory, although uninformed observers who do not realize the soldier’s motive of pride would of course never be aware of this. More generally, if we alter any of the remaining four cases so that the agent is manifesting positive vice, the cases cease to be central, and it is at least not clear that they continue to involve supererogation.
positive virtue—fidelity—to a high degree. So, the intentional activity of the agent—standing up to terrorists in the face of torture—will qualify as (highly) supererogatory.

*Candlesticks, Babette, and Grenade* all manifest positive virtue (at the least, magnificence in the case of *Candlesticks*; generosity and the love of friendship in *Babette*; and generosity and loyalty in *Grenade*). They also do not involve the manifestation of vice. The real question for the virtue ethicist will be whether the agents could have acted differently, but non-viciously, without manifesting the high degree of virtue that they do manifest. For each of these three cases, the answer will be affirmative. Babette could have used her lottery winnings to return to France and reestablish her life there, leaving a small gift for the sisters, without manifesting vice (hence the surprise of the sisters that she does not do just that) or a high degree of virtue. In *Candlesticks*, although it may be that the Bishop would have acted callously (and hence viciously) had he simply turned Jean Valjean over to the police for theft, the Bishop still had alternative courses of action that would have been neither callous, nor otherwise vicious, nor (as) magnanimous, nor as impressive on the score of some other virtue. Myriel could have, for instance, given the same semblance of having given Valjean the silverware (prior to Valjean’s departing with it) without giving Valjean the candlesticks as well; or, Myriel could have waited for the police to leave and then insisted that Valjean return the items. In *Grenade*, the soldier has two non-vicious intelligible options: run to save his own life (in a non-vicious way, e.g., out of self-love, but not pride, selfishness, or hatred of his comrades) or jump on the grenade (from the right motive, e.g., out of loyalty to his comrades). Since the former would not have involved the level of virtue that the latter does (in the form of high degrees of generosity and loyalty), the soldier’s activity turns
out to be supererogatory, on OV. OV can thus accommodate all five of the central cases, versions of Kocheleff in which Kocheleff has a serious vice that so infects his project that it overall manifests a vice excepted. And we have seen reason (in chapter four) to doubt that these (sufficiently) vice-infected versions continue to qualify as central cases.

It is tempting here to object that while the agents in Candlesticks, Babette, and Grenade all manifest a high degree of virtue, non-vicious alternative actions would have manifested equally high degrees of virtue; the relevant virtue manifested to a high degree would just have been different. To be specific, had Babette (non-viciously) moved to France and set up a new life for herself, had Myriel (non-viciously) kept the candlesticks for himself, and had the solder let one of his comrades jump on the grenade, each of our agents would have manifested high degrees of the virtue of self-love. So, the objection goes, their actual actions are no higher on the score of virtue than these less self-sacrificial alternatives; per condition (iii) of OV, their actual self-sacrificial actions are thus not supererogatory.

Two responses to the self-love objection are in order. First, it is not at all obvious that every licit (non-vicious) alternative our agents have would have manifested as high a degree of self-love as the degree of the virtue actually manifested. Because the candlesticks are so much more valuable to Valjean, in the context, than they are to the Bishop (even though he values them highly), it is hard to say that the Bishop’s keeping them would have manifested as high degree of virtue (in the form of self-love) as his giving them away manifests (in the form of generosity). And in the case of Babette, it is not as if, in staying and giving the feast rather than starting a new life in France, she is choosing between self-love and friendship. Precisely because Babette loves the sisters as
she does staying with them does not require that she sacrifice her own overall good. She can be happy and manifest as high a degree of self-love if she stays as she would if she were to go, but only by staying and giving the feast (or doing something else for the sisters) will she manifest such high degrees of generosity and loyalty.

The second response to the self-love objection is more general because it has to do with an asymmetry in self- and other-directed virtues implied—although the implication is not usually acknowledged—by many prominent forms of virtue ethics. On most of the prominent forms of virtue ethics, the standards for the virtue of self-love, if self-love is recognized as a virtue, are higher than the standards for the virtue of generosity (and most other-directed virtues) in the following sense: it takes more for a self-benefiting act to qualify as exemplifying a moderate or high degree of the virtue of self-love than it does for an other-benefiting act to qualify as exemplifying an equally high degree of most other-directed virtues, such as generosity. This double standard likely obtains on Slote’s view of virtues as admirable character traits (due to self-other asymmetries in the admirability of various character traits), but its truth is not only an outgrowth of admirability theories like Slote’s. It also is an outgrowth of the very common model of the virtues as correctives of human nature. Hursthouse explicitly endorses this model, and the corrective model is also true on a standard reading of Aristotelian and Thomistic ethics that see the virtues as taking human-nature-as-it-is and guiding it so that it becomes human-nature-fulfilling-its-telos. On such a picture, the


virtues are corrective in that they are not simply traits of the flourishing human but traits of the flourishing human that humans lack in their natural, starting state. Because pursuing self-interest comes more naturally to humans than serving others’ interests, versions of virtue ethics that see virtues as corrective will thus have asymmetrical standards for self- and other-directed virtues. They may not even recognize self-love as a virtue at all, for the self-directed virtues recognized will be limited to what does not come naturally to humans. The self-love objection we have been considering thus dissolves, at least for most forms of moderate virtue ethics.

5.2.3 Specificity

Even though OV is better at accommodating the central cases than NMW or BD, we will still have the question of whether the explanation for its ease at accommodation is that its standard for supererogation is too low, allowing in too many cases. We have already seen where BD stands on this issue. Because contemporary wealthy societies are not at high risk for devolving into a Hobbesian state of nature or even a state as bad (in terms of level of utility) in which people live in relative peace but face the sorts of health, pain, and safety problems people tended to face until relatively recent developments in medicine, it turns out that members of these societies act supererogatorily on a fairly regular basis. For insofar as present human behavior is responsible for the high standard of living we now enjoy, most people’s everyday ways of living are at least partially

300 This corrective aspect of the virtues may explain why Aquinas does not even seem to recognize a virtue of self-love, though he does recognize self-directed virtues such as prudence and moderation. The challenge for humans in their natural (non-fulfilled) state is less that of trying to do what serves their own good and more a matter of understanding in what their good consists and gaining the tools needed to pursue it.
responsible for keeping us at this relatively (by historical standards) high state. Because people only lie once or twice a day, on average, but not twenty times, because they are often selfish but do not usually resort to thievery, we enjoy a high level of utility. It cannot be denied that developments in medicine and technology have been crucial in making this high standard of living possible for us; but this does not change the fact that now that we have the technology, little is required of us to avoid the *summum* malum, even interpreting that level of society fairly liberally. Most people regularly surpass these (now) low basic duty requirements, and in this way most people (now) regularly act in a supererogatory manner, on BD. Given the way in which technology insulates us from the *summum malum*,\(^{301}\) BD thus turns out to be far too liberal in what it recognizes as supererogatory.

In the case of Jokic’s account, we do not need to worry about morally wrong acts turning out to be supererogatory on NMW and in this way providing a problem for Jokic. For we have already said that, in assessing where the various moral theories stand on the supererogation question, we are ignoring traditional objections to the theories (e.g., the objection that utilitarianism allows the killing of the innocent in dubious circumstances). In the case of Kantianism, these tabled objections will include traditional objections that see Kant’s account of obligation as too lenient, whether generally or in particular cases. Since Jokic preserves a typical Kantian account of moral duty (and in this way differs from the way Urmson approaches utilitarianism or I approach virtue ethics, since both

---

\(^{301}\) We tend to think of technology as enabling us to trade labor for leisure. Interestingly, if Urmson’s way of understanding duty is correct, technology enables us to trade *moral* labor for leisure. It helps insulate us from the *summum malum*, and the moral burden on individuals accordingly becomes lower.
Urmson and I alter our respective home theories’ usual conception of obligation, and because NMW precludes morally wrong (from the perspective of Kant) acts from being supererogatory, any argument that NMW sees wrong acts as supererogatory would be rooted in the sorts of general worries about Kantian morality being too lenient that we are setting those sorts aside for the present discussion.

What will concern us is whether NMW incorrectly classifies any morally neutral acts as supererogatory, and it turns out to be much more difficult to answer this question than it was to assess whether or not Urmson’s account of supererogation was too liberal. The problem is that it is difficult to discern what sorts of motives are low-luck motives and when they are present in a particular case. We know (from Jokic) that such motives as self-love, indirect inclination, and love are not sufficiently low-luck to render so-motivated acts supererogatory. What sort of motive would be even more likely than these to prompt morally acceptable acts without being equivalent to the motive of duty? Jokic does not tell us, but if such a motive has to require less luck than even love, the only specific motives that I can see as being capable of doing the trick will be motives that are very similar to the duty motive not merely in their low-luck status but in their content or first-person “feel” as well: an extreme conscientious concern with something other than, but very close to, duty, perhaps, such as an obsession with benefiting humanity in general (which we will distinguish from respect for humanity, which arguably is identical to the duty motive, and love of humanity, which we already know is not a sufficiently low-luck motive).

Will the acts that such a motive prompts be the sort that are intuitively supererogatory, at least when Kantian duty neither forbids nor requires the prompted act?
In situations in which Kantian duty requires nothing of a person, the person has an obsession with benefitting humanity on which she acts, and in doing so she does not violate Kantian duty, the person’s obsession will almost always produce an act that helps either self or others. Many, but not all, of the self-regarding cases will not be intuitively supererogatory, which is a mark against NMW, but these cases will not be so common as to make NMW’s classification of all such cases as supererogatory deeply problematic. As for the relevant cases in which the obsession prompts other-benefitting activity, classifying (most of) these cases as supererogatory will not be problematic at all; these are precisely the sorts of cases that so many people see as, intuitively, beyond duty.

There is, however, one problem with the way NMW classifies the relevant other-benefiting cases as supererogatory. They are so-classified based on the agent’s motive. But the obsessive motive under discussion can occur in both fairly mundane cases of benefiting others (as when I grant you a small favor because I am obsessed with satisfying everyone, to the extent that I can) and in more extraordinary cases (as when I donate my entire multi-million dollar inheritance to Oxfam because I am obsessed with saving as many humans from the ravages of poverty as possible). It does not hurt NMW on the sensitivity score that it classifies both of these cases as supererogatory and beyond duty; plenty of people want to so-classify even trivial favors as beyond duty. It does, however, hurt NMW that (because the same motive is involved) it classifies them as of equal near-moral value and hence equally supererogatory. If an account of supererogation is going to classify small favors as supererogatory, it needs to do so in a way that distinguishes such near-trivial acts of supererogation from the extraordinary ones that are the paradigms of the category.
At first blush, OV appears positioned to score quite poorly on the specificity criterion, for manifesting virtue to some extent and otherwise avoiding vice initially seems a standard easily satisfied. Here, it needs to be remembered, as I argued in chapter four, that the latter requirement is not as easily met as may at first appear. The list of the vices is extensive: sloth, crudeness, callousness, selfishness, intemperance, shortsightedness, rudeness, pride, and vanity are just the beginning. Since most of us suffer from a number of these vices, acting in a way that manifests none of them and also manifests a positive virtue (and not merely by virtue of the absence of these vices, as honesty might be manifested simply because a speaker manages to avoid dishonesty by uttering a true statement) will not be nearly as easy or common as first appears. It is true that we can have “small” cases of supererogation on my view, cases in which what a person does is supererogatory simply because he manifests a small degree of virtue and avoids vice, in a particular situation in which avoiding the manifestation of vice is fairly easy. But in this regard, OV is no worse than NMW, which also recognizes “small” cases, such as near trivial favors and easy acts of self-service, as supererogatory. Indeed, OV is superior to NMW on this score because, unlike NMW, OV’s standard gives it a threshold account of supererogation, so that it does not have nearly to equate the total value of extraordinary acts with that of more mundane supererogatory acts. And both OV and NMW are better positioned than BD on this issue. For reasons we have seen, in contemporary wealthy societies, BD’s standard for basic duty is so low—the sort of pattern of behavior needed to keep a well-insulated society from reverting to a few steps above subsistence level—that people can more than satisfy it despite acting rather (but
not extraordinarily) selfishly, and such behavior will never come out as supererogatory on OV or NMW.

5.2.4 Additional Considerations

Our final criterion for evaluating the three accounts is the not terribly informative "miscellaneous" criterion, which is really just a check on what other special advantages or disadvantages the individual accounts have as accounts of supererogation for partisans of their respective moral theories. BD, we have already seen (in chapter two), has a special theoretical difficulty. Because it blends act and rule utilitarianisms, setting moral obligation in terms of the latter and moral value in terms of the former, it ends up, paradoxically, classifying some morally wrong acts as of greater moral worth than acts that satisfy basic moral duty.

Despite the above problem, BD does have one theoretical advantage over NMW and OV. A very intuitive way of conceiving of supererogation is in terms of a moral accounting metaphor: we all owe a certain amount (morally), and supererogation is a matter of giving—whether to particular other people, society in general, or the universe—more than this amount. To think that this accounting metaphor exhausts what we can say about supererogation, what it is, and why it captures our moral imagination is, I think, naïve. The last section of this chapter highlights one way in which a view of supererogation that equates supererogation with moral extra or advance payment misunderstands some of the most striking examples of supererogation. Even if the moral accounting metaphor is inadequate to the moral phenomenon of supererogation, though, the thought that the metaphor expresses truth, that there is a sense in which, whenever a
person acts in a supererogatory manner, she has given more morally than she is obliged to give, remains an intuitive one. And it is an intuition that BD clearly affirms.

OV will be seen by some as less satisfactory than BD on this score. OV is consistent with the moral accounting metaphor expressing truth. OV recognizes a minimal core moral requirement (to avoid vice) and way of doing more, morally, than this requirement (to manifest virtue). But whereas BD can point to something concrete—utility—that a person “overpays” to the universe (or, perhaps more correctly, to some sentient beings in that universe) when she performs an act of supererogation, OV’s source of excess moral worth—the manifestation of virtue—is less satisfying when considered as a form of payment to the universe or to anyone else we might morally “owe” payment. So, those who want to take the moral accounting metaphor of supererogation more literally, if not entirely literally, will thus find BD more satisfactory an account of supererogation than OV.

NMW, meanwhile, does not capture the moral accounting metaphor at all. On NMW, supererogation is not moral over payment; supererogatory acts have no moral value whatsoever. This inability of NMW to capture the moral accounting metaphor is related to some of the counterintuitive implications, already noted, that NMW has for the relative values of various acts. To recall the discussion in chapter three, NMW implies that (i) no supererogatory act that is prompted by a given (low-luck) motive is superior to an obligatory act prompted by the same motive, (ii) while one supererogatory act may have a greater worth than another, the difference between any two such acts will be small on an absolute scale, and (iii) any obligatory act performed out of duty is (both morally and absolutely) superior to any supererogatory act. It is the third of these implications
that makes it difficult for NMW to affirm the moral accounting metaphor of supererogation as moral overpayment, and it is likewise the implication that, for reasons we have seen (in chapter three), makes it difficult for NMW to affirm, with a straight face, that supererogatory acts are “above and beyond the call of moral duty.”

Despite the above counterintuitive aspects of NMW, NMW does have one advantage over BD. The advantage is one that it shares with OV, and it is the last “miscellaneous” advantage or disadvantage of the accounts that I think is worth here discussing: NMW and OV both pay ample attention to motive and the agent’s psychological state, and intuitively these factors are highly relevant to the value of supererogation and indeed the status of an act as supererogatory. For one way in which the moral accounting metaphor cannot fully capture all that is of moral significance in the supererogatory is precisely that it is not just the acts and their consequences, but the people doing the acts, the saints and heroes, with their ways of seeing the world and other people, that capture our attention and make the supererogatory an important part of our ethical world. BD is blind to this aspect of the supererogatory, although this blindness is admittedly mitigated somewhat by the fact that it is utilitarian account, and utilitarianism has its own independent reasons for downplaying the significance of motive.

5.3 A Victory for Virtue Ethics

It should be becoming clear that, by accepting OV, some of the most prominent forms of virtue ethics can make good sense of supererogation, better sense than we have seen utilitarians or Kantians make. In comparing the best accounts we could find that have been offered on behalf of utilitarianism, Kantianism, and virtue ethics (BD, NMW,
and OV, respectively), OV was found to fit better with virtue ethics than NMW fits with Kantianism or (even) BD fits with utilitarianism. Moving beyond the issue of fit, OV also provides a better account of supererogation than BD or NMW does. OV beats out BD on a number of criteria: ability to accommodate the central cases, success in weeding out non-supererogatory cases (i.e., low false-positive rate), and (as a slight advantage, given utilitarianism’s reasons for putting little stock in motives) recognition of the relevance that motive and character have to the supererogatory and moral saints and heroes. The only advantage BD has over OV is that it better reflects the intuition that supererogation involves ‘paying’ more than what one morally owes. This advantage, however, is not significant enough to outweigh the advantages that OV has over BD. For OV can also see supererogation as giving more than what is owed—it is just that the form of payment, manifested virtue, is less concrete than the form of payment, utility, involved in BD. Moreover, BD has its own theoretical difficulty, the problem of implying that some morally wrong acts are morally better than some morally neutral (i.e., neither wrong nor supererogatory) acts.

NMW, meanwhile, is, I would argue, even worse off than BD. NMW does a good job of not seeing supererogation in the wrong places (as BD sees it whenever an agent exceeds Urmson’s very low standard for basic duty), but NMW does so at the cost of not recognizing supererogation in most of the cases in which it is to be found. Any extraordinary acts motivated by love or direct inclination cannot be seen as supererogatory, on NMW, and most actual cases of supererogation will involve one of these motives, not the sort of motive that is similar enough to the motive of duty to nearly guarantee that it will prompt an act that conforms to the CI. NMW also gives us the
strange implication that an act of supererogation is always worse, morally and absolutely, than any act of duty that is motivated by respect for the CI. NMW thus allows Kantianism to recognize supererogation, but its picture of supererogation is much altered from the standard one, and the alterations are not well-motivated.

What all of this suggests is that virtue ethics, at least in its moderate forms, is better positioned than Kantianism or utilitarianism to recognize and account for supererogation. Of course, it does not follow that virtue ethics is better able to account for supererogation than any other moral theory is. It may be that Rossian intuitionism could offer an equally satisfying account of supererogation, although there are reasons to be cautious about such a possibility.³⁰² And considering the historical roots of the concept of supererogation, it would not be at all surprising if supererogation were to be shown to be more at home in a divine command theory of ethics than it is in virtue ethics, at least if the distinction between divine commands and counsels is a tenable one. For now, though, we can set aside Rossian intuitionism, divine command theory, and the other moral theories that we have not investigated. Pending the introduction of new utilitarian or Kantian accounts of supererogation that are arguably superior to OV, the conclusion we should draw is that moderate virtue ethics has this advantage over its utilitarian and Kantian rivals: of the three, it makes the best sense of the supererogatory.

---

³⁰² One such reason is that Ross’s list of *prima facie* duties includes *beneficence*. Since almost every action will have effects on others’ well-being, and since a *prima-facie* duty generates an absolute duty unless overridden by another duty, it appears that almost every situation in which we find ourselves will be one in which we have some duty, whether arising from the *prima facie* duty of beneficence or one of the other prima facie duties. We will have latitude in only those rare cases in which even beneficence cannot adjudicate among the options that we have that have not been eliminated by other *prima facie* duties.
In some ways, virtue ethics’ overall advantage on this issue should not be surprising. When we think of the big, messy category of the morally extraordinary, we can see very quickly that it is not just actions or consequences, extraordinary things that are done or accomplished, that capture our attention. Much goes into whether or not an action, lifestyle, or person is morally extraordinary—much more than the sorts of things (rational or utility-promoting actions) that Kantianism and Utilitarianism tend to measure. We see this most clearly when we think of the associated terms “saintly” and “heroic”; the applicability of these terms seems obviously to depend on more than what concerns the utilitarian or the Kantian. My own diagnosis of the mistake that Kantians and utilitarians make is that they affirm action as being at the heart of the ethical life, a principle that is true but misleading. Action is at the heart of the ethical life in the sense that ethics is about how to live, and we live by and through action. But living is more than a collection of actions, construed as individual, intentional bodily movements, and when we turn to reflect on moral saints and heroes, it is these persons’ ways of living, their striking approaches to and perspectives on life, that most capture our attention. My OV, with its focus on intentional activities rather than discrete actions, gets closer to these ways of living, and virtue ethics in turn allows for and even encourages this focus by paying such careful attention to all of the psychological states that determine and are manifested in various ways of living.

5.4 What about the Radicals?

At the outset, I said that one of my goals, indeed my primary goal, was to develop an account of supererogation that virtue ethicists, or a significant subset of them, could
endorse. My OV is this account, and while a significant subset of virtue ethicists can endorse it without abandoning any of their core ethical commitments, there remains another significant subset of virtue ethicists who clearly cannot do so. Radical virtue ethicists, as we have seen, consider moral obligation a fiction, and surrendering this view would change the shape of their moral philosophy dramatically. Insofar as OV is committed to moral obligation, OV is thus not an option for these radicals. To use the earlier language of this chapter, OV “departs too much from the heart” of radical virtue ethics to be acceptable to radicals. The natural question that thus arises is whether, given their rejection of moral obligation, radical virtue ethicists can make room for the supererogatory in some other way.

Space does not permit me to offer a thorough treatment of this issue of whether and how radical ethicists might make room for the supererogatory. I do want to gesture at a response to this question, though, saying something about the sorts of things radicals likely will and will not be able to say about saints, heroes, supererogation, and the morally extraordinary. Since radicals eschew moral obligation, they will not be able to recognize supererogation in anything like the traditional way in which it has been understood. The absence of moral obligations implies the absence of ways of surpassing moral obligation, at least in any non-trivial way. As radicals are quick to point out, they do recognize non-moral forms of obligation: legal obligations, the obligations of social etiquette, and obligations rooted in games, for instance. But while there may be ways of surpassing (non-moral) duty in each of these (non-moral) realms, surpassing duty in some realm, any realm, is not plausibly equivalent to the category we denote when we speak of “the supererogatory.” Nor is it likely that a significant (by radical virtue ethics’ own
lights) subset of these non-moral duties could be shown to make room for the supererogatory, for if this were possible, we would expect for the relevant subset of duties to be able to do the same for the morally obligatory. Thus, while radical virtue ethics will be able to make sense of the surpassing of various types of duties, no form of such surpassing will have the special moral (or quasi moral) status that we traditionally attach to the supererogatory. The radical virtue ethicist’s picture of supererogation, if she has one, will thus be quite different from the traditional picture.

Radical virtue ethics will probably have somewhat more to say about saints and heroes, the nearly larger-than-life characters typically depicted as firmly committed to goals or ways of living that are typically beyond the rest of us. Like the moderate virtue ethicist, the radical will be able to speak of the extraordinary virtue of these people. The radical may also have much to say about what a saint or hero’s own first-person perspective reveals about the value of her project. Moreover, the school of thought that brings to our attention the import of the narrative unity of a life that is itself a quest for the good life may be particularly well-prepared for accounting for how the lives of saints and heroes tower over more ordinary ones.303 For the radical, however, insofar as the terms “saint” and “hero” denote significant categories containing particularly admirable members and insofar as these terms have not only a narrow but also a broad use (so that “saint” refers to more than religious figures and “hero” to more than those who risk life and limb to save others), the category of saints and heroes will either (i) include not only Kocheleff and the soldier in Grenade but also Beethoven and Usain Bolt or (ii) be such

303 See MacIntyre, After Virtue, 214-21.
that membership in it will not connote the same special worth that it connotes in the
contemporary moral discussion. For while there will be a distinction between
Kocheleff’s greatness and Beethoven’s, most forms of radical virtue ethics will (likely)
see this distinction as no more significant than the distinction between Beethoven’s
greatness and Bolt’s. Kocheleff, Beethoven, and Bolt will all be extraordinary figures,
for the radical, but Kocheleff will not be “overall” or “morally” best. So, the radical can
have saints and heroes and see them as extraordinary, but the radical will see these
figures in very different terms than most do.

5.5 Extravagant Love in the Morally Extraordinary

In the first chapter, I said that this dissertation had three goals: to develop an
account of supererogation that virtue ethicists, or at any rate a large subset of them, could
endorse; to show that that account of supererogation is at least as strong, as an account of
supererogation, as the accounts that have been offered on behalf of utilitarianism and
Kantianism; and to “shed some light along the way” on the nature of supererogation. It is
this third goal about which I have so far said the least. I would like to close by reflecting
on it more openly.

By saying that I wish to “shed light on the nature of supererogation,” I mean that I
want to bring out some features of the supererogatory that the contemporary discussion of
the topic tends to miss. I think the virtue ethics account of supererogation that I put
forward makes a very small measure of progress toward this goal, largely for the reasons
I detailed in section 5.3: it reminds us that the supererogatory is not merely a matter of
performing extraordinary actions but of living in extraordinary ways. I see more
significant progress being made, however, by my selection of “five central cases,” particularly my choice to include *Babette* and *Candlesticks* in that group.

The five central cases I chose are all central in the sense that, when we consider them, they seem obviously and paradigmatically to count as supererogatory. Despite this feature of all of the central cases, however, they are not all the sort that one finds prominently displayed in the contemporary discussion of supererogation. *Grenade* and *Kocheleff* are the sort that dominate that discussion insofar as each involves a significant measure of self-sacrifice that secures some good whose great utility is clear. Indeed, these two cases are intentionally modeled on two that Urmson introduced nearly sixty years ago when he redirected moral philosophy’s attention to the category of the supererogatory and set the tone for that discussion. *Torture*, while not the sort of case featured prominently in the contemporary discussion, is one that at least appears in that discussion. As I have said, Mellema is responsible for its introduction, and I include it as one of my five central cases for two reasons: to serve as a reminder that not all central cases are of the ‘great utility’ variety and to show that one way a central case can depart from the ‘great utility’ tradition is by demonstrating either a strong commitment to an ideal or extraordinary defiance of evil. The two central cases the like of which is nearly absent from the discussion (but notably, quite conspicuous in literature and film), however, are *Babette* and *Candlesticks*. Like *Torture*, these are not ‘great utility’ cases, but the reason they capture our attention and count as central cases is different from the reason that *Torture* does. What is striking about both *Babette* and *Candlesticks* is their (very different) depictions of what I would call “extravagant love.” In *Babette*, we see an extravagant love that emerges from deep gratitude and friendship; in *Candlesticks*, we see
the opposite, an extravagant loving-mercy. Extravagant love does not appear in every example of the morally extraordinary, but it does feature prominently in many such examples.

The contemporary discussion’s lack of examples like Babette and Candlesticks is, I submit, a blindness (if a curable one) to this extravagant love, which is significant since, as I say, extravagant love is often enough an important aspect of supererogation. I also suspect that a major source of the blindness is some narrowly consequentialist and deontological ways of thinking that have dominated the supererogation discussion since Urmson (and interestingly may have dominated parts of the medieval discussion as well). I don’t have an airtight argument for either of these suspicions, but I hope to give them some credence by reflecting on a final pair of cases, a pair that Babette and Candlesticks can be seen as imaging.

The pair of cases to which I am referring appear in adjacent chapters in the Gospel of John. The first, which Babette images, is the story of Mary of Bethany anointing Jesus’ feet:

Six days before the Passover Jesus came to Bethany, the home of Lazarus, whom he had raised from the dead. There they gave a dinner for him. Martha served, and Lazarus was one of those at the table with him. Mary took a pound of costly perfume made of pure nard, anointed Jesus’ feet, and wiped them with her hair. The house was filled with the fragrance of the perfume. But Judas Iscariot, one of his disciples (the one who was about to betray him), said, “Why was this perfume not sold for three hundred denarii and the money given to the poor?” (He said this

---

304 Thus, Myriel treats with highest regard a “dangerous” criminal who just stole from him and otherwise has no history, personal or otherwise, to recommend himself to Myriel.

305 I say this because I do not mean to imply that either the cause of the blindness or the blindness itself is unique to the contemporary or modern discussions. Shallow moral ways of thinking existed prior to Urmson and the modern period more generally, although these ways of thinking may have taken different shapes and been more or less dominant at various times and places.
not because he cared about the poor, but because he was a thief; he kept the common purse and used to steal what was put into it.) Jesus said, “Leave her alone. She bought it so that she might keep it for the day of my burial.”

The second, which *Candlesticks* images, is the story of Jesus washing the disciples’ feet:

And during supper Jesus, knowing that the Father had given all things into his hands, and that he had come from God and was going to God, got up from the table, took off his outer robe, and tied a towel around himself. Then he poured water into a basin and began to wash the disciples’ feet and to wipe them with the towel that was tied around him. He came to Simon Peter, who said to him, “Lord, are you going to wash my feet?” Jesus answered, “You do not know now what I am doing, but later you will understand.” Peter said to him, “You will never wash my feet.” Jesus answered, “Unless I wash you, you have no share with me.” Simon Peter said to him, “Lord, not my feet only but also my hands and my head!”… After [Jesus] had washed their feet, had put on his robe, and had returned to the table, he said to them, “Do you know what I have done to you? You call me Teacher and Lord—and you are right, for that is what I am. So if I, your Lord and Teacher, have washed your feet, you also ought to wash one another’s feet. For I have set you an example, that you also should do as I have done to you. Very truly, I tell you, servants are not greater than their master, nor are messengers greater than the one who sent them. If you know these things, you are blessed if you do them.

The parallels between Mary and Babette, on the one hand, and Christ and Myriel, on the other, should be obvious. Mary and Babette give nearly all they have in order to treat as near royalty the persons they love, with whom they have long-standing relationships, and to whom they are extremely grateful. Christ and Myriel, meanwhile, both give or serve people who have or will wrong them (here it needs to be remembered that Jesus washes the feet of both Peter and Judas and that the text indicates that Jesus is

---


307 John 13:2b-9, 12-17

308 Mary’s gratitude is (at least) for the raising of Lazarus; Babette’s is for the sisters giving her a home fourteen years earlier. The parallel is of course imperfect given that the recipient of Mary’s gift is the blameless, soon-to-be-executed Christ.
aware of both Judas’ coming betrayal and Peter’s coming denial). Both Jesus and Myriel, too, subsequently command their recipients to follow their example in the future precisely because of the favor they have been shown.309

I think, however, that there is another parallel to be noticed, a parallel that is not internal to the four cases under discussion but involves the relationship that the contemporary discussion of supererogation bears to those cases. I have characterized that relationship as a sort of blindness: the literature on supererogation, enamored with the “great utility” cases, is blind to the value and supererogatory status of cases like Babette and Candlesticks, hence the absence of such cases from the literature. A parallel can be drawn between this blindness and the way Judas and Peter respond to Mary and Jesus, respectively. Judas’ response to Mary’s anointing Jesus’ feet, we are told, is prompted by greed. Judas does not actually care about the poor, and he complains about Mary’s use of the perfume because he sees a lost opportunity for his own thievery. Judas’ response, however, is not only a function of his greed. It is also a function of his inability to see the value of what Mary is doing. Judas does not understand Mary’s gift. If he understood it, he would not bother making the argument that he makes (“Why was this perfume not sold for three hundred denarii and the money given to the poor?”) for he would already know the response, even if that response would not please him, thief that he is. Judas’ moral perspective is instead limited to the calculative terms of utility. As far as he can tell, Mary is simply throwing away three hundred denarii.

Peter also betrays a lack of understanding, albeit of a very different sort than Judas’. We have no indication that Peter’s motives in initially refusing the foot-washing involve anything like the level of moral corruption that Judas’ response does. Peter may be cowardly and immature, but he is no thief, and his cowardice has little to do with his response to the foot-washing. Still Peter, we are told, does not understand what Jesus is doing (although he will later). The reason appears to be that Peter is so fixated on roles and the rights and they imply that he revolts against Jesus’ seeming reversal of those roles. Whereas Judas’ lack of understanding was a function of a narrowly utilitarian moral perspective, Peter’s is a function of a narrowly deontological one. This is not to say that Peter lacks understanding because he recognizes that roles imply rights or that Judas lacks understanding because he realizes the value of giving to the poor—Jesus arguably affirms the same.\textsuperscript{310} The problem is that neither Judas nor Peter can see past utility and rights, respectively, and this produces a blindness to the extraordinary love that they witness.

And this problem, I submit, is also the problem that infects modern treatments of supererogation and that accounts for their ignoring central cases like Babette and Candlesticks. Urmson noted a blind spot in modern moral thought in that it made no place for acts beyond duty. On my view, however, recognizing the moral category of the supererogatory has not fully remedied the problem because the problem was greater than Urmson realized. The failure was not just a failure to recognize the existence of acts that surpass duty; the failure was an inability to recognize all that is morally extraordinary.

\textsuperscript{310} John 12:7-8 and 13:13-16.
Recognizing the existence of the category of the supererogatory remedies part of the failure, but it does not remedy all of it because important forms that the morally extraordinary takes, forms we see in the extravagant love of Babette and Candlesticks, are still being missed, and this because, at least in the literature on supererogation, narrowly deontological and, especially, utilitarian ways of thinking continue to dominate that discussion. Until the literature on supererogation can see past utility, rules, and even principles and ideals, that literature will remain inadequate for its subject matter. This is not to say that a recognition of extravagant love will make it fully adequate to its subject matter—I suspect that there is much more that the contemporary discussion is missing that I am missing as well—but it is to say that one way the contemporary discussion can be improved is to move beyond utility, roles, and rights. I hope that I have “shed some light” on the nature of supererogation by bringing out one way in which it can do so.
SELECTED BIBLIOGRAPHY


Luther, Martin. “Sermon for the Eighth Sunday after Trinity; Matthew 7:15-23.” In *Luther’s Church Postil*, trans. by John Nicholas Lenker, 1905.


