



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**Beyond Duty: An Examination and
Defence of Supererogation**

Alfred Archer

PhD in Philosophy

The University of Edinburgh

2014

I have read and understood The University of Edinburgh guidelines on Plagiarism and declare that this written dissertation is all my own work except where I indicate otherwise by proper use of quotes and references.

Abstract

Many accept that there are some acts that are ‘supererogatory’ or ‘beyond the call of duty’. Risking one’s life to save others or dedicating one’s life to helping the needy are often thought to be examples of such acts. Accepting the possibility of acts of this sort raises interesting problems for moral philosophy, as many moral theories appear to leave no room for the supererogatory. While these problems are increasingly recognized in moral philosophy, there remain a number of debates that have failed to pay sufficient attention to the existence of acts of this sort. In this thesis I investigate the implications of accepting the possibility of supererogation for three of these debates.

The first issue I investigate is the relationship between morality and self-interest. One popular view is that supererogatory acts are those that demand too much sacrifice from those who could perform them. However, I argue that looking at self-reported accounts and empirical psychological studies of moral exemplars gives us good reason to reject this view, as it has the implausible implication that those with less developed moral consciences are excused from obligations that apply to those with more developed moral sensibilities. We should accept, then, that performing an act of supererogation may be in line with an agent's self-interest.

The next debate I examine concerns the connection between moral judgements and motivation. Motivational judgement internalists claim that there is a necessary connection between moral judgements and motivation. However, it is often unclear which moral judgements this view is supposed to cover. The claim is made about judgements of 'moral goodness', 'moral rightness' and 'moral requirement'. I argue that internalists need to restrict their claim to moral obligation judgements.

I then examine how to give an account of the relationship between moral obligations and reasons for action. It is often claimed that moral reasons are overriding. A related view is moral rationalism, which holds that agents have most reason to act in line with their moral requirements. I start by examining the differences between these two views before looking at what form of either view it is plausible to hold if we accept the existence of supererogation.

I finish by looking at whether accepting the existence of supererogatory acts goes far enough or whether there is a need to make room for additional deontic categories, such as suberogation, quasi-supererogation or forced supererogation. I will argue that none of the arguments put forward in defence of these claims show that there is a need to make room for these additional categories.

Contents

Acknowledgements	6
Introduction	9
Chapter One: Defining Supererogation	25
Chapter Two: Are Acts of Supererogation Always Praiseworthy?	83
Chapter Three: Supererogation and Intentions	111
Chapter Four: Supererogation, Sacrifice and The Limits of Duty	137
Chapter Five: Supererogation and Motivational Judgement Internalism	163
Chapter Six: Supererogation, Moral Rationalism and Overridingness	191
Chapter Seven: Do We Need Room For More Deontic Categories?	217
Concluding Remarks	249
Bibliography	253

Acknowledgements

I have had many helpful discussions about the ideas in this thesis. Thanks to the following for helpful questions or suggestions about my work: Natalie Ashton, Anne Baril, Ana Barrandella, Claire Benn, Cameron Boulton, Michael Brady, Campbell Brown, Evan Butts, Justin Caouette, Robert Cowan, Ben Davies, Michael Ferry, John Fitzgerald, Guy Fletcher, Allan Hazlett, James Humphries, Sebastian Kohler, Tim Kunke, David Levy, Kevin Lynch, Amanda MacAskill, Fiona Macpherson, Neil McDonnell, Brian McElwee, Robin Mckenna, Chris Mills, Joey Pollock, Brian Rabern, Catherine Robb, Shane Ryan, Stephen Ryan, Geoff Sayre-McCord, Kyle Scott, Selena Sedat, Rowland Stout, Benjamin Sworn, Ashley Taylor, Travis Timmerman, Nick Traynor, Lauren Ware, Lani Watson, Lee John Wittington, Silvan Wittwer and Elizabeth Drummond Young.

I would like to express particular thanks to the following who have read and commented upon sections of the thesis: Shlomo Cohen, David Collins, Christopher Cowley, Richard Rowland, Caj Strandberg and Alan T. Wilson.

In addition to those who have provided substantive comments I would also like to thank Audrey Kordecki and Urszula Kordecki for proofreading sections of this thesis and Simon Frith and Jennifer McKay for proofreading the whole thesis.

Several sections of this thesis have already been published. Chapter 3 is based upon my (2013) article in *Philosophia*. Chapter 6 is an expanded version of my (2014a) article in *Ratio*. Section 7.3 is based on my (2014b) article in *Journal of Ethics and Social Philosophy*. I would like to thank the editors and anonymous referees for these journals for their helpful comments and suggestions and for allowing me to reprint this material here.

Finally, and most importantly, I would like to express particular thanks to my excellent supervisors, Elinor Mason and Mike Ridge.

I dedicate this work to my parents: John Archer, Jenny McKay, Simon Frith and Clara Glynn.

Introduction

In 2007 a man collapsed at a New York Subway station and fell onto the tracks as a train approached. Wesley Autrey, a fellow passenger, jumped down onto the tracks and held the other man down as the train came to a stop above their heads, saving the man's life. This act seems especially morally valuable but does not seem to be morally required, rather it seems like an act that we would describe as being 'beyond the call of duty'. Acts like these that go beyond what is morally required are called 'supererogatory'. Despite receiving a great deal of attention from theologians (particularly Christian theologians) the concept did not attract the attention of moral philosophers until relatively recently.¹

J. O. Urmson opened the contemporary discussion of the concept in moral philosophy.² He argued against traditional accounts of the deontic categorization of acts that assume a three-fold classification whereby acts are divided into the morally obligatory, morally indifferent and morally forbidden. Moral obligations are acts that are good to perform and bad not to perform. Morally indifferent acts are those that are neither good nor bad to perform. Forbidden acts are those that are bad to perform and good not to perform. Urmson argued that this account leaves no room for acts that are good to perform but are not bad

¹ For an overview of the theological origins of the concept see Heyd (1982 pp.15-34).

² Urmson (1958). However, while it is true to say that Urmson's paper opened up the discussion of the concept in modern western philosophy, his paper was not the first to discuss the concept. As Rabinowicz notes (2000 p.79), Kotarbinski published a short note in 1914 in which he argues that utilitarianism is unable to accommodate the supererogatory (1914/ 2000). Both Mill (1865 10.337-8) and Sidgwick (1907 p.220) also seem to recognize the problem.

not to perform, such as the actions of saints and heroes. Urmson described saints as those who regularly go beyond duty by performing acts for the benefit of others rather than pursuing their self-interest and heroes as those who go beyond duty in protecting others rather than themselves. Such actions, and less spectacular cases of kindness and generosity, are morally valuable but go beyond what is demanded by ordinary morality and cannot be fitted into the tripartite model that had been assumed by moral philosophers.

My aim in this thesis is to investigate supererogation and the implications of accepting the existence of acts of supererogation for moral philosophy. Before I begin this, though, I will look at the reasons why we might think that we should accept that acts of supererogation exist.

0.1 Why We Should Accept That Supererogation Exists

Urmson argued that for a normative moral theory to be acceptable it must make room for supererogation.³ He gave the following example to support this claim:

We may imagine a squad of soldiers to be practising the throwing of live hand grenades; a grenade slips from the hand of one of them and rolls on the ground near the squad; one of them sacrifices his life by throwing himself on the grenade and protecting his comrades with his own body.⁴

Urmson claimed that the soldier has clearly acted in a way that was good but not morally required.

What is it about this act and others such as Autrey's that leads us think that they are not morally required? There are two features of these acts that give us reason to think that they are not obligatory.

³ Urmson (1958 p.66).

⁴ (1958 pp. 63).

First, it would be inappropriate to demand that someone perform either act. As Urmson points out, it would have been inappropriate for a superior to demand this act from the soldier.⁵ This gives us reason to think that this act is not morally required because when someone has a duty to act in a certain way it does seem appropriate to demand that they do so. In the words of John Stuart Mill:

It is 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.⁶

If we accept that we can demand that people perform their duties but that it would have been inappropriate to demand that of either Autrey or Urmson's soldier that they act as they did then this gives us good reason to think that their acts were not morally required.

The second feature of these acts that suggests that they are not morally required is that in both cases it would be inappropriate to blame someone who did not perform the act. Neither the other subway passengers nor the other soldiers seem to be blameworthy for failing to act in the way in which Autrey or Urmson's soldier did.⁷ This is important because it is generally thought that there is a tight connection between moral obligations and blameworthiness.⁸ Mill makes something like this point in the following:

We do not call anything wrong unless we mean to imply that a person ought to be punished in some way or other

⁵ (1958 pp. 63).

⁶ (2001 p.49).

⁷ Urmson (1958 pp. 63) makes this point about his example. It is also endorsed by Mellema (1991 p.5).

⁸ Darwall (2006 p.97), Gibbard (1990 p.45), McElwee (2010 p.400), Portmore (2011 p.43), Shafer-Landau (2003 p.192), Stroud (1998 p.176), Williams (1995 pp.40-44).

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.⁹

The point Mill is making here is that violating a moral obligation makes someone the fitting subject of negative reactions such as punishment, blame or guilt. Stephen Darwall is more explicit on the link between moral obligation and blame in the following:

If I fail to act as I am morally required without adequate excuse, then distinctively second-personal responses like blame and guilt are thereby warranted.¹⁰

If we accept this and think that those who were in a position to act as Autrey or Urmson's soldier did but who failed to do so are not blameworthy, even in the absence of an excuse, then this gives us good reason to think that these acts were not morally required.

We can summarize these two arguments in the following way:

THE ARGUMENT FROM DEMAND:

P1 If an act is morally obligatory then the agent can legitimately be demanded to perform it.

P2 No one could legitimately demand that the soldier dive on the grenade.

C Diving on the grenade is not morally obligatory.

THE ARGUMENT FROM BLAME:

P1 If an act is morally obligatory then the agent is blameworthy for failing to perform it (in the absence of an excuse).

⁹ (2001 p.48).

¹⁰ (2006 p.26).

P2 The soldier would not be blameworthy for failing to dive on the grenade, even if he lacked an excuse.

C Diving on the grenade is not morally obligatory.

What we have, then, are two claims about the appropriateness of demands and blame (P2) which, combined with two claims about the concept of moral duty (P1), provide two arguments in support of Urmson's claim that the soldier does not have an obligation to jump on the grenade.

These arguments, combined with the assumption that the soldier's act is morally good, show that a tripartite view of deontic categories involving the required, the forbidden and the indifferent is insufficient. The soldier's act is neither required nor forbidden nor indifferent. In order to accommodate such acts in our deontic framework we must make a room for a new category of acts that are morally good but not morally required.

So far we have looked at why we might think that these acts are not morally required. However, the existence of acts that are not morally required does not by itself give us any reason to reject the tripartite view of the range of deontic options. Acts that are neither required nor wrong can be easily accommodated in this view by the category of 'moral indifference'. The reason that the acts we have considered create a challenge to the tripartite view is that as well as being neither required or forbidden, these acts are not morally indifferent either. Far from being morally indifferent, both of these acts seem to be acts of great moral importance and indeed to be the morally best course of action in these cases. There is, then, good reason to think that there are acts, such as Autrey's and that performed by Urmson's soldier, that are neither morally required, forbidden or indifferent.

Another reason to be interested in the supererogatory is that this term picks out a recognizable concept in commonsense moral discourse. Most clearly, the phrase ‘beyond the call of duty’ is a familiar part of everyday moral talk. Moreover, while the term ‘supererogation’ derives from Roman Catholic tradition,¹¹ the idea that it is possible for an act to be morally good without being morally required is one that is present in many moral traditions. In Judaism, the concept of ‘beyond the letter of the law’ applies to acts of fairness and charity that exceed what is prescribed by the law.¹² The idea is also present in Islamic ethics.¹³

To sum up, the concept of supererogation is one that picks out a recognizable feature of commonsense moral discourse, is present across different ethical traditions and seems to be required to handle cases like Autrey and Urmson’s soldier.

0.2 Supererogation and Normative Ethics

Supererogation creates a challenge for normative ethical theories. In order to be seen as compatible with commonsense morality, a normative ethical theory must allow room for acts of supererogation. Urmson pointed out that it is far from clear that any of the normative ethical theories popular in the 1950s, when he wrote his paper, (Kantianism, utilitarianism and Moorean intuitionism) could do so.¹⁴

Since Urmson raised this challenge there have been many attempts to reconcile normative ethical theories with the existence

¹¹ For an overview of the theological origins of the concept see Heyd (1982 Chapter 1)

¹² Moghaddam, Novoa and Warren (2012 p.798).

¹³ See Carney (1983), Reinhart (1983) and Moghaddam, Novoa and Warren (2012 p.798).

¹⁴ Though Urmson did argue that of these three, utilitarianism was best placed to do so.

of acts of supererogation.¹⁵ It is worth making clear at this stage that there are ways in which both consequentialists and deontologists can reconcile their views on the nature of moral value with the existence of acts that go beyond what is required. In this section I will survey some of the suggested ways in which consequentialists and deontologists might accommodate the supererogatory. My aim is not to show any advantages for one view over the other but simply to make clear that this is an issue we should take seriously regardless of our favoured view in normative ethics.

Consequentialists hold that the moral value of an action is determined solely by the consequences of that action. Acts are morally valuable for consequentialists in so far as they bring about good consequences and bad in so far as they bring about bad consequences. The reason why it would be good for me to donate a kidney to someone who needs it is that this would bring about the good consequence of saving that person's life. For consequentialists, then the good comes before the right. What it is right to do in any situation is determined by what it would be good to do, which is in turn determined by the consequences that will be brought about by the various actions available to the agent. Consequentialists give various accounts of what might count as a good consequence. Most famously, utilitarians hold that

¹⁵ Consequentialist attempts include Dorsey (2013), Portmore (2003; 2011), Slote (1984), Vessel (2010), Kantian attempts include Hill (1971) and Heyd (1982 Ch.3). For an attempt to reconcile virtue ethics with the possibility of supererogation see Brännmark (2006) and Kawall (2009). For an attempt to reconcile a form of Care Ethics with the existence of supererogatory acts see Slote (2007). See New (1974), Pybus (1982), Feldman (1986), Hale (1991) and Lockhart (2000 p.107) for attempts to deny the existence of supererogatory acts.

consequences are good in so far as they bring about pleasure and bad in so far as they bring about pain.

This kind of value can be easily reconciled with the claim that acts of supererogation must bring about more moral value than the acts of obligation they go beyond. If an obligatory act brings about consequences with a certain value then acts of supererogation must bring about consequences that are more valuable. To return to the case of Urmson's soldier, it is reasonable to expect that the consequences of the soldier jumping onto the grenade will be better than those that would result from his not doing so. Of course it may be difficult to reconcile this account with certain consequentialist accounts of moral rightness. Maximizing consequentialist theories, for example, might be thought to leave little room for acts that are of greater moral value than other permissible alternatives.¹⁶ Nevertheless, my point is simply that the consequentialist understanding of moral value seems to be easily compatible with the claim that acts of supererogation are of greater moral value than other permissible non-supererogatory alternatives.

Deontologists give a very different account of moral value. For deontologists, whether or not an act is right or wrong is not determined by the value of its consequences. There is no one deontological account of moral value but one feature shared by deontological accounts is that the right is more important and should be determined independently from the good. What makes an action right is that it conforms to the moral law. There are, of course, various accounts of what the moral law is and where it comes from. Kant held that the moral law comes from principles

¹⁶ Though the two are not necessarily incompatible. See Portmore (2011) and Vessel (2010) for attempts to reconcile maximizing forms of consequentialism with the possibility of acts of supererogation.

of reason.¹⁷ Divine command theorists, on the other hand, hold that God determines the moral law.¹⁸ What makes both theories deontological is that they both hold that certain actions are right or wrong no matter how much good their performance will bring about.

What unites deontologists, then, is that the right is prior to and independent of the good.¹⁹ By itself this seems to be perfectly compatible with what has been said so far about supererogation. If what is right is not defined in terms of what is good then wherever more than one permissible act is available there is room for one or more act to be more valuable than the others. As the act's rightness is not a result of its goodness this view can easily allow that some right acts might be better than others. It is worth pointing out, then, that deontology is perfectly compatible with the existence of acts that are morally better than other permissible alternatives.

However, what is less clear is whether a specifically deontological conception of moral value is compatible with acts of supererogation being more valuable than obligatory acts. One way to understand a specifically deontological conception of moral value is that moral value is the value in performing one's duty.²⁰ This might seem to make deontology incompatible with giving an account of supererogation in terms of moral value. If one act is obligatory because it respects the moral law it is unclear how another act could be supererogatory by respecting that law to

¹⁷ For attempts to reconcile Kant's moral theory with the existence of supererogatory acts see Heyd (1982 Ch.3) and Hill (1971). Others reject the claim that Kant's moral theory leaves room for the supererogatory. See, for example, Baron (1995), Guevara (1999) and Timmermann (2005).

¹⁸ For example Quinn (1978).

¹⁹ Fried (1978 p.9).

²⁰ See Kant (1993 p.13).

a greater extent than the obligatory act. We might think that it would speak against any definition of supererogation if it were found to be incompatible with deontological accounts of value.

A simple response to make at this point would be to accept that any deontological account that only finds moral value in following the moral law would be incompatible with supererogation. If moral value is solely brought about by observance of the moral law then there does not seem to be any means of ranking permissible acts as better or worse than each other. This, after all, is the reason why deontologists often seek to analyze supererogation in terms of duties.²¹ It is no failure of this account of supererogation, then, if it cannot be accommodated with an account of moral value that is incompatible with any unqualified account of supererogation. Perhaps deontologists wishing to accommodate supererogation should be revisionists about the concept and defend a qualified version of supererogation²² or explain why the acts that we think are supererogatory are actually best seen as duties.²³

In addition to this defence, though, it is worth pointing out that there are some deontological approaches that are compatible with the claim that acts of supererogation are more valuable than the available non-supererogatory alternatives. For example, Derek Parfit has argued that the most plausible version of Kant's theory is Kantian contractualism, which he outlines as follows:

Kantian Contractualism: Everyone ought to follow the principles that everyone could rationally will to be universal laws.²⁴

²¹ See Hill (1971).

²² As both Hill (1971) and Richards (1971) do.

²³ Baron (1995) defends this view.

²⁴ Parfit (2011 p.410).

Someone who holds this view could think that in addition to agreeing to a set of principles concerning the morally required, those in an original contracting position would agree to a set of principles about the supererogatory.²⁵ Such acts would be ones that the original contractors would agree to encourage but not require. This would allow a deontologist to say that acts of supererogation are particularly valuable, as they are acts that would be encouraged by rational contractors.

Likewise forms of deontology that allow consequences to play some role in determining the value of an action will be compatible with the claim that acts of supererogation are more valuable than obligatory acts. On some deontological accounts, the consequences of an act are relevant in assessing its moral value and may even play some role in determining what our duty is in a particular situation. For example, W. D. Ross held that if the consequences of performing a prima facie duty were sufficiently bad then this might make it permissible to omit the act.²⁶ Such a theory remains a deontological one, for the most part it is our prima facie duties that determine what ought to be done, not the consequences of our actions. Any account that, like Ross's, recognizes some form of consequential value will be able to say that when we have several different permissible acts to choose from, some of these might realize more consequential value than others. These acts would count as supererogatory.

In response to the claim that specifically deontological accounts of moral value will be incompatible with a definition of supererogation that holds such acts to be more valuable than acts of obligation we can say two things. First, while this is true for

²⁵ See Ashford (2003 p.282) for a similar account of supererogation for contractualists.

²⁶ Ross (2003 p.35).

deontological accounts that recognize only this kind of value, these accounts will be incompatible with any unqualified account of supererogation. Second, this account of supererogation will be compatible with deontological accounts that allow some room for moral value other than that of acting out of conformity to the moral law.

0.3 The Plan Ahead

So far we have seen why we should make room for supererogation and the challenges this presents for those attempting to give a normative ethical account of what makes acts right or wrong and good or bad,

However, the challenge for moral philosophy does not stop here. Supererogation also creates problems for standard forms of deontic logic. This problem has motivated the search for alternative systems of deontic logic that can accommodate the supererogatory.²⁷

Supererogation also presents a challenge to those seeking to examine the relationship between moral reasons and moral requirements. The challenge of how to reconcile the existence of acts that are good but not required with the thought that there must be some connection between moral requirements and moral goodness is sometimes called ‘The Puzzle of The Good Ought Tie Up’ and is also known as ‘The Problem of Supererogation’ or ‘The Paradox of Supererogation’.²⁸

²⁷ See Chisholm and Sosa (1966a; 1966b), Joerden (2012), McNamara (1996a; 1996b; 1996c; 2011).

²⁸ Heyd (1982 p.4) uses the term The Good-Ought Tie-Up. Chisholm and Sosa (1966b) call this The Problem of Supererogation and Horgan and Timmons (2010) call this The Paradox of Supererogation.

Both of these issues have generated substantial philosophical literatures. My aim in this thesis is to investigate previously unexplored or underexplored implications of accepting the existence of supererogatory acts. Given this aim I will not be addressing either of these issues directly, though where my discussion has implications for these two issues I will make this clear. Instead, I will explore three debates that up to now have paid insufficient attention to the supererogatory.

The first issue I investigate, in Chapter 4, is the relationship between morality and self-interest. While discussions of this issue tend to acknowledge the importance of the supererogatory, I will argue that they frequently appeal to an implausible claim about the nature of supererogation. One popular view is that supererogatory acts are those that demand too much sacrifice from those who could perform them. However, I will argue that looking at self-reported accounts and empirical psychological studies of moral exemplars gives us good reason to reject this view, as it has the implausible implication that those with less developed moral sensibilities are excused from obligations that apply to those with more developed moral sensibilities. We should accept, then, that performing an act of supererogation maybe in line with an agent's self-interest.

The next debate I will examine, in Chapter 5, concerns the connection between moral judgements and motivation. Motivational judgement internalists claim that there is a necessary connection between moral judgements and motivation. However, it is often unclear which moral judgements this view is supposed to cover. The claim is made about judgements of 'moral goodness', 'moral rightness' and 'moral requirement'. I argue first that there is no necessary connection between supererogation judgements and

motivation and second that this means that the internalist needs to restrict her claim to moral obligation judgements.

The final debate I will examine, in Chapter 6, is how to give an account of the relationship between moral obligations and reasons for action. It is often claimed that moral reasons are overriding. A related view is moral rationalism, which holds that agents have most reason to act in line with their moral requirements. I start by examining the differences between these two views before looking at what form of either view it is plausible to hold if we accept the existence of supererogation.

I finish, in Chapter 7, by considering whether making room for supererogation goes far enough. Some have argued that there is a need to make room for even more deontic categories, such as the suberogatory, the quasi-supererogatory and forced supererogation. I point out that many of the arguments for these claims fail to withstand serious scrutiny.

However, before I begin to address these issues I first need to investigate how the term 'supererogation' should be defined. The first section of my thesis will be devoted to this project. I will start, in Chapter One, by giving a definition of supererogation. In the next two chapters I will examine two challenges to the sufficiency of this definition. The first challenge, which I will examine in Chapter Two, is that my definition allows for the possibility of acts of supererogation that are not praiseworthy. This is contrary to what has been called The Standard View of supererogation, according to which all acts of supererogation are praiseworthy. If we accept The Standard View then my definition will have failed as an account of the sufficiency conditions for supererogation. However, I will argue that The Standard View ought to be rejected and so my definition is unthreatened. Next, in Chapter Three I will examine David Heyd's claim that only acts

performed with altruistic intent can be classed as supererogatory. Again, if we accept Heyd's claim then the definition I defend in Chapter One will fail as an account of the sufficiency conditions for supererogation. However, I will argue that Heyd's argument fails to give us reason to restrict supererogation to acts performed with altruistic intent.

Part One

Chapter One: Defining Supererogation

My aim in this chapter is to provide a definition of supererogation. My goal is to give an account that avoids making unnecessary commitments to controversial positions in moral philosophy. The reason for this is that these issues are ones that should be decided on their own merits rather than being prejudged by the definition of supererogation.

There have been many attempts to provide a formal definition of supererogation. Roderick Chisholm, for example, defines supererogation as, “something which it would be good to do and neither good nor bad not to do.”¹ M.W. Jackson describes supererogatory acts as acts that are “right to do but not wrong not to do”.² David Heyd, meanwhile, says that an act is supererogatory if and only if:

1. It is neither obligatory nor forbidden.
2. Its omission is not wrong, and does not deserve sanction or criticism – either formal or informal.
3. It is morally good, both by virtue of its (intended) consequences and by virtue of its intrinsic value (being beyond duty).
4. It is done voluntarily for the sake of someone else’s good, and is thus meritorious.³

Gregory Mellema offers another alternative, claiming that an act is supererogatory if and only if:

1. The performance of the act fulfils no moral duty or obligation;
2. The performance of the act is morally praiseworthy;

¹ Chisholm (1963 p.10).

² (1986 p.294).

³ (1982 p. 115)

3. The omission of the act is not morally blameworthy.⁴

Clearly, then, there is much disagreement about how supererogation should be defined. Heyd even suggests that the concept is one “that cannot be captured by a strict formal definition.”⁵

In this chapter, I will outline my positive claims about supererogation. I will be defending the following definition:

Supererogation: An act, ϕ , is supererogatory for an agent A, at time t , if and only if ϕ -ing is better than the minimum that morality demands of A at t and there is no other obligation that forbids A from ϕ -ing or for which ϕ -ing is the minimum that morality demands of A.

I will be defending my positive claims about supererogation against alternative ways of understanding supererogation. In the following two chapters I will examine and respond to arguments in support of adding extra necessary conditions to the definition.

Before I get started on this it is worth making clear a point of methodology. In this chapter I wish to provide an analysis of ‘supererogation’. However, the term ‘supererogation’ is a term of art, not one that is used in everyday language. As a result, our intuitions regarding the use of this term are likely to be unclear. Nevertheless, I think that whatever disagreements philosophers may have about the definition of the term, it is widely agreed that it is a technical term that is equivalent to ordinary language

⁴ (1991 p.13)

⁵ (2011).

phrases such as ‘beyond the call of duty’ and ‘above and beyond’.⁶

While this point may seem uncontroversial, it is not universally accepted. Paul McNamara has argued recently that ‘beyond the call of duty’ should not be seen as equivalent to ‘supererogatory’.⁷ The reason that McNamara gives to support his claim is that standard analyses of supererogation make praiseworthiness a necessary condition while this does not seem to be a necessary condition for being beyond the call of duty. This is the wrong way to approach the task of trying to understand what we mean by ‘supererogation’ and I will explain why in Chapter 2. For now though I wish to say only that our analysis of ‘supererogation’ should be informed by the way that the equivalent terms in ordinary language are used. If we find a clash between our analysis and the use of the ordinary language term then we should question our analysis before questioning the equivalence of the terms. In this chapter, then, I will be looking at what it is to judge something to be supererogatory by considering our intuitions regarding the correct uses of the ordinary language phrases such as ‘beyond the call of duty’ and ‘above and beyond’.

The first point to note about the phrase ‘beyond the call of duty’ is that this contains both a positive and a negative claim. The negative claim is that an act that meets this description will not be a duty. The positive claim is that an act that meets this description will go beyond duty. In other words, the definition of ‘supererogation’ must contain both a positive and a negative element.⁸ I will start this chapter by looking at what the negative

⁶ See Dorsey (2013 p.356), Ferry (2013), Heyd (1982 p.1), Mellema (1991), Portmore (2011 p.91), and Weinberg (2011 p.274).

⁷ McNamara (2011 p.208).

⁸ As Heyd (2011) notes.

element of this definition should be. I will then investigate what the positive element in the definition should be.

1.1. Are Acts of Supererogation Morally Optional?

In this section I will investigate whether being morally optional is a necessary condition of supererogation. Those who seek to give a reductionist account of supererogation have challenged this view. Qualified supererogationists seek to give an account of supererogation that makes such acts obligatory, albeit in a weaker sense than standard duties. I will argue that this approach should be rejected and we should seek an account of supererogation that makes these acts completely optional to perform or to omit. I will start by looking at why we might think that acts of supererogation are morally optional before going on to consider the objections that have been raised against this view.

Before looking at the two competing positions it is worth getting clear on what is meant by ‘morally optional’. A morally optional act is one that it is both permissible to perform and to omit. In other words:

Morally Optional: An act A is morally optional if and only if:

- 1) Performance of the act is morally permissible (the act is not morally forbidden).
- 2) Non-performance of the act is morally permissible (the act is not morally required).

The first part of this definition seems uncontroversial. For an act to be supererogatory clearly it seems reasonable to think that it must be permissible to perform. However, as we shall see later in the section, this view has been challenged by those that think it is possible for some acts of supererogation to be wrong for some

agents to perform.⁹ I will be responding to this argument later but it is worth noting at this point that there is at least a presumptive case in favour of the claim that acts of supererogation are always permissible to perform.

The initial focus of this section will be on the second part of this definition. I will be investigating whether or not acts of supererogation are permissible to omit. There are some theorists who claim that supererogatory acts are obligatory, albeit in a weaker sense than other obligations. I will argue that we should reject this view. I will then defend the optional nature of supererogation against those who claim that some acts of supererogation may be wrong for some agents to perform.

1.1.1 The Challenge of Qualified Supererogation

I will start by outlining the competing positions. The debate about whether or not acts of supererogation are morally optional takes place between those who give a qualified account of supererogation and those who give an unqualified account. These two positions can be defined as follows:

Qualified Supererogation: Acts of supererogation are not fully optional. They are acts that are required, albeit in a weaker sense than obligations.

Unqualified Supererogation: In order for an act to be supererogatory both its performance and nonperformance must be fully morally permissible.

To settle the debate between these two positions we must ask ourselves whether or not acts of supererogation are morally obligatory.

⁹ Straumanis (1984) and Stanlick (1999).

At first the question of whether acts of supererogation are morally obligatory may seem to be an odd one. If acts of supererogation go beyond duty then it may seem as if there is no question about their being anything other than morally optional. Being morally optional seems to be an essential part of what we mean when we say that an act is supererogatory.¹⁰ Initially, at least, unqualified supererogation has greater intuitive support than qualified supererogation. At the very least, then, the onus is on those who defend qualified supererogation to give us good reason to reject this.

In addition to being in tension with how we think of supererogation, qualified supererogation also conflicts with how we think about cases of supererogation. When we consider people who perform acts of heroism there does not seem to be any sense in which we think that these acts were obligatory. Imagine a passerby who runs into a burning building to save the life of a stranger. Those who judge that this act is supererogatory would surely want to say that there is no sense in which the agent had a duty to perform this act.

We must look then to the arguments given by qualified supererogationists to see if they give us good reason to reject unqualified supererogation. I will start by looking at Thomas Hill's qualified supererogationist account and the reasons given to reject unqualified supererogation. Hill argues that we should accept a qualified account of supererogation, as this is compatible with his favoured normative ethical theory. I will argue that this does not give us reason to accept qualified supererogation.

1.1.2 An Argument in Favour of Qualified Supererogation

¹⁰ This point is made by Zimmerman (1996 p.234).

The most prominent qualified supererogationist view is that defended by Thomas Hill.¹¹ Hill argues that we can reconcile Kant's moral theory with the existence of acts of supererogation if we hold that such acts fulfil wide scope imperfect duties.¹² An imperfect duty is one that leaves some room for the agent to determine how it shall be fulfilled. To have an imperfect duty is to have a duty to adopt a maxim that leaves considerable room for the agent to determine how this maxim should be acted upon.¹³ For example, if someone has an imperfect duty to help the poor then she is able to choose whether she does so by giving money to charity or by volunteering at a soup kitchen etc. These imperfect duties can vary in the width of their scope. The wider the scope, the more room that the agent has to determine how to fulfil her duty. Hill claims that when we have a wide scope imperfect duty to adopt a particular maxim (such as the maxim of beneficence) the act that conforms to this maxim can be both meritorious to perform and not demeritorious to omit.¹⁴ For my purposes, I am not interested in how Hill interprets Kant to fit with his interpretation of supererogation.¹⁵ Rather, I am interested in how Hill interprets supererogation to fit his understanding of Kant. The point Hill is making is that acts of supererogation are obligatory in some sense. While any individual act of supererogation may not be required we do have a duty to adopt maxims that will lead us to perform some acts of supererogation. This, then, is a qualified

¹¹ (1971). Strangely, Heyd (2011) classes both Richards (1971) and Raz (1975) as defending qualified supererogation. This does not seem like the right way to class these views, as neither thinks that supererogatory acts are a special kind of duty. Rather, both understand these acts as ones that would be obligatory were it not for the presence of exclusionary permissions.

¹² Hill (1971 p.71).

¹³ Hill (1971 p.62).

¹⁴ Hill (1971 p.68).

¹⁵ For criticism of Hill's interpretation see Baron (1995 Ch.1).

account of supererogation, as it holds that there is a sense in which we do have a duty to perform acts of supererogation.

There are two ways in which we might understand Hill's argument here. We might take Hill to mean that when we can have an obligation to adopt a maxim of this sort and an obligation to act on that maxim but any individual act of this sort will be morally permissible. If we interpret Hill in this way then his view is perfectly compatible with Unqualified Supererogation, as both the performance and nonperformance of the supererogatory act would be fully morally optional.

Alternatively, we might think that Hill's account means that acts of supererogation are not fully optional. We might think that if we have an obligation to adopt a maxim and an obligation to fulfil it then there is a weak sense in which the possible ways of fulfilling the maxim are required. They are required in the sense that if we choose not to perform any of the other available ways of acting on the maxim then we are obliged to perform this act.

Whichever way we interpret Hill, the reason given to support this account of supererogation is that it makes the concept compatible with his preferred ethical theory. This gives us one reason in favour of accepting qualified supererogation, it is claimed to offer a better fit with normative ethical theories than unqualified supererogation. More generally, we might think that qualified accounts of supererogation will be easier to accommodate in moral philosophy, as these accounts remove the need to make room for an entirely new category of action.

1.1.3 In Defence of Unqualified Supererogation

I will start by responding to the claim that qualified supererogation offers a better fit with certain normative ethical theories. This claim may well be true. However, I do not think that

this gives us good reason to accept qualified supererogation. Although Hill aims to analyze supererogation in terms of duty, what he actually achieves is something subtly but importantly different. What his argument shows is not that his ethical theory can accommodate supererogation but that it can allow for the existence of acts that are similar to supererogatory acts. What Hill seems to ignore is that acts of supererogation have a unique moral status that is not reducible to that of duty.¹⁶ This account abandons the idea that acts of supererogation are *beyond* duty.

Of course, as I mentioned in the introduction to this chapter, ‘supererogation’ is a philosophical term of art and is open to being used in different ways. We might think that it is perfectly acceptable to say that being beyond duty is not part of being supererogatory. Nevertheless, as I pointed out in the introduction, while ‘supererogation’ is a term of art there is a general agreement about the sorts of phrases in ordinary language that are thought to pick out these acts. The phrase ‘beyond the call of duty’ is taken to be a typical example of such a phrase.

It seems hard to deny that when an act is described as being ‘beyond the call of duty’ this means that it is not morally obligatory, as it goes beyond what is morally required.¹⁷ This strikes me as an uncontroversial assumption to make about what is meant by this phrase. Accounts of supererogation that seek to deny unqualified supererogation should, then, be seen as revisionary. Those who claim that qualified accounts of supererogation are a better fit with existing normative ethical theories should be seen as attempting to show that the ethical theory defended by the author is compatible with the existence of

¹⁶ This point is made by both Heyd (2011) and Horgan and Timmons (2010 p.37).

¹⁷ This point is made by Zimmerman (1996 p.234).

acts that are similar to supererogatory acts. Of course, this need not count against these ethical theories. If we see accounting for the intuitive appeal of the existence of supererogatory acts as a challenge that normative ethical theories should meet then a reasonable response to this challenge would be to show that an ethical theory is compatible with the existence of acts that are similar to supererogatory acts. What they fail to show is that being required is part of the meaning of 'supererogatory'.

Perhaps, though, it is right to defend a revisionary view. After all, for many low-key acts of supererogation the revisionary view seems plausible. It is reasonable enough to suggest that we have a general standing duty to help others though most individual acts of helping will not be required. Unfortunately for the defender of qualified supererogation, there are some supererogatory acts for which there does not seem to be any standing duty to perform. Sacrificing one's life for a stranger when there is no special obligation to help, for example, does not seem like an act that there is a standing duty to perform. This act seems to go beyond any standing duties we may have.

At the start of this discussion I argued that, given that supererogatory acts go beyond the call of duty, a reasonable initial assumption is that these acts are not in any sense obligatory. In addition, when we consider cases of supererogation it seems counter-intuitive to think that people are obliged to perform these acts. The onus, then, is on qualified supererogationists to show that this is not how we should think of supererogation. It is up to defenders of qualified supererogation to show us why we should abandon unqualified supererogation. I have looked at Hill's argument in support of qualified supererogation and found that it fails to give us reason to abandon unqualified supererogation. I conclude that we should accept unqualified supererogation.

1.1.4 Is It Ever Wrong to Perform an Act of Supererogation?

The final challenge to the claim that acts of supererogation are always morally optional comes from a different direction. Joan Straumanis and Nancy Stanlick both argue that there are some acts of supererogation that are morally wrong for members of certain groups to perform.¹⁸ In this section I will defend the optional nature of supererogation against this attack.

The starting point of Straumanis's argument is that in many of the world's cultures women are assumed to have many more caring and child-rearing duties than men and that this is not morally justified.¹⁹ These societies condition their members to have these expectations through, "a 'hidden curriculum' of sex role socialization."²⁰ Women are expected to prioritize the needs and wants of their husband and children over their own goals while men are not, or at least not to the same extent. As a result, many women find themselves in a situation where they are viewed, by both themselves and others, as having a duty to perform acts that are actually supererogatory. Straumanis argues that in this situation when women conform to these expectations they reinforce the expectation that this is what women should do. The reason for this is that:

We are each exemplars of the groups to which we belong. Whenever we perform according to the norms that are applied to that group, we reinforce those norms. To behave in exactly the expected ways, even for private reasons, is to validate the expectations of others that this is the right, natural, way for women or slaves or Blacks or poor people to behave.²¹

¹⁸ Straumanis (1984) and Stanlick (1999).

¹⁹ (1984 p.3).

²⁰ (1984 p.4).

²¹ (1984 p.10).

In other words, by exceeding familial duties women increase the pressure on other women to conform to these expectations.

Straumanis argues that a woman who does this, “commits a tiny act of treason against some woman, somewhere, whose welfare depends upon that man's understanding of the truth.”²² As a result, women owe it to each other not to perform these acts of supererogation.

Stanlick’s argument takes a different approach to arrive at the same conclusion. Drawing on the work of Jean Hampton²³, Stanlick argues that we should reject the idea that selfless actions are always morally good.²⁴ Moral behaviour must, according to both Stanlick and Hampton, involve self-regard; the recognition by the agent of her own standing as a person deserving moral recognition.²⁵ As a result, we should not regard an act as morally good if it is performed by someone who fails to pay proper attention to her own status as a valuable human being. This leads Stanlick to conclude that:

Supererogatory actions may not be permitted when they derive from or arise from character traits that are either not consistent with virtues relevant to supererogatory actions or when they are wrenched from a person by others, whether by their expectations or by their praise of “selfless” behavior in their own interest.²⁶

In other words, when an agent performs a supererogatory act without a proper regard for her own standing as a human being and source of moral claims then it is morally wrong.

Both of these arguments create a problem for my claim that acts of supererogation are morally optional. If there are some

²² (1984 p.11).

²³ Hampton (1993).

²⁴ (1999 p.215).

²⁵ Hampton (1993 pp.147-149) and Stanlick (1999 p.216).

²⁶ Stanlick (1999 pp.217-8).

supererogatory acts that it would be morally wrong for some people to perform then we must reject the claim that being morally optional is a necessary condition of being supererogatory. To defend this claim we must find some problem with the arguments we have just considered. It is to this task that I will now turn.

It is not important for my purposes to assess the strength of Straumanis's or Stanlick's argument for their practical conclusion that women have a duty not to exceed their familial duties or to recognize their own standing as human beings worthy of moral recognition. All I wish to claim is that if these arguments are successful then the acts performed by someone who acts wrongly in either of these ways cannot be supererogatory. What both Straumanis and Stanlick ignore is that an act cannot be supererogatory, at least in the moral use of the term, unless it is worthy of some form of positive moral evaluation. To see why, consider a teacher who has a duty to punish a disobedient pupil. Suppose the teacher punishes the pupil more than the required amount. In doing so he might exceed his duty to discipline appropriately but he cannot be said to have performed an act of supererogation. The reason for this is that to describe an act as supererogation is to make a positive assessment of the act. As I shall show in the next section, there are different accounts of what this evaluation might be. The positive component of supererogation could be analyzed in terms of 'moral rightness', 'moral goodness' or 'moral betterness'. For my purposes here, though, it is enough to note that whichever way we analyze the positive aspect of supererogation, Straumanis' and Stanlick's arguments can be shown to be unsuccessful.

With this in mind let's look again at the structure of both Straumanis's and Stanlick's arguments. The reason that

Straumanis thinks women owe it to each other not to exceed their familial duties is that by exceeding these duties they act in a way that will have a negative impact on the welfare of other women. The point that Straumanis is making here is that it is this negative impact that creates the duty not to exceed familial duties. This, though, shows the problem with Straumanis's conclusion. If we have a duty not to exceed familial duties in these cases then it would be wrong to do so. This being the case, exceeding these familial duties clearly cannot be morally right. Nor it would seem, can it be morally good or morally better to do so. This, at least, fits with the traditional understanding of the link between moral wrongness and moral value according to which performing a wrong act can be neither morally good nor morally better than a permissible alternative.²⁷ If exceeding these duties is not worthy of positive moral evaluation then it cannot be supererogatory to do so.

Similarly, the reason that Stanlick thinks acts of supererogation can be morally wrong is that when they are performed by agents who do not have a sufficient sense of their own self worth these acts cannot be considered to be morally good.²⁸ It would be morally better for such a person not to perform the act that exceeds her duty and this makes it morally wrong for her to perform the supererogatory act. This argument suffers from the same problem as that of Straumanis. The reason we are given for thinking that the act would be morally wrong is that it would be morally worse than the alternative acts available.²⁹ Again, though, if this is the case then the act is not supererogatory as it is not one

²⁷ See, for example, Chisholm (1963 p.7).

²⁸ (1999 p.218).

²⁹ It is worth noting that Hampton does not make the same mistake. At no point does she claim that a selfless act that is lacking in moral value would nevertheless continue to be supererogatory.

that can be considered good, right or better than the permissible alternatives. This is certainly the case if we accept that morally forbidden acts are morally worse to perform than morally indifferent or obligatory acts.

Of course, it is open to someone who is sympathetic to either position to reject this scheme but there is no reason given as to why we should do so. Moreover, such a view is going to appear somewhat bizarre. If one act is morally wrong and the other morally obligatory how could it be right, good or better to perform the first act? In addition, rejecting this account of the relationship between obligation and moral value does not seem to fit comfortably with what is motivating Straumanis's search for a justification for the claim that women owe these duties to one another. Straumanis is searching for such a justification because she believes that this will be important in achieving the morally valuable goal of women's liberation. Given this it seems reasonable to think that Straumanis's view is that it is morally good for women not to exceed their familial duties. Similarly, the structure of Stanlick's argument leaves little room for the rejection of this account of the relationship between moral value, moral requirement and moral prohibitions. The reason Stanlick gives for accepting that the selfless act is morally wrong is that it is morally worse than not performing this act. In doing so Stanlick seems to be implicitly appealing to this relationship between the deontic and the evaluative.

To sum up, I have responded to Straumanis's claim that it may sometimes be wrong to perform an act of supererogation by pointing out that this conflicts with the plausible thought that supererogatory acts must be better than the non-supererogatory alternatives and the thought that it is always morally worse to perform a morally wrong act than a morally obligatory act. Note

that my response to both claims is not limited to the particular cases she considers but would apply to any attempt to defend the possibility of acts of supererogation that it would be wrong to perform by appealing to the possibility that some examples of such acts may be morally bad.

1.1.5 Summary

In this section I have defended the optional nature of supererogation against the challenge raised by both qualified supererogation and the claim that it may sometimes be wrong to perform an act of supererogation. In the next section I will outline the second component of my definition of supererogation.

1.2. Supererogatory Acts as Morally Better Alternatives

In the previous section I argued that supererogatory acts are morally optional, that is neither obligatory nor forbidden. This, though, is far from a complete definition of supererogation, as morally indifferent acts are also morally optional. In addition to this negative claim, we must add a positive component that distinguishes these acts from morally indifferent acts.³⁰ While there has, for the most part, been widespread agreement about how to understand the negative component, attempts to define the positive component have been less widely accepted. Chisholm's account of the positive component is that for an act to be supererogatory it must be, "something which it would be good to do".³¹ Similarly, Heyd, puts the point in terms of moral goodness claiming that in order for an act to be supererogatory it must be, "morally good, both by virtue of its (intended) consequences and by virtue of its intrinsic value (being beyond duty)."³² Others,

³⁰ As Heyd (2011) notes.

³¹ Chisholm (1963 p.10).

³² (1982 p. 115). Dorsey (2013 p.356) and Horgan and Timmons (2010 p.31) also put the point in terms of 'moral goodness'.

however, put the point in terms of praiseworthiness. Mellema, for example, says that if an act is supererogatory then, “The performance of the act is morally praiseworthy.”³³ Jackson offers another alternative, saying that supererogatory acts are, “right to do”.³⁴ Finally, Michael Ferry defines a supererogatory act as one that “is better than the minimally permissible act”.³⁵

There are two key issues here. The first is whether to opt for a comparative account of the positive component or a non-comparative account. While Ferry’s account defines the supererogatory in comparison to the other available acts, Chisholm, Heyd and Mellema do not. Once I have settled this issue, I must then decide which of the possible positive terms offers the best account of the positive component of supererogation. I will start, in §1.2.1, by addressing the first issue, arguing that we must define the positive component of supererogation in comparative terms. I will then, in §1.2.2, argue that ‘moral betterness’ is the best account of the positive component of supererogation. Finally, a comparative account must explain what the relevant object of comparison is. I will finish, in §1.2.3 by addressing this issue.

Before we start this discussion it is worth making explicit a few points of methodology. First, our definition ought to make sense of the thought that acts of supererogation go beyond what is required by duty. For the purposes of the present discussion, what

Dancy (1993 p.127), Montague (1989 p.102) and Zimmerman (1996 p.234) appear to be making the same claim, though they use the term ‘moral value’ rather than ‘morally good’.

³³ (1991 p.13). Attfield (1979), Cohen (Forthcoming), and Peterfreund (1978 p.54) also put the point in terms of praiseworthiness.

³⁴ (1986 p.294). John Ladd defines the positive component in the same way (1957 p.127).

³⁵ (2013 p. 574). Similar definitions are given by Hansson (Forthcoming) and McNamara (1996a p.426).

this means is that we should seek an analysis of supererogation that presents an ethical classification that is distinct from both morally indifferent acts and acts that fulfil a duty in a minimal way. In addition, we want to give an analysis that picks out the central features of the concept rather than an analysis that picks out incidental features. Finally, we want to give an account of supererogation that does not involve commitments to substantive ethical positions. The goal in this section is to analyze the concept of supererogation so that we can then go on to look at the implications of this concept for ethical theory. It is important at this stage that the definition that we give is not needlessly prejudicing these later debates.

1.2.1 The Need for a Comparative Account

In this section I will argue that we should look for a comparative account of supererogation. As we saw above, some definitions of supererogation are clearly comparative. Ferrys' definition, for example, defined supererogatory acts as those that are better than the minimally permissible act. Paul McNamara defends a similar definition.³⁶ This, though, is not the only way in which people have defended a comparative analysis of supererogation. Dancy for example says that supererogatory acts are acts that "enjoy a very high degree of value, probably more value than any other act available to the agent."³⁷ Even though Dancy does not commit himself to the claim that supererogatory acts are always more valuable, this would be another way of giving a comparative account. Another way in which we might give a comparative account would be to say that supererogatory acts are 'more praiseworthy'.

³⁶ (1996), (2011).

³⁷ (1993 p.127).

However, many theorists working on the concept of supererogation do not define the concept in a comparative way. Mellema's definition, for example, is that an act is supererogatory if and only if:

1. The performance of the act fulfils no moral duty or obligation;
2. The performance of the act is morally praiseworthy;
3. The omission of the act is not morally blameworthy.³⁸

This definition is not comparative. Similarly, Chisholm's definition of supererogatory acts as acts that are "good to do and neither good nor bad not to do,"³⁹ does not involve any comparison. The purpose of this section, then, will be to determine whether or not the definition of supererogation should involve a comparison.

The first point to note in favour of giving a comparative account of supererogation is that it seems to be suggested by the ordinary language phrase 'beyond the call of duty'. Clearly, 'beyond' is a comparative concept, and so we might think that only a comparative account of supererogation will be able to capture this part of the ordinary language phrase that is equivalent to 'supererogatory'. At best, though, this provides only *prima facie* support for the claim that supererogation is a comparative concept. While it is worth paying attention to the ordinary language phrase used to pick out the supererogatory we should not let this settle the matter. It is possible, after all, that the phrase 'beyond the call of duty' could have become a dead metaphor the meaning of which has been transformed from the natural reading of the original metaphor. Certainly, if good reason can be found to give a non-

³⁸ (1991 p.13)

³⁹ Chisholm (1963 p.10).

comparative account then it is unlikely that the wish to respect the ordinary language phrase will carry much weight.

In fact, Chisholm claimed that there is good reason to reject a comparative account. Chisholm argued that it is implausible to think that supererogatory acts are always morally better than other acts.⁴⁰ Chisholm supports this claim by pointing out that we may be morally obliged to perform acts of great moral value. Acts of supererogation, on the other hand, can be small favours.⁴¹ For example, let's assume that firefighters have a moral obligation to run into a burning building to save someone's life when they are on duty. It seems clear that a firefighter who risks his own life to save someone else's will have performed a morally better act than someone who performs a small favour, such as helping a friend with her gardening. This seems incompatible with the claim that acts of supererogation must be morally better than other acts.

Chisholm is right to point out that obligatory acts can be better than supererogatory acts. There are some duties that are morally better to perform than some supererogatory acts. However, it would be wrong to conclude, as Chisholm does, that, "We cannot say that the supererogatory must be better than the obligatory."⁴² What we should instead conclude, as Julia Driver acknowledges, is that supererogatory acts are better than the other acts available in that particular context.⁴³ By focusing on the alternatives available to the agent at the time of acting this allows for the

⁴⁰ Chisholm (1963 p.8), Driver (1992 p.290) also makes this point but does not see it as a problem for a comparative account for the reasons given below.

⁴¹ Both Heyd (2011) and Horgan and Timmons (2010 p.32) claim that small favours can be supererogatory.

⁴² Chisholm (1963 p.8).

⁴³ Driver (1992 p.290). This point is also made by Dorsey (2013 p.356), Hansson (2013), McNamara (1996a p.426; 1996b p.175) and Zimmerman (1996 p.235).

possibility of highly valuable obligatory acts and acts of supererogation that are of low value. Chisholm's objection, then, does not cause problems for accounts that compare supererogatory acts to the other acts available to the agent in that situation rather than non-supererogatory acts in general.

In fact the comparison with other acts available to the agent appears to be essential to adequately distinguishing acts that merely fulfil a duty from those that both fulfil a duty and go beyond it. To see why we should first note that there are often a range of ways in which we can fulfil a duty. Suppose I promise to get in touch with an elderly neighbour once a week. Given this promise it seems reasonable to think that I have a duty to do so. There are a number of ways I can do this. I can send a text message, send an email or go round and visit. Let's stipulate that while the first two are equally acceptable, the final option will make my neighbour very happy, much more so than if I take one of the other options. In this case it seems reasonable to say that while all three acts fulfil my obligation to my neighbour, the final option goes beyond this and should be considered supererogatory. Without a comparative account it is far from clear how we could make sense of this. Presumably, all three of these acts are praiseworthy and good to perform. They are also morally optional in exactly the same way. For each of these acts we must either perform that act or one of the two other acceptable options. What is it, then, that makes the visiting option supererogatory? The answer has to be that although the other two options are both good and praiseworthy, the final option is morally better.

Note that this problem is not one that is raised by the particular example under consideration. Rather, it is a problem that arises from the need to accommodate cases of 'supererogatory

oversubscription'.⁴⁴ Given that there are some cases of supererogation that also fulfil a duty, we need some way to distinguish these acts from other optional ways of fulfilling that duty. The relevant point for our purposes is that no non-comparative term will allow us to do this. As both Claire Benn and McNamara point out, in order to make sense of cases of supererogatory oversubscription we need to make reference to the minimum that morality demands.⁴⁵ By referring to the minimum that is required we can distinguish between those acts that merely fulfil a duty and those that not only fulfil a duty but also go beyond it.

Mellema defends his account from this problem by offering what might appear to be a defence of a non-comparative account against this criticism. In order to handle cases of supererogatory oversubscription Mellema introduces a distinction between acts that fulfil a duty directly and those that do so indirectly.⁴⁶ An act fulfils a duty directly only if the performance of the act is praiseworthy for no other reason than the fact that it fulfils a duty, while it does so indirectly if and only if it fulfils a duty and is praiseworthy for other reasons as well. Mellema then says that while acts of supererogation can fulfil duties, they can only do so indirectly. This proposal would allow Mellema to get the right result in the neighbour case. Visiting the neighbour rather than sending a text message or email is praiseworthy not only for fulfilling a duty but also because it helps to improve the neighbour's mood. In Mellema's terminology, this act fulfils a duty indirectly and, as a result, is supererogatory. This response

⁴⁴ Benn uses this phrase to describe this kind of case (2014 p. 57).

⁴⁵ Benn (2014), McNamara (1996a p.426).

⁴⁶ (1991 p.37).

gets the right result in these cases and does so without including comparative terms in the definition of supererogation.

However, this response is only plausible if it is understood in a comparative way. We can see why by looking again at the distinction Mellema draws between fulfilling an obligation directly and doing so indirectly. An act fulfils a duty directly if it is praiseworthy for no other reasons than the fact it fulfils a duty. There are two ways in which ‘the fact that it fulfils a duty’ can be understood. The first is in a *de dicto* sense. On this reading an agent fulfils a duty directly if the only reason she is worthy of praise is that she has acted in a way that has fulfilled a duty. Mellema’s proposal can also be understood in a *de re* sense. On this interpretation an agent who fulfils a duty directly is someone who is praiseworthy only for the reasons that it is praiseworthy to perform the specific act that is her duty. To see the difference imagine a firefighter who fulfils a duty to save a child from a burning building. On the *de dicto* reading, the firefighter fulfils the duty directly if the fact that she has fulfilled a duty is the only reason that she is praiseworthy. The *de re* reading, on the other hand, would class the firefighter as fulfilling a duty directly if she is only praiseworthy for the reasons that saving the child is praiseworthy (that it saves a life or that it brings about more happiness or better consequences etc.).

Once we have distinguished between these two readings it should be clear that only the *de re* reading is plausible. After all, there are usually many reasons that make someone who performs a duty praiseworthy over and above the fact that they have performed a duty. Moreover, often the fact that they have performed a duty is not the most important reason why an agent deserves praise. For example, while the firefighter in the previous example may be praiseworthy in virtue of the fact that she has fulfilled some duty,

the main reason she is praiseworthy is that she has saved a life. The *de dicto* reading, then, would give the bizarre result that most, if not all, acts that fulfil a duty are supererogatory.

If we accept this then we can see that Mellema's response to the problem of supererogatory oversubscription turns his account into a comparative one. For in order for an act to fulfil a duty indirectly it must be praiseworthy for reasons that go beyond the reasons that performing the required act is praiseworthy. This means that this definition is making an implicit comparison to some minimally permissible way of fulfilling a duty.

I have argued that we need a comparative account of supererogation in order to distinguish supererogatory acts from other optional ways of performing a duty. I will finish this section by briefly mentioning a puzzle that has arisen out of a failure to pay attention to the need for a comparative definition of supererogation.

Jason Kawall has argued that a promise to perform an act of supererogation is one that is impossible to fulfil.⁴⁷ The reason for this is that a supererogatory act cannot be one that merely fulfils a duty. However, if we accept that promises create obligations, then once we have made the promise we cannot perform an act that would otherwise have been supererogatory without fulfilling our duty to perform a supererogatory act. As a result, this act will now be merely fulfilling an obligation and so cannot be supererogatory. However, as Benn points out, if we pay attention to the fact that supererogation involves a comparison with the minimally permissible way of fulfilling an obligation then this puzzle disappears.⁴⁸ It follows from a comparative account of

⁴⁷ (2005).

⁴⁸ (2014).

supererogation that in order to perform an act of supererogation there must be some alternative way of acting that is morally permissible but not supererogatory. The reason, then, that this promise is not one that can be kept is that there is no non-supererogatory way to fulfil this promise. As a result, there cannot be a supererogatory way to fulfil this promise either. It follows that there is no way that this promise can be fulfilled.⁴⁹ This, though, should not be seen as puzzling, it simply follows from the fact that supererogation involves a comparison to non-supererogatory ways of acting.

1.2.2 Evaluating Alternative Comparative Accounts: In Defence of The Betterness Analysis

In the last section I argued that an adequate account of supererogation must be comparative. This raises two questions. First, we need to know what kind of comparison is being made. Second, we need to know what the relevant object of comparison is. In this section I will address the first question.

As I have already shown, there are a number of terms that people have used to define supererogation. Some define supererogation in terms of ‘goodness’⁵⁰, others in terms of ‘rightness’⁵¹, ‘praiseworthiness’⁵² or ‘betterness’⁵³. This gives us four accounts of the kind of comparison being made.

⁴⁹ Benn (2014 p.59).

⁵⁰ See Chisholm (1963 p.10), Driver (1992 p.286), Heyd (1982 p.115), Horgan and Timmons (2010 p.31). Dancy (1993 p.127), Montague (1989 p.102) and Zimmerman (1996 p.234) appear to be making the same claim, though they use the term ‘moral value’ rather than ‘morally good’.

⁵¹ Ladd (1957 p.127), Jackson (1986 p.294).

⁵² Cohen (Forthcoming), Mellema (1991 p.13), Peterfreund (1978 p.54).

⁵³ Ferry (2013 p. 574), Hansson (Forthcoming), McNamara (1996a p.426).

The Rightness Analysis: Supererogatory acts are those that are morally optional and more right than the minimally permissible acts.⁵⁴

The Praiseworthiness Analysis: Supererogatory acts are those that are morally optional and more praiseworthy than the minimally permissible acts.

The Goodness Analysis: Supererogatory acts are those that are morally optional and bring about more good than the minimally permissible acts.

The Betterness Analysis: Supererogatory acts are those that are morally optional and morally better than the minimally permissible acts.

I will argue that The Betterness Analysis offers the best account of the comparison.

We might think that the issue of which term we use to define the positive component of supererogation is an unimportant one.

Given the long tradition of defining at least some of these different concepts in terms of each other we might think that it is of little consequence which term is used to define supererogation. For example, there is a long tradition of defining ‘goodness’ in terms of ‘betterness’.⁵⁵ However, the thought that it does not matter which term we use to define supererogation is misguided for two reasons. First, saying that we can define one concept in terms of the other does not by itself give us a definition of either

⁵⁴ Of course, we might worry that ‘right’ is not a graded predicate. This is just one of the concerns for this view that I will discuss below.

⁵⁵ See Åqvist (1968), Brogan (1919), Chisholm and Sosa (1966a p.246) Hansson (1990), Mitchell (1950 p.103) and Von Wright (1963). See van Benthem (1982) for an attempt to define ‘better’ in terms of ‘good’.

concept. Those who agree that ‘goodness’ is definable in terms of ‘betterness’ disagree about what the correct definition is.⁵⁶

Second, the view that our ethical concepts can be defined in terms of each other is a controversial one.⁵⁷ Those seeking to give a definition of ‘supererogation’ should set this question aside and focus on which term provides the best account of the positive component of supererogation.

The first point to note is that we can dismiss The Rightness Analysis with little effort. ‘Moral rightness’ is ambiguous as between ‘morally required’ and ‘morally permissible’.⁵⁸ If we take ‘morally right’ to be equivalent to ‘not morally wrong’ then it seems that the term should be understood to mean the latter. When people talk about ‘doing what is right’ we might take this to be closer to the former meaning. However we understand ‘moral rightness’, The Rightness Analysis is an unacceptable account of the kind of comparison being made when an act is described as supererogatory. If we take ‘morally right’ to mean ‘morally required’ then this account will be claiming that supererogatory acts are more required than obligatory acts. Clearly, this is an unacceptable account. If it makes sense to say that one act is ‘more required’ than another then clearly it is the non-supererogatory act that is more required than the supererogatory act. The other interpretation, that ‘morally right’ means ‘morally permissible’ does little better, as permissibility does not seem to be the kind of thing that comes in degrees. Even if permissibility does come in degrees, this does not seem to be the relevant kind of

⁵⁶ Contrast, for example, the definition given by Brogan (1919) to that given by Chisholm and Sosa (1966a p.246).

⁵⁷ Gustafsson (2014) argues against defining either of ‘good’ and ‘better’ in terms of the other.

⁵⁸ Ross (1930 p.6) understands ‘morally right’ in terms of ‘morally required’ while Scanlon takes ‘morally right’ to mean ‘morally permissible’ (1998 p.153).

comparison between supererogatory and merely obligatory acts. Suppose I promise to give ten pounds to charity and that as a result I have an obligation to do so. Now imagine that instead I give fifteen. Assuming that doing so does not fulfil any other requirements that I have, this looks like a clear case of supererogation. However, it seems wrong to say that the difference between the two acts is that giving fifteen pounds is 'more permissible'. The Rightness Analysis, then, does not seem to adequately capture the relevant comparison that is being made between supererogatory and non-supererogatory acts.

At first look, The Praiseworthiness Analysis is a marked improvement on The Rightness Analysis. First, unlike rightness, praiseworthiness is a property that clearly comes in degrees. Second, it has a far better claim to be picking out the relevant difference between supererogatory and non-supererogatory acts. To return to the previous example, it seems reasonable to describe the giving of fifteen pounds as more praiseworthy than giving ten pounds. On closer inspection though, The Praiseworthiness Analysis should also be rejected. The reason why is that praiseworthiness is primarily a feature of agents. It is agents, after all, that are the subject of praise. That is not to say that we cannot describe an act as praiseworthy but when we do so what we really mean is that the agent is worthy of praise for performing the act. Of course, how someone acts plays an important role in establishing whether she is praiseworthy or not. However, it is not alone in determining how praiseworthy someone is. It is commonly thought that the motivation behind an action contributes to the praiseworthiness but not the moral status of the act. As John Stuart Mill puts the point, "the motive has nothing to

do with the morality of the action, though much with the worth of the agent.”⁵⁹

Bearing this in mind allows us to see why The Praiseworthiness Analysis does not pick out the relevant form of comparison. To return to the previous example, let’s imagine two different scenarios in which I fulfil my duty to give ten pounds to charity. In the first I give ten pounds for the wrong reasons, perhaps to try and appear generous to impress those around me. In the second I do so for the right reasons, perhaps out of genuine concern for those that the money will help. In both cases I do no more than the minimum that is required of me. However, the second act is more praiseworthy than the first. If we accept that the first act is enough to fulfil the minimum that morality requires in this case then The Praiseworthiness Analysis will class the second act as supererogatory. This, though, is clearly the wrong result. The second act is more praiseworthy than the first, as doing the minimum that morality demands for the right reasons is more praiseworthy than doing so for the wrong reasons. However, the second act should not be considered supererogatory, as the act I perform in this case does not go beyond what I am required to do.

This leaves us with two options: The Betterness Analysis and The Goodness Analysis. Like The Praiseworthiness Analysis, both of these analyses involve properties that come in degrees. Likewise, both seem plausible candidates for picking out the relevant difference between supererogatory and non-supererogatory acts. In the charity donation case it seems reasonable to think that

⁵⁹ Mill (2001 p.18). What motives we must look for is subject to debate. We might think that the agent must be acting from the motive of duty as Kant (1993) does. Alternatively, we might think that an act is praiseworthy if it stems from good will, as Arpaly (2003 Ch.3) does. Finally, we might think, as Markovits (2010) does, that an act is praiseworthy if the agent’s motivating reasons coincide with the reasons justifying the act’s performance.

donating fifteen pounds is both morally better than donating ten and that it brings about more good. Finally, unlike The Praiseworthiness Analysis, both of the analyses involve properties of acts rather than agents.

How, then, should we choose between the two? We might think that there is just no difference between the two. If we accept John Broome's claim that, 'more good than' is synonymous with 'better than' then there will be no difference between the two analyses and it will not matter which we choose.⁶⁰ However, there is good reason to think that The Betterness Analysis should be preferred as it is the more neutral of the two. A strict deontologist who makes no consequentialist concessions, for example, might be wary about allowing for any discussion of the goodness of acts. Nevertheless, if someone who held this position allowed that some acts can be morally better than others then they might be able to accommodate supererogation. Someone else might disagree with Broome's claim by arguing that there are two ways in which one act can be better than another. It can either bring about more goodness or it can bring about less badness (where these are not synonymous). If there is a distinction to be made here, there seems no reason to restrict supererogation to the former. It seems no less plausible to say that someone who sacrifices a great deal to prevent something bad from happening has performed a supererogatory act than to say the same about someone who brings about a great deal of good. Of course Broome's claim may well be right. If it is then The Betterness Analysis is no different to The Goodness Analysis. However, if there is a distinction to be drawn between the two then The Betterness Analysis is the more plausible. In order to avoid pre-judging these issues, then, we should prefer The Betterness Analysis to The Goodness Analysis.

⁶⁰ (2004 p.50).

1.2.3 How to Formulate The Betterness Analysis

So far I have argued that we should define supererogation in terms of acts that are better than the minimum that morality demands. However, this definition leaves it unclear exactly how to understand ‘the minimum that morality demands’. There are three options that have been proposed. Michael Zimmerman suggests a time-relative account.⁶¹ According to this account, supererogation is defined in relation to what is the minimum that morality requires at a particular time. McNamara, on the other hand, defines supererogation in relation to the lowest ranked set of worlds in which the agent acts permissibly from the time of acting onwards.⁶² Finally, Sven Ove Hansson claims that we should understand supererogation in terms of the minimal way of fulfilling a particular obligation.⁶³ There are two ways in which any account could be criticized. First, an account may class acts as supererogatory that should not be classed as such. Second, an account may fail to class acts as supererogatory that should be classed as such. In the remainder of this section I will argue that none of these ways of spelling out what is meant by ‘The Minimum that Morality Demands’ are able to avoid both objections. I will then give my own account of how this should be understood that avoids both of these objections.

One way in which we might be tempted to define the relevant notion of the minimum for supererogation is in terms of the minimum that morality demands from us at a particular moment in time. In Zimmerman’s discussion of supererogation, he points out that by making supererogation agent and time relative enables us to explain why some obligatory acts are morally better than some

⁶¹ (1996 p. 236).

⁶² (1996a.; 1996b; 2011).

⁶³ (Forthcoming).

supererogatory acts.⁶⁴ If we make supererogation agent-relative then this allows us to say that while it would be supererogatory for a passerby to save a child from the burning building, it would not be supererogatory for a firefighter to do so. Similarly, making supererogation time-relative allows us to say that it would be supererogatory for the firefighter to help her friend with his gardening while off duty but not supererogatory to do so when she is supposed to be working. We might, then, be tempted by the following account of supererogation:

Time-Relative Betterness Analysis: An act ϕ is supererogatory for an agent, A, at time t , iff it is better than the minimum that morality demands of A at t .⁶⁵

This account avoids the problem of being overly restrictive.

Clearly, an act cannot be supererogatory if it is no better than the minimum that morality demands from the agent at that time.

There will, then, be no acts that should be defined as supererogatory that this definition will not class as such.

However, this account does not avoid the problem of being overly permissive, as it will allow fulfilling imperfect duties in a minimal way to count as supererogatory. For example, suppose my friend is going to be in hospital for a week. I promise to visit at some point and as a result I now have a moral obligation to do so.

Suppose that it makes no difference what day of the week I visit on. I make my visit on the first day of the week. According to The Time-Relative Betterness Analysis I have performed a supererogatory act. After all it was morally better to go to visit my friend than not to do so and morality did not require that I visit my

⁶⁴ (1996 p.235).

⁶⁵ This account is suggested by Zimmerman's discussion of supererogation (1996 p. 236), though, he does not set it out formally like this.

friend at that particular time. Clearly, however, this act is not supererogatory, as it merely fulfils my duty in a minimal way. We might think that we could avoid this problem if we supplement The Time-Relative Account with further necessary conditions. This is, I think, the right place to look for a satisfactory definition. However, before I examine how best to supplement this account, I will first explain why the alternative accounts are problematic.

First let's look at McNamara's account. According to this proposal, an agent has a range of possible worlds that are accessible to her. Within this set of worlds is the subset of 'acceptable worlds', those where she conducts herself permissibly from now on.⁶⁶ The acceptable worlds are then ordered, such that the morally better the world, the higher it will be ranked. A supererogatory act, then, is one that occurs in some acceptable world but fails to occur in any of the lowest ranked acceptable worlds.⁶⁷ This allows McNamara to give the following version of The Betterness Analysis:

McNamara's Betterness Analysis: An act ϕ is beyond the call iff it occurs somewhere among the permissible alternatives but not among any of the minimal ones (and thus above them).⁶⁸

⁶⁶ We might worry that this is a misinterpretation of McNamara's view. McNamara describes the acceptable worlds as the subset of those (accessible worlds) where, "she comports herself permissibly *as of now*," (2011 pp.212-213 emphasis added). Perhaps, this could be interpreted as meaning those worlds where the agent comports herself permissibly *at this very moment* rather than *from now on*. However, the latter reading is certainly the more natural one and McNamara has confirmed in personal correspondence that this is how it should be read.

⁶⁷ (1996 p.436).

⁶⁸ This is the exact wording McNamara gives for the concept of 'Action Beyond Morality's Call' (2011 p. 215). This is formally represented as follows $SUp =df PEP \ \& \ MI\neg p$ (1996a p.429; 1996b p.177) where Sux : *x-ing* is supererogatory. PEx : *x-ing* is morally permissible. MIx : meeting morality's demands minimally involves

Like The Time-Relative Analysis, McNamara's account explains why an obligatory act may be morally better than a supererogatory act. When we make clear that a supererogatory act need only be better than the minimally acceptable accessible worlds, then we can see that these examples are no longer problematic. When an agent has an obligation to save someone from a burning building, choosing to help a friend with her gardening will not feature in any of the acceptable worlds. As a result, this would not count as an action that goes beyond duty.

This account also avoids the problems facing The Time-Relative Analysis. The problem with this account was that it looked only to what is required at a particular time and so wrongly categorized the fulfilment of an imperfect duty in a minimal way as supererogatory. By looking to sets of worlds, McNamara's account avoids this problem, as the worlds in which I fulfil my duties in a minimal way will be equally ranked. The world in which I visit my friend in hospital on day one would not be ranked any higher than the world in which I visit her on day two and so McNamara's account will not class this act as supererogatory.

However, McNamara's account does face the problem of being overly restrictive.⁶⁹ In McNamara's account in order for an act to

x-ing. In earlier work McNamara took the concept of 'beyond morality's call' to be equivalent to 'supererogation' (1996a). However, in later work, McNamara (2011) claims that 'supererogation' requires the additional necessary condition that the act also be praiseworthy. In Chapter 2 I will argue that we should reject this claim.

⁶⁹ Hansson (Forthcoming p.4) claims that it is problematic that McNamara's account is open to a version of 'Ross's Paradox' for standard deontic logic. According to standard deontic logic, if p entails q and p is obligatory then q will be obligatory as well. This is problematic, as it generates odd obligations. For instance, suppose I have an obligation to post a letter and it would be forbidden to burn it. It follows from this that I have an obligation to either post the letter or burn it, as for any p , p entails p or q .

be supererogatory it must occur in at least one acceptable world that is accessible to the agent. In other words it must occur in a world where the agent conducts herself permissibly from now on and one that is accessible to her, given her abilities and disabilities. However, this rules out the possibility of an act that is supererogatory but only occurs in a world where the agent does not act permissibly from then on. To see why this is problematic, imagine that someone who responds to praise by becoming temporarily arrogant is passing by a burning building and hears cries for help coming from inside. The passerby's arrogance consistently leads him to behave in a rude and short-tempered way. Thankfully, though, this arrogance only ever lasts a day. Let's stipulate that in all of the possible worlds where he rescues the child, he is praised for doing so and, as a result, behaves in a rude and short-tempered way to those around him for a day. Let's further stipulate that this behaviour is morally wrong, though only in a minor way.

Now we must ask ourselves whether the fact that saving the child will lead the passerby to perform some minor wrongs should

This is problematic, as it looks like I can fulfil this obligation by burning the letter. McNamara's Betterness Analysis is open to a similar objection. Suppose a mailman is in a position to rescue a child from a burning building and that doing so would be supererogatory. Fanning the flames, on the other hand, is morally forbidden. However, given that it is permissible to rescue the child, it will also be permissible to either to rescue the child or fan the flames. Moreover, both acts are precluded by doing the minimum. As a result, McNamara's Betterness Analysis gets the odd result that it would be supererogatory either to rescue the child or fan the flames. However, McNamara has convincing responses to this problem (1996 p.431). First, as McNamara points out, this is a problem not for his account but for the combination of his account and standard deontic logic. Moreover, McNamara argues that the standard replies to this paradox are plausible here as well. Finally, we can avoid this problem altogether by stipulating that in order for an act to be supererogatory, it must not involve the performance of any impermissible acts.

prevent this act from being supererogatory? There doesn't seem to be any reason to think that it should. We would think it perfectly permissible for the passerby to rescue the child, even though this will lead him to perform some minor wrongs. The world in which these acts occur also seems to be morally preferable to any in which the child is not saved. The case might be different if the passerby would perform major wrongs as a result of saving the child. If the passerby were to respond to the praise by embarking on a killing spree then perhaps it would not be permissible to save the child (though even then we might say that the saving of the child is permissible while the killing spree is not). In this case, though, this is an acceptable trade-off to make. This is not to say, though, that it ceases to be wrong for the passerby to behave in a rude and short-tempered way. The fact that it is permissible for the passerby to perform the supererogatory act even though he knows that this will result in his performing a wrong act does not change the fact that the wrong act is impermissible. Given that an acceptable world is one where the agent acts permissibly *from then on*, it seems possible for an act to be supererogatory even if there is no acceptable world in which it occurs. This is problematic for McNamara's account as, on this account, an act is only supererogatory if it occurs in an acceptable world.

Of course, it is open to McNamara to respond to this point by claiming that an act cannot be supererogatory if the only the accessible world in which the agent performs the act is one where she goes on to perform an impermissible act. Perhaps it could be claimed that it is never permissible for someone to act in a way that will lead her to act impermissibly. However, this approach involves making a commitment to a substantial first-order moral claim that a definition of supererogation would be better off avoiding. The question of whether it can ever be permissible for

an agent to act in a way that will lead her to act impermissibly is not one that our definition of supererogation should pre-judge.

Another account of the relevant version of The Minimum That Morality Requires comes from Sven Ove Hansson, who offers the following analysis:

Hansson's Betterness Analysis: a supererogatory act is an optional action that is a better variant of another, obligatory action.⁷⁰

By looking to lesser ways of fulfilling particular duties Hansson's analysis avoids the problems facing The Time-Relative Analysis. On this account, an act cannot be supererogatory simply because it is better than that which morality requires from an agent at that particular time. Visiting my friend in hospital on day one would not count as supererogatory, as it is no better than visiting on any of the other days. This account also avoids the problems facing McNamara's account. By avoiding talk of 'acceptable worlds' Hansson does not commit himself to the claim that an act cannot be supererogatory if the only accessible world in which the agent performs the act is one where she goes on to perform an impermissible act.

However, unlike McNamara's account, this account is only able to speak about supererogation in relation to particular duties. There are two problems with this. First, this is unsatisfying, as we want to be able to say that that an act goes beyond any duty the agent had, not just that it has gone beyond one particular duty. Take the passerby example; we want to be able to say that saving the child in this case is just supererogatory. This is particularly important when we consider that whatever duty the passerby's rescue is

⁷⁰ (Forthcoming p.6). Hansson symbolizes this in the following way: p is supererogatory with respect to q if and only if $p \vdash q, Oq, \neg Op, \neg O\neg p$, and $p > (q \& \neg p)$.

supererogatory relative to (an imperfect duty to help people to some extent perhaps) would also apply to a firefighter. Relative to this duty, a firefighter would also be performing an act of supererogation, though of course relative to the duties of a firefighter she would not. What we want is a way of saying that while the passerby's act would be supererogatory, the firefighter's would not. If all Hansson's account can offer is an account of what it is for an act to be supererogatory in relation to a particular duty then it will be unable to provide this.

Hansson responds to this concern by introducing the new concept of 'a substandard variant of an obligatory action'. He defines this as follows:

An act p is substandard variant of an obligatory action, q , if and only if p is a variant of q , q is required, p is optional and q is better than p and q .⁷¹

Having outlined this concept Hansson tentatively suggests that an act is supererogatory simpliciter if and only if it is supererogatory with respect to at least one other obligatory act, and substandard with respect to none.⁷²

However, this account will not work, as there will be some acts that are the minimally permissible way of performing an obligation that this account classes as supererogatory. To see why this is problematic, compare the cases of the passerby and the firefighter saving a child from a burning building. In order for the passerby's act to be supererogatory, on this account, it needs to be a better version of some obligation (perhaps a general imperfect obligation to help those in need). However, the firefighter will

⁷¹ The more formal version Hansson gives is as follows: p is substandard with respect to q if and only if $p \vdash q, Oq, \neg Op, \neg O\neg p$, and $(q \& \neg p) > p$. (Forthcoming p.6).

⁷² (Forthcoming p.7).

presumably be performing a better version of this obligation as well. The difference between the firefighter's act and the passerby's is that while both will perform a better version of this duty, the firefighter's act also fulfils another duty, a special firefighter's duty to protect those endangered from fires. Now imagine that there is only one possible way for the firefighter to fulfil this additional duty in this context. This would mean that there is no substandard variant of the obligation available to her. This means that she has performed an act that is supererogatory with respect to some duty and substandard with respect to none. On Hansson's account of supererogation simpliciter this act will have to be classed as supererogatory, even though all she has done is fulfil her duty in a minimal way.

Note that this problem will arise in any situation in which relative to one source of duty we are acting in a minimally permissible, though not substandard, way and relative to another source we are going beyond what is required. Take, for example, special obligations that are generated by family or friendship relationships. If there is only one way of fulfilling these obligations and they all exceed our impersonal obligations to others then performing the minimally permissible act will be supererogatory. This shows that Hansson's account of supererogation simpliciter is unacceptable. Supererogatory acts are supposed to go beyond duty; this is incompatible with these acts fulfilling duties in a minimal way.

It might be thought that a simple amendment to Hansson's view would allow it to avoid this problem. Perhaps we should say that in order to be supererogatory in a way that is not relative to a particular duty an act must be supererogatory with respect to the most important duty in play and substandard with respect to

none.⁷³ This would get the right result in the previous case, as the most stringent duty in play for the firefighter is to save the child from the fire. As a result saving the child would not count as supererogatory.

However, while this result gets the right result in this case it does so at a price. This new account of supererogation is overly restrictive, as it will prevent any act from counting as supererogatory whenever a more important duty is in play. This is the wrong result, as it looks like there are cases where we perform supererogatory acts that do not go beyond the most stringent duty in play. Suppose I have borrowed a large amount of money from my friend and that she needs to be repaid this money by the end of the week in order to avoid defaulting on her mortgage. Does the existence of this stringent duty mean that any act I perform this week that is not a better version of this duty cannot be considered supererogatory? Clearly not. To see why, suppose that on the way to meet my friend to repay her money I see someone struggling to carry home his shopping and offer to help him. This act seems like a plausible case of supererogation. However, this act does not seem like a better version of the most stringent duty that is in play. Whatever duty this goes beyond it does not seem plausible to think that it is the duty to repay my friend by the end of the week. However, on the revised version of Hansson's account this act can only be classed as supererogatory if it is a better version of the duty to repay my friend, as this is the most stringent duty in play. Note that the problem for Hansson's account is not simply that it gets the wrong result in this case but that it does so for a very odd reason. In this case the presence of the more stringent duty prevents the act of helping the stranger with his shopping even

⁷³ Thanks to Elinor Mason for suggesting this response.

though doing so at this time is entirely compatible with the performance of that more stringent duty.

A second problem with Hannson's account is that it appears to be overly restrictive. While this account does well at handling cases of supererogatory oversubscription not all cases of supererogation fit this model. As both Julia Driver and Joel Feinberg point out, favours such as offering a stranger a match or helping a friend paint his house seem like cases of supererogation but do not fit easily into the over-subscription model.⁷⁴ As Hannson acknowledges, this is a challenge for his account as it appears that there are some acts that are clear cases of supererogation but are not better variants of other obligatory acts.⁷⁵ He gives the example of someone who passes an upset looking stranger in the park and spends a couple of hours comforting him. This act appears to be supererogatory but it is far from obvious what obligatory act it is a better version of. If this objection holds then Hannson's definition is an overly restrictive account of supererogation.

Hannson responds to this problem by claiming that the obligation that the act in the previous example goes beyond is a general obligation, perhaps to help people in need or to be kind to strangers.⁷⁶ For our purposes, it is not the content of the obligation that is important but the resources it offers Hannson to respond to the problem. If we accept that there is a general obligation to be kind to strangers, and that this act exceeds this obligation, then this case is no longer problematic for Hannson's account.

There are, though, two problems with this response. First, it just seems odd to say that this act is a better version of the general obligation to help strangers or be kind to people. The more natural

⁷⁴ Driver (1990 p.290), Feinberg (1968 pp.279-280).

⁷⁵ (Forthcoming p.7).

⁷⁶ (Forthcoming p.7).

way of describing acts like this one is that the agent was under no obligation whatsoever to help. Second, as Hansson acknowledges, there may be some cases of supererogation where there is no lesser obligatory option available.⁷⁷ For example, in Urmson's example of the soldier who jumps on a live grenade to save his comrades, there does not appear to be any other course of action available that would help to save the other soldiers. Given that there is nothing short of this that the soldier could have done to protect his comrades it looks like if we accept that this act is supererogatory we must also accept that the soldier is under no obligation to protect his comrades here. Hansson responds to this objection in the following way:

Arguably, our use of the term 'supererogatory' in such (rather unusual and untypical) cases is derived from analogies with similar cases in which there is an obligation to help that one goes beyond.⁷⁸

In other words, these cases are not really cases of supererogation they just appear to be so because they are similar in some ways to acts that are supererogatory. This response is unsatisfying, though, as the case we have considered appears to be a paradigm case of supererogation, not a marginal case. After all, it was this case that Urmson used to persuade others of the need for moral philosophers to address this concept. Unless Hansson can provide some independent reason to think that acts like this should not be considered supererogatory then this is a significant problem for his account.

So far we have looked at three versions of The Betterness Analysis and found all to be wanting. McNamara's Analysis was found to be overly restrictive. The Time-Relative Analysis, on the other hand, was too permissive. Finally, Hansson's Analysis was

⁷⁷ (Forthcoming p.7).

⁷⁸ (Forthcoming p.7).

a mixture of both, preventing some acts that should count as supererogatory from doing so and including other acts that should not count as supererogatory. However, of these three accounts The Time-Relative Analysis appears to be the most promising. The problem with this analysis is simply that it fails to provide the correct sufficiency conditions for supererogation. Of the three accounts though, it is only this one that is a plausible necessary condition for supererogation. It is worth investigating whether this account could be supplemented with further necessary conditions to give an acceptable account of the sufficiency conditions for supererogation.

The problem with The Time-Relative Analysis is that there are some obligations which the agent has some choice over the exact time at which to fulfil them. This is problematic, as fulfilling these obligations in a minimal way is not supererogatory but it does satisfy the criteria given by this analysis. The solution to this problem is to add an additional necessary condition that prevents any acts that fulfil a duty in a minimal way from counting as supererogatory, as the following analysis does:

The Revised Time-Relative Analysis: An act, ϕ , is supererogatory for an agent A, at time t , if and only if ϕ -ing is better than the minimum that morality demands of A at t and there is no other obligation that forbids A from ϕ -ing or for which ϕ -ing is the minimum that morality demands of A or a way of partially fulfilling the minimum that morality demands.

This account avoids the problems facing the original Time-Relative Analysis as the additional condition rules out the fulfilment of a duty in a minimal way as being an act of supererogation. However, it also retains the spirit of the original analysis in looking, primarily, to what is required from an agent at a particular time for the comparison case. We might worry that

including a necessary condition just to rule out the problem cases is somewhat ad hoc. This concern, though, is misplaced. The new condition is one that is suggested by the phrase ‘beyond the call of duty’. To go beyond duty, it is not enough that an act be a minimal way of fulfilling a duty. This condition simply captures this thought.

This account also avoids the problems facing McNamara’s account, as it does not rule out the possibility of acts of supererogation that only occur in worlds where the agent does not act permissibly from then on. While this analysis does rule out the possibility of acts of supererogation that are impermissible, it does not rule out the possibility that an act may be permissible even if performing it will lead the agent to perform an impermissible act later. Nor does this analysis declare definitively that this is possible. On this account this issue will come down to whether it is morally permissible to make this kind of trade off. This is as it should be. Instead of being prejudged by our definition of supererogation, this issue is one that should be determined by first order moral theorizing.

Finally, this account also avoids both problems facing Hannson’s analysis. The first problem for Hannson’s account was that it was unable to give an acceptable account of what it is for an act to be supererogatory simpliciter. We should not take this advantage too seriously, however, as it seems plausible that adding the same additional necessary condition to Hannson’s account would provide the same results. The real advantage for this analysis over Hannson’s is that it can accommodate cases of supererogation that do not fit the over-subscription model. Hannson’s Analysis needed to make the implausible claim that all acts of supererogation are better variants of other obligatory acts. The Revised Time-Relative Analysis, on the other hand, can allow that

an act can be supererogatory without being a better variant of an obligatory act. So long as an act is morally better than what the agent has to do at that moment and is neither forbidden nor the minimal way of fulfilling some duty, then this is enough for it to be supererogatory.

In this section I have looked at problems for three ways of understanding The Betterness Analysis. I have argued that The Revised Time-Relative Analysis is able to avoid the problems facing all of the other accounts we have looked at. In the next section I will consider two objections that might be raised against the definition I have given by those who think we need to add extra necessary conditions to this definition.

1.2.4 Objections and Responses

One challenge that could be raised against the account I have been defending is that it does not go far enough. We might think that in order to be considered supererogatory an act must be not only be morally better than the minimum morality demands but *significantly* better. To see why we might think this, let's imagine that I borrow some money from a friend and promise to repay her within a year. Let's suppose that, although she has no immediate need for the money, it is slightly better from my friend's point of view if the debt is repaid as soon as possible, perhaps because she is uneasy about a friend owing her money. Let's suppose there are two possible acts available to me. I can repay the debt at the end of the year as agreed or I can repay her one day before the end of the year. We might think that it would be morally better, though only slightly so, to repay her a day earlier. According to my definition of supererogation, repaying a day early would count as supererogatory. This might strike some as odd. It would, after all, seem strange to say that I have gone beyond the call of duty here. This objection could be avoided by requiring acts of

supererogation to be *significantly* better than the minimum morality demands.

This modification does however come at a cost. Responding to the objection in this way seems to open up space for a new category of acts that are like acts of supererogation in every way but are not considerably better than the minimum permissible acts they go beyond. Opting for this response creates a new category of acts that are permissible and better than the permissible alternatives but not by enough to be considered supererogatory. If we accept that, all else being equal, we should prefer a simpler theory to a more complicated one then it seems that creating a new category of acts should be considered a cost of this approach. Again, this is not a decisive objection to this line of response but it does give us reason to prefer an alternative response if one is available. Given this cost, if we can find an alternative way of responding to the problem that does not bring with it this cost, or any of comparable or greater significance, then we should prefer that way of responding to the theory.

A better way to respond is to debunk the intuition. The first step of this response involves identifying the reason why it seems strange to describe the act that is only marginally better than the minimum that morality demands. I think that the reason why it would be strange to describe this act as beyond the call of duty is that we would think it odd to praise someone for acting in this way. Given that the obligation in this case has only been exceeded by a minimal amount it would seem strange to praise the agent in this example. This means that this objection should only worry us if we think that acts of supererogation are necessarily praiseworthy. Given that praiseworthiness does not form part of my definition of supererogation this is unproblematic for me. Of course, some explanation still needs to be given as to why it seems strange to

describe this act as supererogatory. However, if we accept that the reason that this seems odd is that this act would not be praiseworthy then the explanation that is needed will be one that explains why it seems counter-intuitive to describe an act as beyond the call of duty but not praiseworthy (for whatever reason). In Chapter 2 I argue that this intuition should be explained in terms of conversational implicature rather than truth conditional content. If this argument is successful and my debunking explanation is right then the claim that it seems odd to describe an act that is marginally better than a merely permissible alternative as supererogatory will cause no problem for my view.

Another objection that might be raised against The Betterness Analysis is that because there is no requirement that a supererogatory act be morally good, it would be possible for an act to be supererogatory and morally bad. This may strike some as odd. It would certainly be weird for someone to describe an act as 'beyond the call of duty but bad'. This might tempt someone to add an extra necessary condition. This would give us the following definition:

The Goodness and Betterness Analysis: An act, ϕ , is supererogatory for an agent A, at time t , if and only if ϕ -ing is better than the minimum that morality demands of A at t and there is no other obligation that forbids A from ϕ -ing or for which ϕ -ing is the minimum that morality demands of A and ϕ -ing is morally good.

The difference between this hybrid account and The Revised Betterness Analysis is that if it is possible for an act to be morally optional and morally better than the minimum that morality demands without being morally good then this analysis will yield different results. While The Betterness Analysis will class such acts as supererogatory, The Goodness and Betterness Analysis

will not. For example, imagine a situation where we must choose to perform one of two morally bad acts.⁷⁹ Suppose a murderer breaks into a school and tells the teacher that she will either kill five students in his class or kill three different students along with the teacher. The teacher must now make a choice. Either he tells the murderer to kill the five children or he tells the murderer to kill himself and three others. In this case we might be tempted to say that both acts are bad. Of course we might also think that one of these options is morally wrong and the other morally required. Alternatively, we might think that this is a genuine moral dilemma in which both acts are morally wrong. Nevertheless, I take it that it is not wildly counter-intuitive to think that although neither act is morally wrong both are morally bad, though the latter is better than the former. After all, we might think that to let five children die when there is a chance to intervene to allow less children to die is bad, while to dirty one's hands by choosing which children will be killed is also bad. On top of this we might think that, though the latter act is better than the former, this is not the kind of act that we can demand from someone; perhaps because this act involves the teacher giving up his own life or perhaps because we think that as the teacher has made the education of children a major life project, it would be inappropriate to demand that he actively engage in helping the murderer.⁸⁰

The question we must now ask is if we accept this description of the situation then should we say that it would be supererogatory

⁷⁹ Note that, though similar, this is different from a 'moral dilemma', at least as these are traditionally defined, as cases where an agent must choose between two *wrong* acts rather than two *bad* acts. For a defence of the claim that moral dilemmas exist see Marcus (1980), Nussbaum (1986 Ch. 2, Ch.3) and Williams (1973 Ch. 11).

⁸⁰ We might agree with Bernard Williams that it would be absurd to demand that someone act with such disregard for their own life projects (1973b p.116).

for the teacher to take the latter option? The answer is that we should. This is because there are two problems with accepting the alternative. First, there is nothing in the phrase ‘beyond the call of duty’ that seems to warrant this restriction. This phrase is a simple comparative one. When we describe an act in this way we are making a claim about that act compared it with what is morally required. If an act is morally better than what is required then this seems sufficient for the use of this phrase to be appropriate.

Second, this approach creates the need for a range of new deontic categories. First, we must make a new category for acts that satisfy the other necessary conditions for supererogation but are morally bad. This, though, is just the beginning. If we accept that the goodness and badness of an act can vary so dramatically from the acts deontic status then presumably it will also be possible for an obligatory act to be morally bad and a prohibited act to be morally good. In the interests of parsimony, then, there seems good reason not to accept this extra condition.

These points give us sufficient reason to think that the simpler definition is superior until given good reason to prefer the more restricted definition. However, the only reason that there seems to be to add this necessary condition is the oddness of saying that an act is ‘beyond the call of duty but morally bad’. Note, though, that it seems equally odd to describe an act as ‘morally wrong but morally good’ or ‘morally obligatory but morally bad’. There seems little reason, though, to think that we should include ‘goodness’ and ‘badness’ in our definitions of ‘moral obligation’ and ‘morally wrong’. Moreover, if we did include ‘morally good’ in our definition of ‘moral obligation’ then there would be no need to do so for supererogation as well. If this were the case then supererogatory acts would also count as morally good, as these acts are better than at least one obligation.

1.3 Dorsey's Radical Alternative

In the previous section I investigated how to give a precise formulation of the claim that acts of supererogation go beyond duty. As became clear in the previous section, there are many different ways in which theorists have attempted to give such a definition and the differences between these competing accounts are often fairly subtle. All the accounts examined in the previous section agreed that supererogation should be understood as acts that are worthy of some positive moral evaluation but that are not morally required. In this section I will investigate a definition of supererogation given by Dale Dorsey that departs from this orthodoxy. I will start by presenting Dorsey's definition. I will then explain why this definition fails to provide a satisfactory definition of supererogation.

1.3.1. Dorsey's Proposal

Dorsey begins his discussion of supererogation by outlining what is puzzling about supererogation. Dorsey begins by outlining a traditional definition of supererogation as involving the following necessary conditions:

Permissible not Required: If an act ϕ is supererogatory, ϕ is morally permissible, but is not morally required.⁸¹

Morally Good: If an act ϕ is supererogatory, ϕ is especially morally good or meritorious in comparison to acts that fulfil obligations in a minimal way.⁸²

⁸¹ This is how Dorsey characterizes this part of the traditional view of supererogation (2013 p.356).

⁸² This is slightly different from the way that Dorsey formulates the traditional view but fits with how he introduces the puzzle (2013 p.358).

Of course, as I argued above, this definition is not the best account of the traditional view but it is close enough to be accepted for the purposes of evaluating Dorsey's proposal.

The final piece of the puzzle is the thought that, "in any collection of potential

actions a person might perform, it seems right to say that this person ought to perform the action that is supported by the strongest balance of moral reasons."⁸³ The puzzle, then, is that if what we are required to do is the act that is best supported by the moral reasons then it looks like there can be no acts that meet both *Permissible not Required and Morally Good*, at least if we accept that the act that it would be most good or meritorious to perform is the act we have most reason to perform.

Dorsey's proposed solution to this puzzle is to reject the definition of supererogation given by *Permissible Not Required and Morally Good*. In its place Dorsey proposes the following definition of supererogation:

Permissible not Required II: If an act ϕ is supererogatory, ϕ is rationally permissible, but is not rationally required.

Morally Good II: If an act ϕ is supererogatory, ϕ is *especially* morally good or meritorious in comparison to other rationally permissible actions.⁸⁴

According to this new definition, supererogatory acts are morally good but not *rationally* required. As a result, this definition is compatible with The Good Ought Tie Up, as acts that the *morally* obligatory are those that are best supported by moral reasons. It is the *rationally* obligatory acts that can be morally surpassed.

⁸³ (2013 p.359). Dorsey seems to be using 'ought' as equivalent to 'morally required' here. It is worth noting that many reject that these are equivalent (eg. Chrisman (2012)).

⁸⁴ Dorsey (2013 p.371).

Dorsey also claims that this response to the problem does better than its rivals at handling cases of supererogation that would have been obligatory were it not for the fact that they require a non-trivial sacrifice on the part of the agent.⁸⁵ In the remainder of this section I will argue that Dorsey's account ought to be rejected, regardless of whether it has these advantages. The reason for this is that it is unable to capture the intuitions that push us towards accepting the need for the category of the supererogatory. I will start by examining the reasons that we might accept that acts of supererogation exist. I will then explain why Dorsey's account of supererogation fails to accommodate these intuitions.

1.3.2 Evaluating Dorsey's Proposal

In the introduction we looked at why there is a need to make room for the category of the supererogatory. By making room for this additional category we are able to reconcile the moral goodness of the soldier's act with the thought that it would be inappropriate to demand this act from him or blame him for failing to act in this way. In this section I will argue that, unlike the traditional view of the supererogatory, Dorsey's definition is unable to reconcile the intuitions that created the need for a new deontic category. This, I will argue, creates a dilemma for Dorsey's view.

Let's begin by looking at why Dorsey's definition of the supererogatory fails to reconcile the thought that the soldier's act is morally good with the thought that it would be inappropriate for anyone to demand this act from him or to blame him if he had failed to act in this way. As we saw in §1.3.1, Dorsey defines supererogatory acts as acts that are rationally permissible, not rationally required and especially morally good in comparison to other rationally permissible actions. However, this category of actions fails to make sense of the intuitions that Urmson claims

⁸⁵ See Dorsey (2013 pp. 365-369 and 379-381).

give rise to the need to make room for the concept of supererogation. Accepting that there are acts that are rationally optional and especially morally good in comparison to other rationally permissible actions fails to explain how it is possible for the soldier's act to be morally good yet for it to be inappropriate to demand that he perform it or to blame him for nonperformance. As we saw in §0.1, if we accept the conceptual claims about moral requirements then the inappropriateness of demands or blame gives us reason to think that the soldier's act is not morally required. The puzzle then arises as to how the act can be morally good without being morally required. Dorsey's proposed category fails to provide a solution to this puzzle. It can only make room for actions that are morally good but not *rationally* required. What we need to make sense of though are acts that are morally good but not *morally* required. If proposing a new deontic category is going to enable us to reconcile these intuitions then it is going to have to be a category of acts that are both morally good and *morally* optional. The category of actions that Dorsey proposes fails to resolve the problem that led us to seek out a new deontic category.

It seems, then, that if Dorsey accepts that the reason to make room for the supererogatory is to enable us to make sense of the intuitions considered in §0.1 then he should also accept that his solution fails to accomplish this task. This, though, is not the end of the discussion. Dorsey could respond to this objection by denying that these arguments provide us with good reason to make room for the supererogatory. However, this approach faces a dilemma.

First, taking this approach involves taking on the additional burden of giving some reason to deny both of the arguments put forward by Urmson. For as long as we accept these arguments we

are going to need to make room for a concept fitting the traditional definition of supererogation. Unless Dorsey provides some reason for us to reject these arguments then the need for a category fitting the traditional definition will remain. This is problematic, as the intuitive claims made by the second premise in each argument seem very plausible. It would be hard to deny that no one could reasonably demand that the soldier act as he did or blame him had he failed to do so. Similarly, the conceptual claims found in the first premise seem very plausible. One only need look at the number of supporters of each of the views to see this. In order to respond to these two arguments, then, Dorsey is going to have to deny at least two intuitively compelling claims.

However, even if Dorsey were able to provide good reason to reject these arguments, a further problem remains. If we reject the arguments given in the previous section then there seems little reason to accept Dorsey's account of the supererogatory. By denying these arguments Dorsey would also remove the need to make room for the concept. If we were to find a way to respond to these arguments then the next move would not be to find some other way to understand the supererogatory. Rather, it would be to simply abandon this category, as it would not be doing any work. This then is a dilemma for Dorsey's account. Either Dorsey endorses the claims examined in the previous section or he doesn't. If he does then he must accept that his account fails to solve the problem that the category of the supererogatory is supposed to solve. If he does not then he not only takes on the burden of providing a reason to reject both arguments, he also removes the need to make room for any category of the supererogatory and so his proposed reinterpretation of the concept is redundant.

1.3.3 Objections and Replies

We might think that there is a way of making the second horn of the dilemma more acceptable. Perhaps Dorsey could embrace the latter option by claiming that although Urmson's arguments should be rejected there are similar arguments that could be given to show the need for the category he proposes. If we reinterpret the claims made in the premises of both arguments to refer to rational requirements rather than moral requirements then this would provide us with the following two arguments to show that Urmson's soldier is not rationally required to act as he did.

THE REVISED ARGUMENT FROM DEMAND:

P1 If an act is rationally required then the agent can legitimately be demanded to perform it.

P2 No one could legitimately demand that the soldier dive on the grenade.

C Diving on the grenade is not rationally required.

THE REVISED ARGUMENT FROM BLAME:

P1 If an act is rationally required then the agent is blameworthy for failing to perform it (in the absence of an excuse).

P2 The soldier would not be blameworthy for failing to dive on the grenade, even in if he lacked an excuse.

C Diving on the grenade is not rationally required.

Of course for this response to work it would need to be shown that there is good reason to accept the first premise of these arguments.

This is problematic, as it is far from clear that there is good reason to accept the first premise of either argument. However, even if we think that it can be shown that blame and demands are always appropriate for rational requirements a further problem with this approach remains. The problem is that even if the revised arguments are effective they fail to provide any reason to create a new deontic category that fits the definition that Dorsey provides. While the original arguments showed that we could not divide morality into the required, the permissible and the forbidden these arguments pose no threat to this division. The fact that an act can be morally better than a rational requirement does not give us any reason to propose a new deontic category, it simply tells us that moral requirements are not rational requirements. While this result is interesting, it does not provide any reason to create a new moral category. The tripartite division of morality is unthreatened by either argument. Even if we accept the revised versions of these arguments, there is no reason to accept Dorsey's account of the supererogatory. This response, then, fails to prevent Dorsey's account being impaled on the second horn of the dilemma.

Another response that Dorsey might make to this dilemma is to provide some other reason in favour of making room for this additional deontic category. Perhaps Dorsey could argue that he is not proposing such a category in order to solve the problem posed by Urmson's argument. Rather he is trying to capture a feature of folk moral discourse. It is, after all, a feature of ordinary moral discourse that acts can be 'beyond the call of duty'. Someone sympathetic to Dorsey's view might be tempted to suggest that this account should be seen as an attempt to make sense of this feature of ordinary moral discourse.

The problem with this response, though, is that if this is what Dorsey's account of the supererogatory is setting out to achieve

then it is unsuccessful. In order for Dorsey's account to be a plausible account of the concept being picked out by 'beyond the call of duty' we need to accept both that requirements of practical reasons are a feature of ordinary normative discourse and that these requirements are picked out by the term 'duty'. It could plausibly be argued that the concept of a rational requirement plays some role in ordinary talk and thought. It seems reasonable enough to think that the concept of a requirement of practical reason is being appealed to when people make utterances like, "You must do your homework," or "You must take your medicine." However, it is implausible to claim that it is this concept that is being picked out by the term 'duty'. This can be seen if we substitute 'have a duty to' for 'must' in the previous utterances to read: 'You have a duty to do your homework' and 'You have a duty to take your medicine.' It is clear that this substitution creates a change in meaning between the two sentences, at least if we took the original utterances to refer to rational requirements. These new utterances suggest a moral requirement not a rational requirement. Dorsey's account then is not a plausible articulation of what is meant by the ordinary language phrase 'beyond the call of duty'. This leaves Dorsey's account back in the grip of a dilemma. If Dorsey's account is supposed to articulate what is meant by the ordinary language phrase 'beyond the call of duty' then it is unsuccessful. If this is not what Dorsey's definition is designed to achieve then we are again left without a motivation for making room for this concept in the first place.

In this section I have argued that Dorsey's attempt to solve the problem of supererogation by offering a new definition of supererogation faces a dilemma. I started by looking at the reasons why we might be tempted to make room for supererogation in our ethical scheme. I then argued that these considerations provide no

support for the existence of a deontic category fitting Dorsey's definition. This creates a dilemma for Dorsey's account. If the proponent of this account accepts that these considerations give us reason to make for supererogation then it should also be accepted that this account fails to solve the problem that the category of the supererogatory is supposed to solve. If the proponent rejects these considerations then there is no longer a need to make room for any category of the supererogatory and the proposed reinterpretation of the concept is redundant.

Conclusion

In this chapter I have outlined my account of the definition of supererogatory. I started by arguing against qualified accounts of supererogation that seek to deny the claim that acts of supererogation are optional. I then sought to give an account of the comparative component of the definition. I have claimed that an act, ϕ , is supererogatory for an agent A, at time t , if and only if ϕ -ing is better than the minimum that morality demands of A at t and there is no other obligation that forbids A from ϕ -ing or for which ϕ -ing is the minimum that morality demands of A. I finished by defending this approach against a radical alternative proposed by Dale Dorsey. In the next two chapters I will investigate whether the definition I have given here is an adequate account of the sufficiency conditions for supererogation. I will consider two challenges to the sufficiency of my definition and argue that both challenges fail to provide any reason to supplement it with further sufficiency conditions.

Chapter Two: Are Acts of Supererogation Always Praiseworthy?

Introduction

In April 2012 Cory Booker, the Mayor of Newark New Jersey, returned home to find his neighbour's house on fire. After fighting off a member of his security staff who attempted to restrain him, Booker entered the burning building and saved a woman trapped on the second floor of the house. Both Booker and the woman survived.¹ Like many acts of supererogation it seems to be one that the agent is praiseworthy for performing. Many who have written on the subject of supererogation have claimed that all acts of supererogation are praiseworthy. The view that supererogation and praiseworthiness are necessarily connected is so widely accepted that it has been described as 'The Standard Account' of supererogation.² In this chapter I will investigate whether there is any good reason to hold that such a connection exists. The issue is an important one. If true it will limit the range of acts that can rightfully be classed as supererogatory. However, I will argue that there is no necessary connection between the two.

I will start, in §2.1, by laying out the different views in the literature on this topic. In §2.2 I will examine whether the claim that acts of supererogation are always praiseworthy follows from the analysis of supererogation I gave in Chapter One. I will show

¹ BBC (2012).

² McNamara says that this is part of 'The Standard Analysis' (2011 p.203). Mellema says that this is part of 'The Standard Account' (1991 p17). The view is endorsed by Horgan and Timmons (2010 p. 32), Jacobs (1987 p.97), Mellema (1991 p.17, Montague (1989 p.102), Peterfreund (1978 p.54) and Raz (1975 p.164).

that it does not. I will then, in §2.3, consider and dismiss what I will call ‘The Argument From Stipulation’ in favour of holding that acts of supererogation are always praiseworthy. As part of my response to this argument I will argue that there is a prima facie advantage for allowing acts of supererogation to be unworthy of praise. Given this advantage, the onus is on those who wish to defend the view that acts of supererogation are always praiseworthy to provide a justification for their view. In §2.4 I will consider one such justification in the form of Philip Montague’s argument from intuition that there exists a necessary connection between supererogation and praiseworthiness. However, I will argue first, in §2.5, that there is a debunking explanation that can be given for these intuitions. I will argue that this debunking explanation should be preferred to the claim that there is a necessary connection between supererogation and praiseworthiness for two reasons. First, as I will argue in §2.6, this explanation is favoured by the methodological principle known as ‘Grice’s Razor’. Second, as I will argue in §2.7, the view that there is a necessary connection between supererogation and praiseworthiness gets implausible results in cases of practical deliberation, while the debunking explanation does not.

2.1. The Options

Following Gregory Mellema and Paul McNamara I will call the view that there is a necessary connection between supererogation and praiseworthiness ‘The Standard View’. I will define this claim as follows:

The Standard View: Necessarily, if an act ϕ is supererogatory then an agent who performs ϕ will be worthy of praise for ϕ -ing.

According to this view, in order for an act to be supererogatory the performance of the act must make the agent worthy of praise.

It is important to get clear on what those who defend some form of The Standard View are claiming. First, this is a view about *moral* praiseworthiness. There are many ways in which someone may be worthy of praise as the result of performing some act. It might, for example, have involved an impressive display of skill or willpower. The Standard View should be understood as making a claim about moral praiseworthiness or what is commonly called ‘moral worth’. Second, we might understand this claim to mean that only agents who have all round praiseworthy characters are capable of performing acts of supererogation. This is not how this claim should be understood. The point is that the agent must be considered praiseworthy with regard to the performance of the act in question and what leads the agent to perform it (her motivation, dispositions etc). It is quite possible, on this account, for a villainous person to perform an act of supererogation (providing they are not acting in character).³ All that The Standard View requires from the agent in order for the act to be supererogatory is that she is worthy of praise for having performed it.

Not all accounts of supererogation include an assessment of the agent as well as the act. Chisholm, for example, defines supererogatory acts as those that are, “good but not obligatory to do.”⁴ I will call the view that there is no necessary connection between judgements of supererogation and assessments of the agent The No Necessary Connection View. I will define this view as follows:

The No Necessary Connection View: It is possible for an act ϕ to be supererogatory and for an agent to be unworthy of praise for ϕ -ing.

³ Montague makes this point (1989 p.102).

⁴ Chisholm (1963 p.14).

The purpose of this chapter will be to defend The No Necessary Connection View. I will start by looking at whether praiseworthiness is entailed by the basic components of the definition of supererogation.

2.2. Does Supererogation Entail Praiseworthiness?

In this section I will argue that it does not follow from the analysis of supererogation that I gave in Chapter One that acts of supererogation are always praiseworthy.

Simplifying slightly, in Chapter One we saw that the basic features of supererogation are as follows:

Morally Optional: If an act ϕ is supererogatory then ϕ -ing is neither morally forbidden nor morally required.

Morally Better: If an act ϕ is supererogatory then ϕ -ing is morally better than the minimum morality demands.⁵

The question I will be answering in this section is whether it follows from this that supererogatory acts are always praiseworthy. It is worth noting that this rules out one way in which we might seek to argue against The Standard View. As we saw in Chapter One, Joan Straumanis and Nancy Stanlick both argue that some acts of supererogation are morally wrong for members of certain groups to perform.⁶ I have already argued against these views and so for my purposes in this chapter I will be assuming that an acceptable analysis of supererogation must accommodate both Morally Optional and Morally Better.

We might think that The Standard View follows straightforwardly from Morally Better. If the supererogatory acts are the morally

⁵ The more complex account defended in Chapter One includes a precise account of the minimum that morality demands.

⁶ Straumanis (1984) and Stanlick (1999).

best acts available to an agent then we might think that it follows that they will be praiseworthy to perform. Indeed this would follow if we assume that it is always praiseworthy to perform the morally best act available.

However, while this claim may seem right at first look it should, nevertheless, be rejected. An act can be morally good, even the morally best act available, without being praiseworthy, as praiseworthiness is primarily a feature of agents. It is agents, after all, who are the subject of praise. That is not to say that we cannot describe an act as praiseworthy but when we do so what we really mean is that the agent is worthy of praise for performing the act. Of course, how someone acts plays an important role in establishing whether she is praiseworthy or not. However, in order to establish whether or not someone is praiseworthy we must look to more than simply the act she has performed. It is commonly thought that the motivation behind an action contributes to the praiseworthiness but not the deontic status of the act. As John Stuart Mill puts the point, “the motive has nothing to do with the morality of the action, though much with the worth of the agent.”⁷

Although it is only agents that can be worthy of praise we might think that agents will always be worthy of praise whenever they perform a morally good action. This is also mistaken. It is after all a familiar thought that someone can perform the right action for the wrong reasons.⁸ For example, suppose John is in a situation

⁷ Mill (2001 p.18). What motives we must look for is subject to debate. We might think that the agent must be acting from the motive of duty as Kant (1993) does. Alternatively, we might think that an act is praiseworthy if it stems from good will, as Arpaly (2003 Ch.3) does. Finally, we might think, as Markovits (2010) does, that an act is praiseworthy if the agent’s motivating reasons coincide with the reasons justifying the act’s performance.

⁸ As pointed out by Arpaly (2003 p.69) and Scanlon (2008 pp.56-57). Kant also distinguishes acts that are in line with duty from those that possess moral worth. In order to have moral worth, an

where the right thing for him to do is to save a child who is drowning in a pond. John saves the child from the pond but he is motivated by the thought that doing so will bring about positive media attention. Despite the fact that John performs a good action in this case we would not think that he is praiseworthy. This is because John's action, while morally desirable, is performed for the wrong reasons.⁹

Similarly, as Paul McNamara points out, it is possible to perform an act that goes beyond duty but fails to be praiseworthy.¹⁰ To see this consider the following example:

Cunning Candidate: Jane is a power hungry misanthrope standing for election as mayor. She wants to become mayor so she can more effectively enact her evil plans. Jane passes a burning building and hears a child scream for help. Inspired by the positive media reaction to Booker's act of heroism, Jane recognizes that by running into the burning building she could save the child and also that doing so is likely to win her some votes in the forthcoming election. Jane has no concern whatsoever for the safety of the

act must be performed by an agent who is motivated in the right way (1993 p.11).

⁹ In response to this it might be suggested that there are two ways in which an agent can be deemed to be praiseworthy. One is in virtue of her character traits and the other is in virtue of an act that she has performed. Gregory Trianosky suggests a proposal along these lines for moral blameworthiness. According to Trianosky an agent can be worthy of *deontic* blame for performing a wrong action or failing to perform an obligatory action and worthy of *aretaic* blame for possessing character traits that are the fitting subject of criticism (1986 pp.29-30). Whatever the strengths of this proposal for moral blame I take it that it is not plausible to posit the equivalent claim for moral praiseworthiness. We would not in the case of someone who performs the right act for the wrong reasons think that he is the fitting subject of any form of praise.

¹⁰ (2011 p.208).

child in itself and is motivated solely by a desire to gain votes in the election.

It seems reasonable to think that this act is morally optional and morally better than the alternative acts Jane could perform, such as phoning the fire brigade. Nevertheless, because of Jane's motives we would not think that she is praiseworthy for acting in this way.

Another way in which an act can fail to be praiseworthy is if it is performed accidentally. Consider the following example:

Accidental Donation: Louise makes a large donation that she can comfortably afford to a famine charity by bank transfer. However, Louise donated the money by accident. She had intended to transfer the money between two of her own accounts to enable her to buy an expensive car. Louise would like to retrieve the money but is unable to do so.

The act of donating the money to charity is morally good. It certainly seems to be morally better than buying the expensive car. However, Louise is not praiseworthy for acting in this way, as she did so accidentally.

We might worry that the basic components of the definition of supererogation we have considered in this section miss an important aspect of an acceptable definition. Perhaps saying that an act is beyond the *call* of duty suggests that the agent is able to hear this 'call' which we might think means that an agent must be aware of the moral reasons that support performing the act. We might think that by rejecting the existence of a necessary connection between supererogation and praiseworthiness, The No Necessary Connection View will be unable to make an awareness of the moral reasons part of the definition.

The first point to make in response is that it would be a mistake to read too much into what is suggested by this phrase. The fact that the ordinary language phrase suggests a subjectivist reading does not give us any conclusive reason to hold a subjectivist view of supererogation.

More importantly, however, this objection rests on a confused understanding of exactly what a supporter of The No Necessary Connection View is committed to. First, this view is compatible with a subjectivist view of supererogation, which holds that whether or not an act is supererogatory is determined by the agent's beliefs about her situation.¹¹ We might worry that, while perhaps not incompatible, it would be strange for a subjectivist to hold The No Necessary Connection View. However, this worry disappears when we keep in mind the point that for an agent to be praiseworthy for performing an act she must be motivated in the right way. Subjectivism about rightness is a view about the relationship between an agent's beliefs and the rightness of an action. A subjectivist about rightness can hold a similar view about supererogation whilst maintaining that an agent need not be motivated in the right way in order to be said to have performed an act of supererogation.

Finally, it is compatible with The No Necessary Connection View that acts of supererogation must satisfy some form of intention requirement. A full discussion of the relationship between intentions and supererogation must wait until Chapter Three but for now it is worth pointing out that an act can be performed with morally worthwhile intentions without being motivated in the right way. In the case of Jane, for example, the act is intentional under the description, 'saving a child's life' but is not motivated in

¹¹ For a defence of the subjectivist view of moral rightness see Smith (2010).

a praiseworthy way. Jane certainly seems to intend to perform a morally good act, as this is why she thinks performing the act will be popular with voters. Louise, on the other hand, would not count as performing an act of supererogation on this account, as her donation would not meet this intention requirement.

To sum up this section, performing an action that satisfies Morally Optional and Morally Better is not sufficient for an agent to be worthy of moral praise. In order to be deserving of such praise an agent must perform the act for the right reason. This means that we cannot conclude from the fact that an act is morally better than the available non-supererogatory alternatives that the agent is worthy of praise for performing it.

2.3. The Argument from Stipulation

In the last section we saw that it does not follow from the basic components of the definition of supererogation that all such acts are praiseworthy. One way in which a supporter of The Standard View might seek to respond to this point is to say that we should reserve the term ‘supererogation’ for acts that are praiseworthy. Rather than thinking that it follows from the basic definition of supererogation that these acts are always praiseworthy we might instead simply stipulate that praiseworthiness is part of the definition of supererogation.¹²

Of course, it is not enough for those wishing to defend this view simply to make this stipulation; they need to provide us with some reason to think that we should accept it. McNamara claims that this stipulation is acceptable due to “the absence of any real use of ‘supererogation’ except as a technical term”.¹³ McNamara’s claim is that because ‘supererogation’ is a technical term used only by

¹² This is how McNamara (2011 p.204) explains this possibility.

¹³ McNamara (2011 p.204).

philosophers we are free to stipulate that it should be reserved for acts for which the agent is praiseworthy for performing. By itself this does not give us reason to include it in the definition of ‘supererogation’, just licence to do so if we find a good reason to. Jason Kawall argues that there is good reason to do so because it is, “useful to have a term for actions in which everything goes right.”¹⁴ The thought here is that reserving the term ‘supererogation’ for cases where the agent is praiseworthy will prove a useful distinction that will allow us to quickly identify whether we are talking about acts that merely exceed duty or the subset of those acts for which the agent is praiseworthy. Combining these two thoughts provides us with an argument in favour of including a praiseworthiness condition in the definition of supererogation. If we accept both these claims then it follows that we should restrict the term in this way.

Again the problem with this argument lies not in the reasoning but in the premises. The claim that technical terms should be defined in whatever way is most useful seems to be an acceptable point of philosophical methodology. If philosophers are going to invent a term then they should use it in whatever way will be most useful. However, the other premises are more problematic. There is good reason to reject both of these claims.

The first claim we should reject is that ‘supererogation’ is a technical term used only by philosophers. The term ‘supererogation’ was not invented by moral philosophers but by Christian theologians to refer to those acts that go beyond what is commanded by God.¹⁵ As the term is used in theology to refer to acts that go beyond what is required by duty we should use the term in the same way in moral philosophy. Moreover,

¹⁴ Kawall (2003 p.495).

¹⁵ Heyd (2011).

‘supererogatory’ is used in ordinary language to mean ‘superfluous’.¹⁶ This use of the term is easier to reconcile with the simple ‘beyond duty’ definition and harder to reconcile with ‘beyond duty and praiseworthy’.

The claim that this is the most useful way to define the term can also be rejected. The first reason to do so is that the term has a long history of being used by moral philosophers to refer to all acts that go beyond duty, not just the acts of this kind for which the agent is praiseworthy. We have already seen that this is how the term is used in Christian theology and it is the most common use of the term in contemporary moral philosophy as well. Dale Dorsey¹⁷, Michael Ferry¹⁸, David Heyd¹⁹, Douglas Portmore²⁰ and Justin Weinberg²¹ all define supererogation as acts that are beyond duty without making reference to these acts being praiseworthy. As a result, those who see the value in having a term that refers to a subset of the acts that are beyond duty should coin a new term.

Of course, supererogation also has a history of being defined in terms of praiseworthiness. However, many who endorse this view do so from the mistaken assumption that if an act is beyond the call of duty it follows that it is also praiseworthy. Both Mellema and Cohen, for example, identify supererogatory acts as those that are ‘beyond the call of duty’ and then claim that these acts are

¹⁶ This use of the term is pointed out by Heyd (2011, Section 2 and Section 4). Further support for this claim comes from The Oxford English Dictionary (2013) which gives the following example of this use of the term, which appeared in *Time Out* in 1996: ‘By the time he gets his head bashed in, you hate him so much you could supply a supererogatory kick in the face.’ Thanks to Richard Rowland for pointing out this usage.

¹⁷ (2013 p.356).

¹⁸ (2013).

¹⁹ (1982).

²⁰ (2011 p.91).

²¹ (2011 p.274).

necessarily praiseworthy.²² However, as we saw in §2.2 this is not the case.

Second, the most philosophically interesting questions about supererogation concern the possibility of acts that are beyond duty, not just those for which the agent is worthy of praise for performing. For example, the issue of how to understand the relationship between moral reasons and moral obligations in a way that makes room for the supererogatory is concerned with all acts that go beyond duty. The same is true for the question of how best to reconcile normative ethical theories with the possibility of supererogation. In both cases the issues are concerned with all acts that are beyond duty and not just the subset of these acts for which the agent is praiseworthy.

Finally, a definition of supererogation that does not include praiseworthiness in the definition has the advantage of offering a unified set of deontic concepts. As we saw in the previous section, it is a recognisable feature of our ordinary moral discourse that people can do the right thing for the wrong reasons and as a result deserve no praise for acting in this way. Likewise, many people think it is possible to perform the wrong act but be in no way blameworthy for doing so, if for example the agent has an excuse. This suggests that praiseworthiness and blameworthiness are not part of the definition of rightness and wrongness. There is nothing to stop defenders of The Standard View from accepting this. However, if they do accept this then on their account

²² Mellema (1991 p.16) argues that being ‘beyond the call of duty’ is the core of any analysis of supererogation, before going on to argue that acts that are beyond the call of duty are necessarily praiseworthy. The root of this mistake seems to be that Mellema sees ‘praiseworthy’ as roughly equivalent to ‘morally good’ (1991 p.24). Cohen (Forthcoming p.1) identifies supererogatory acts as those that are beyond the call of duty before endorsing Mellema’s analysis of such acts.

supererogation possesses a necessary connection to an assessment of the agent's motives that moral obligation and wrongness do not. This presents a *prima facie* disadvantage for defining supererogation in terms of praiseworthiness, as it presents a less unified account of our deontic concepts. This view is committed to this additional connection is not in itself problematic but it does seem like a difference that calls for an explanation. If we are to accept the claim that supererogation is different from obligation and prohibition in this way then I think that we need to be given good reason to do so. In the absence of such a reason we should prefer an account that does not posit this divide.

To sum up, it has been suggested that there is good reason to define 'supererogation' as the subset of acts that are beyond the call of duty for which the agent is praiseworthy for performing. I have argued that there are good reasons to reject this argument and maintain the equivalence of supererogatory acts with those that are beyond the call of duty. In responding to this argument we found that a *prima facie* advantage for The No Necessary Connection View is that, unlike The Standard View, it does not posit a divide between supererogation and other deontic concepts. This should not be considered a decisive blow but as a challenge that defenders of The Standard View should seek to respond to. In the following section I will consider a possible justification that might be given for this divide.

2.4. The Necessary Connection Argument

So far we have seen that the claim that acts of supererogation are always praiseworthy does not follow from Morally Optional and Morally Better. We have also seen that we should not accept the claim that we should simply stipulate that supererogatory acts must be praiseworthy. In this section I will evaluate an argument

given by Montague in favour of the claim that acts of supererogation are always praiseworthy.

Montague accepts that it does not follow from Morally Optional and Morally Better that acts of supererogation are always praiseworthy. However, he argues that this gives us reason to reject this as a full analysis of supererogation rather than to reject the claim that acts of supererogation are always praiseworthy.

Montague explains his view in the following:

Something is surely missing from this account of supererogation. For suppose A is an action which is neither required nor prohibited, but which is nevertheless morally valuable in virtue of being particularly beneficent. Since this characterization of A implies nothing about the motives with which it is performed, we can also suppose that the person doing A acts at his own convenience and for his own enjoyment. Should we regard A as supererogatory? I think not – and this because the person who performs A deserves no special credit or praise for having acted. The point is that supererogatory acts are necessarily praiseworthy – a point which is reinforced by the examples cited by supererogationists as paradigms of supererogatory acts, and by the frequent (even if largely unexplained) references to praiseworthiness in discussion of supererogation.²³

Montague makes three points in this passage to support the existence of a necessary connection between supererogation and praiseworthiness. The first is an appeal to intuition. As Montague points out there is something quite odd about claiming that someone has performed an act of supererogation without being praiseworthy for doing so.²⁴ If we were told that someone has performed an act of supererogation we would usually take this to be evidence that the agent is praiseworthy for having acted in this way. To return to the case of Cunning Candidate, suppose we know nothing about the case and we are told the following, ‘Jane

²³ (1989 p.102)

²⁴ Peterfreund makes a similar point (1978 p.55).

performed an act of supererogation.’ It is quite appropriate to take from this utterance that the speaker views Jane as praiseworthy for acting as she did. We would certainly think it misleading for someone to tell us this if she were aware of what motivated Jane to perform her act. Montague’s next claim is that the best way of accounting for this intuitive oddness is by appealing to the existence of a necessary connection between supererogation and praiseworthiness. Finally, positing such a connection enables us to explain why the paradigm cases of supererogation involve praiseworthy agents and why it is so common for people to define supererogation in terms of praiseworthiness.

Montague takes this intuition to provide conclusive evidence for the existence of a necessary connection between supererogation and praiseworthiness. However, as I will argue in the next section, this is not the only way in which these intuitions can be accounted for.

2.5. A Debunking Explanation

Until now, those seeking to provide an explanation for the intuition examined in the previous section have assumed that the only way to do so is to appeal to the existence of a necessary connection between supererogation and praiseworthiness. In this section I will argue that an alternative explanation for these intuitions is that asserting that an act is supererogatory tends to convey via conversational implicature that the agent is praiseworthy for performing it. I will defend the viability of this explanation by showing that the connection between supererogation and praiseworthiness passes all the tests that are commonly used to identify conversational implicatures. I will then, in the following sections, argue that there is good reason to prefer this explanation to that proposed by Montague.

At this point we should consider a point of philosophical methodology introduced by Paul Grice. Grice noted that a common philosophical manoeuvre is to draw conclusions about a word's meaning from the fact that its use would be inappropriate in certain kinds of situation.²⁵ Grice warned that we should be careful about the application of this method, as there may be other reasons, stemming from the general principles of discourse, that can explain why the use of the word would be inappropriate in a given context.

Grice claims that conversations operate according to what he calls 'The Cooperative Principle'.²⁶ This principle states, roughly, that speakers ought, all things being equal, to make their contribution to the conversation such that it fits with the purpose or direction of the conversation. Grice gives four maxims for this principle:²⁷

Quantity: Make your contribution as informative as required and not more informative than is required.

Quality: Do not say what you believe to be false or lack adequate evidence for.

Relation: Be relevant.

Manner: Avoid obscurity of expression and ambiguity and be brief and orderly.

Grice argued that the assumption that speakers are following these maxims gives rise to 'conversational implicatures'. These he defines as follows:

By saying, p, utterer U *con conversationally implicates* q iff:

1. U is presumed to be following the maxims
2. the supposition q is required to maintain (1)

²⁵ (1989 p.20).

²⁶ Grice (1989 p.26).

²⁷ Grice (1989 p.26-27).

3. U thinks the recipient will realize (2).²⁸

These implicatures then are not part of the conventional meaning of the word or phrase used by the speaker but are communicated by the use of the words in a certain conversational context. In addition to these implicatures Grice claimed that there are also ‘conventional implicatures’.²⁹ Like conversational implicatures, these do not form part of the literal meaning of the term. However, unlike conversational implicatures, the implicature is part of the conventional meaning of the term. I will be arguing that praiseworthiness is not part of the conventional meaning of the term ‘supererogation’ or the phrase ‘beyond the call of duty’. This will show, then, that praiseworthiness is neither part of the truth conditional content nor conventionally implicated by ‘supererogation’.

However, these positive arguments must wait for now, as in this section I merely wish to establish that explaining the intuitions considered in the previous section in terms of conversational implicature is a viable alternative to positing a necessary connection. I will do so by showing that the connection passes the tests that can be used to identify conversational implicatures. Grice provides six tests for identifying conversational implicatures:

- a) Conversational implicatures are capable of being “worked out” on the basis, inter alia, of the Cooperative Principle. That is, they are *calculable*.
- (b) Conversational implicatures are *cancellable*.
- (c) Conversational implicatures are *nondetachable*.

²⁸ Grice (1989 pp.30-31) summarized in this way by Levinson (2000 p.15).

²⁹ Grice (1989 p.26).

(d) Conversational implicata are not part of the meaning of the uttered forms. They are *nonconventional*..

(e) Conversational implicata are not carried by what is said, but by the saying of it.

(f) Conversational implicata may be *indeterminate*.³⁰

According to Grice these tests, while not conclusive, are fairly reliable indicators of conversational implicata.

However, the adequacy of these tests has been challenged. Tests (a), (d), (e) and (f) have all been claimed to be of no help in identifying implicatures.³¹ Whereas, (b), (c) and the additional test of *reinforcibility* have all been claimed to be useful ones that should be used. This is not the place for a discussion what the best available tests for conversational implicature are and it is fair to say that the rejection of Grice's first test is controversial. This test is commonly used to test whether a conversational implicature is present.³² To avoid this controversy I will be testing whether the connection between supererogation and praiseworthiness is: 1. Cancellable 2. Non-detachable 3. Reinforcible and 4. Calculable. We are now in a position to apply these tests to the topic at hand.

Test 1 Cancellability

The first test I will apply to the question is that of cancellability. According to Grice, conversational implicatures can be denied without contradiction or misuse of words.³³ While this test is a useful one in testing for conversational implicature things are not always as straightforward as they are in these examples. As

³⁰ Grice (1989 pp. 39-40), compiled in this way by Saddock (1991 p.367).

³¹ Saddock (1991 p.367).

³² For example, see Strandberg (2012).

³³ Grice (1989 p.39).

Saddock points out, the more generalized the conversational implicature the less clear it will be that cancelling the implicature does not involve some form of misuse of words.³⁴ Nevertheless, even highly generalized conversational implicatures will be more cancellable than conventional implicatures.³⁵ We are now in a position to apply the test to supererogation to see whether the implicature can be cancelled. Consider the following utterance:

It was beyond the call of duty for the election candidate to save the child but she deserves no praise for doing so.

As pointed out in §1 there does seem to be something odd about this. For this reason the connection between supererogation utterances and praiseworthiness cannot be one of particularized conversational implicature. However, the implicature is sufficiently cancellable to be counted as conversational rather than conventional. We can see this by comparing it to the following case where praiseworthiness is part of the conventional meaning of the word:

The election candidate became a heroine by saving the child but she deserves no moral praise for doing so.

In this case the speaker would be guilty of misusing the word 'heroine'. When compared to a case of clear conventional meaning such as this one we can see that, the implicature is cancellable, albeit less obviously cancellable than particularized conversational implicature. As a result, applying the cancellability test supports the claim that the connection between supererogation and praiseworthiness is one of generalized conversational implicature.

³⁴ Saddock (1991 p.373).

³⁵ Saddock (1991 p.373).

Test 2 Non-Detachability

The second test for conversational implicature is that of non-detachability. If the connection between supererogation and praiseworthiness is one of conversational implicature then it will not be possible to say something that shares the core meaning of the first utterance but lacks the implicature.³⁶ We can see this by considering the following utterance:

It was beyond the call of duty for the candidate to save the child

If the claim I am defending is true then this will typically convey the following via generalized conversational implicature:

The candidate is praiseworthy.

Now the question we must ask is whether we can make a claim that shares the same core meaning as the utterance without conveying the implicature. The answer is that we cannot. Consider the following utterance:

By saving the child, the candidate did more than was required of her.

Even though a new phrase is used in this example the implicature, that the candidate is praiseworthy for having acted as she did, remains. The non-detachability test also supports the claim that the connection between supererogation and praiseworthiness is one of conversational implicature.

Test 3 Reinforcibility

The third test for conversational implicature is that of reinforcibility. If an implicature is conversational then it should be

³⁶ Grice (1989 p.39).

possible to make the implicature explicit without this explicit statement of the proposition being redundant.³⁷ Let us now look at whether this test supports the claim that the connection between supererogation and praiseworthiness is one of generalized conversational implicature. This test clearly supports my position. We can see why by considering the following:

It was beyond the call of duty for Julie to give money to charity and she is worthy of praise for having done so.

The second half of this utterance gives an explicit statement of the proposition conveyed by implicature without being redundant. This second half of the above utterance makes absolutely clear that the speaker thinks that Julie's act makes her worthy of praise. Applying this test gives us even more reason to think that the connection between supererogation and praiseworthiness is one of generalized conversational implicature.

Test 4 Calculability

The final test is that of calculability. Grice argued that if an implicature is conversational then it should be possible to work out the implicature using The Cooperative Principle.³⁸ Conventional implicatures, on the other hand, are not calculable. Saying an implicature is calculable does not mean that this is the only process by which we come to understand that there is an implicature conveyed. The point is that, even if we grasp the implicature intuitively, if the implicature is conversational we should be able to rationally reconstruct it. We do this by working out what needs to be presupposed about what the speaker intends

³⁷ Saddock (1991 p.374).

³⁸ Grice (1989 p.31).

to communicate in order for the speaker to be said to be following The Cooperative Principle.

Let's start with the uncontroversial claim that a central purpose of moral discourse is to influence how we behave in different situations.³⁹ This partly consists in working out what act to perform. It also involves working out how to respond to people who have acted in certain ways. So when we ask whether killing is wrong we want to know both whether we should kill and how we should respond to those who kill.⁴⁰ I will call these two features of moral discourse 'Act Directives' and 'Response Directives'.⁴¹

With this distinction in hand we are now in a position to point out the conversational context in which the intuition appealed to in §4 occurs. We are told that it would be odd to describe an act as supererogatory when the agent is not praiseworthy for performing it. In this conversational context the primary goal of the conversation is to assess how to respond to this act. This puts us in a position to give a rational reconstruction of the implicature carried by supererogation assertions. Consider the following assertion:

Cory performed an act of supererogation.

³⁹ For a defence of the claim that influencing behaviour is one of the central purposes of moral conversation see Strandberg (2012 p.105–108).

⁴⁰ Scanlon makes a similar point about the two roles moral principles can play (2010 p.23). He claims that moral principles serve as both standards of criticism and guides for action.

⁴¹ Of course, in so far as our responses are themselves acts, Response Directives are directing us to perform acts as well. The point, though, is that while Act Directives direct us to perform the acts being discussed, Response Directives direct us to respond in certain ways to the acts being discussed.

Given what has just be said, we should accept that the primary purpose of this assertion is likely to be response directing. This makes it clear why these assertions commonly carry the implicature that the act is praiseworthy. When we are assessing the deontic status of an act that has already been performed, we are primarily interested in how we should respond to this act. Given this goal, in order to take a speaker of the above assertion to be complying with the maxim of relation, we must suppose that the speaker views Cory as being worthy of praise for performing the act.

The fact that this implicature does not require any special conversational context provides a response to one line of objection that might be raised against my attempt to explain these intuitions without appealing to a necessary connection. It might be objected that explaining the intuition in terms of implicature can only account for the oddness of utterances that class non-praiseworthy acts as supererogatory. However, the original point Montague made was that it would be odd to *regard* acts as both supererogatory and unworthy of praise. The point then is that we need to explain why this would seem odd even in cases where we are considering the oddness of *thinking* an act to be supererogatory rather than *asserting* that it is. By positing a generalized form of conversational implicature, however, we can say that the reason why it seems odd to regard an act as supererogatory but not praiseworthy is that the implicature is so widespread that we have internalized the pragmatic connection between the two.⁴² This explains why it remains odd to imagine someone regarding an act as supererogatory but not praiseworthy. In addition, it is worth noting that the case we are asked to

⁴² A similar point is made in a different context by Strandberg (2011 p.350).

consider is not entirely devoid of context. We are after all asked to consider the application of ‘supererogatory’ in the context of assessing another person’s act. I will return to this point in §2.7 when I give an argument against The Standard View.

Another challenge that might be raised at this point is that nothing has been said to explain how we can explain the other reasons that Montague gives for accepting The Standard View. As we saw in §2.4, in addition to explaining the intuitions we have about the case of Cunning Candidate, The Standard View also explains why paradigm cases of supererogation involve praiseworthy agents and why it is so common for people to define ‘supererogation’ in terms of praiseworthiness.

However, the view that supererogation assertions convey claims about the praiseworthiness of the agent is able to provide adequate explanations for both of these issues. First, given the existence of a generalized conversational implicature between supererogation and praiseworthiness it would be odd if people used examples of agents who are not praiseworthy when attempting to give paradigm cases. The reason for this is that the aim in giving these cases is to clearly illustrate the kinds of act in question. It would be confusing to do so by using examples of the unusual cases where no implicature is present. Second, there are two reasons why people might commonly define supererogation in terms of praiseworthiness. One reason is that many people mistakenly think that a morally good act is necessarily praiseworthy. The second is that people commonly mistake the content conveyed by these implicatures for part of the truth conditional content of the term.⁴³ Given the frequency with which this mistake is made it should come as no surprise that it has been made here.

⁴³ As Grice points out (1989 Ch.1).

In this section I have suggested that rather than viewing the intuition considered in §2.5 as evidence of a necessary connection, we can instead explain this intuition by appealing to the presence of conversational implicature. In the next two sections I will argue that this explanation is superior to that offered by The Standard View.

2.6. Grice's Razor

The first reason to think we should prefer the debunking explanation over the appeal to a necessary connection is given by a methodological principle that Grice termed 'Modified Occam's Razor' but which I will, following Wayne Davis, call 'Grice's Razor'.⁴⁴ Grice defines this as the principle that, "Senses are not to be multiplied beyond necessity."⁴⁵ What this means is that if there is a plausible pragmatic explanation that can be given for some linguistic phenomena then we should prefer this to an explanation that increases a term's semantic complexity. This can be applied in several ways. First, if to account for some linguistic phenomena we must choose between positing a more restrictive use of a term or a pragmatic explanation then we should choose the latter. The same holds for a choice between positing a brute ambiguity and a pragmatic explanation. The justification for this principle is that it is a principle of parsimony. Applying it helps to provide simpler theories by opposing the unwarranted postulation of entities and by ensuring that no theories are postulated to explain something that can already be explained by a theory we are already committed to.⁴⁶

⁴⁴ Grice (1989 p.47). The label 'Grice's Razor' is given by Davis (1998 p.19)

⁴⁵ (1989 p.47).

⁴⁶ These points are made by Hazlett (2007 p.674).

When we apply Grice's Razor to the two explanations we have for the intuition raised in §2.4 we can see that this principle supports the debunking explanation. Accounting for this intuition by positing a necessary connection to praiseworthiness increases the semantic complexity of the term, as instead of one term to cover all kinds of act that go beyond duty we now have two; one that applies to all acts that go beyond what is morally required and one that covers the subset of these acts that the agent is worthy of praise for performing. The debunking explanation, on the other hand gives a pragmatic explanation for this intuition. If we accept that the pragmatic explanation is a viable one then it would be theoretically redundant to accept a more restrictive use of the term 'supererogation' to explain a datum that can already be explained by existing theoretical commitments.

2.7. An Argument Against The Standard View

So far we have seen that there is an alternative explanation that can be given for the intuition that supporters of The Standard View take to be evidence for their view and that given this, Grice's Razor suggests that we should reject the existence of a necessary connection between the two. In this section I will provide an additional argument against The Standard View.

Until now we have considered uses of the phrase in situations where we are being asked to appraise another agent's action. As we have already seen, in these contexts it would be strange to assert that an act is beyond the call of duty but that the agent is not praiseworthy for performing it. However, when we change the conversational context to one where someone is asking for advice about how to act then the connection between supererogation and praiseworthiness appears far weaker. Suppose Cory Booker had asked one of his aides whether he ought to run into the burning building. In response she says, 'That would be beyond the call of

duty.’ This is not an unusual use of the phrase. However, if we accept The Standard View then the aide would be saying something quite extraordinary, as rather than simply informing Booker of the moral status of the action she has also made a speculative predication about what will motivate Booker’s action. After all, this is part of what will determine the truth or falsity of her assertion according to The Standard View. Given that this is an acceptable use of the phrase it seems as if The Standard View is committed either to accepting that the aide is making a prediction in this case or to attributing error to those who use the phrase in this way. Neither is an attractive option.

On the other hand, if we accept The No Necessary Connection View then the aide’s assertion is perfectly understandable. According to this view the aide is simply making a comment about the deontic status of the action and not making any prediction about what would motivate Booker to perform it. If we want a view of supererogation that is able to make sense of a common usage of the term and its ordinary language equivalent then we seem forced to accept that whether or not an agent is praiseworthy for performing an act does not determine the term’s truth conditions.

What this tells us is that The Standard View brings with it implausible results in cases of practical deliberation. While positing a necessary connection between supererogation and praiseworthiness gets the right result in cases where we are assessing another’s conduct, albeit in a theoretically extravagant way, it is implausible to think that such a connection exists in cases where we are asking for advice about our own conduct. Given that the debunking explanation allows us to explain the intuition without bringing about this strange result in cases of

practical deliberation, it should be preferred to The Standard View.

Conclusion

In this chapter I have investigated whether supererogation and praiseworthiness are connected in some way. I argued first that acts that satisfy Morally Optional and Morally Better may be performed by an agent who is motivated in such a way that he is not praiseworthy for acting. It does not follow then from the common understanding of supererogation as going beyond duty that such acts will always be praiseworthy. Moreover, the fact that other deontic concepts are not tied to praise or blameworthiness gives us reason to favour The No Necessary Connection View. I then considered The Stipulation Argument, which I rejected on the grounds that ‘supererogation’ is not a technical term that is used only by philosophers and more importantly is not most usefully defined in terms of praiseworthiness. I then moved on to consider what I took to be the most plausible argument in favour of The Standard View, The Necessary Connection Argument. I responded to this argument by offering a debunking explanation for the intuition used to motivate it. I then gave two reasons to favour this debunking explanation. First, the methodological principle of parsimony known as ‘Grice’s Razor’ speaks against positing a semantic explanation when a pragmatic one is available. Second, positing a necessary connection between supererogation and praiseworthiness gets implausible results in cases of practical deliberation. The conclusion of this discussion then is that there is no good reason to think that an agent who performs an act of supererogation will always be praiseworthy for doing so.

Chapter Three: Supererogation and Intentions¹

In the previous chapter I investigated the relationship between supererogation and praiseworthiness, arguing that it is possible for an act to be supererogatory without being praiseworthy. In this chapter I will investigate the relationship between supererogation and intentions. In particular, I will be investigating David Heyd's claim that in order for an act to count as supererogatory the agent performing the act must possess altruistic intentions.² According to Heyd, acts of supererogation must be performed with the intention of benefitting people other than the agent.³ I will name this feature of Heyd's definition The Altruistic Requirement. This claim is important for the same reason that the discussion of praiseworthiness was important. Accepting this claim will limit the range of acts that can rightfully be classed as supererogatory, which in turn will have implications for the issues I will consider in the second half of the thesis.

I will start this chapter by looking at why we might think that supererogatory acts are performed with altruistic intent. According to Heyd, this requirement allows us to make sense of the meritorious nature of acts of supererogation. I will argue that, even if we accept the need to make room for the meritorious nature of supererogation, there is also good reason to think that Heyd's requirement is overly restrictive. We can accommodate everything that Heyd wants to say in a way that avoids the problematic restrictions of Heyd's view by replacing Heyd's

¹ This chapter is largely based on Archer (2013).

² (1982 p.115).

³ (Heyd 1982 p.136).

requirement with a moral intention requirement. I will finish by pointing out that the reasons that Heyd gives in support of the view that classing an act as supererogatory involves an assessment of the agent's intentions does not, as Heyd suggests, give us reason to include this as a separate condition in my definition of supererogation. Rather, it is presupposed by the definition I have already given.

3.1. Why include The Altruism Requirement in the definition?

I will start by investigating the reasons for making The Altruism Requirement a necessary requirement for acts of supererogation. Heyd argues that this requirement is needed to make sense of the claim that acts of supererogation are meritorious.⁴ Heyd argues that without this requirement, an act could count as supererogatory even if it is completely lacking in moral merit. In this section I will investigate how we should understand merit and explain the importance of merit for understanding supererogation.

Heyd argues that moral merit is an essential feature of acts of supererogation.⁵ By 'merit' Heyd, means something distinct from 'praiseworthiness'. It is unclear exactly how we should understand this distinction. One way that Heyd distinguishes the two is unconvincing. Heyd argues that praiseworthiness can be influenced by a variety of reasons, such as the agent performing the act, the person doing the praising and the consequences that the praise might have, while merit is related more strictly to what

⁴ (1982 p.136).

⁵ (1982 p.132).

is deserved.⁶ This, though, seems to be a strange way to think of praiseworthiness. Those who agree with Peter Strawson's claim that the desirability of praising or blaming someone is the wrong kind of reason to praise or blame them would be unlikely to accept this claim.⁷ Nevertheless I think Heyd is right to think that there is a distinction that can be made here between acts that are praiseworthy and acts that are worthy of some other form of positive appraisal.

A more promising way of making the distinction is suggested by Heyd's claim that the principles of merit, "take into consideration primarily the act itself and not the agent's motives, virtues, strength of will."⁸ This way of distinguishing between the two makes the merit an agent receives from performing an act distinct from her praiseworthiness. While merit is focused only on the act, praiseworthiness is also concerned with the character of the agent. Merit then is concerned with everything praiseworthiness is concerned with apart from that relating to the character of the agent. Given this we might think that an act is meritorious if and only if it would be praiseworthy if performed for the right reasons. In other words, the act would be praiseworthy if it is the appropriate features of the situation that motivate the agent to perform the act. However, this would not explain why merit is supposed to be a form of positive appraisal. Someone who accidentally performs an act that would have been praiseworthy if performed for the right reasons is not worthy of any form of positive appraisal. There must, then, be an intentional component to merit. Given this, I propose the following definition:

⁶ (1982 p.139).

⁷ (1968)

⁸ (1982 p.139).

Moral Merit: An act is morally meritorious if and only if it is an act that is intentional under a description such that it would be praiseworthy if performed for the right reasons.⁹

According to this definition, an agent can be said to have earned moral merit if under a legitimate description of the agent's intentions, the agent would be morally praiseworthy for performing the act if motivated by the appropriate features of the situation. This definition allows us to clearly see the relationship between praiseworthiness and merit. Merit is separate from praiseworthiness, as it is possible for an act to be meritorious without being praiseworthy (when the agent is motivated by the wrong reasons). Of course, it seems reasonable to think that the two are closely related, as it is plausible to think that whenever an agent performs a meritorious act for the right reasons she will be praiseworthy. Nevertheless, for my purposes in this chapter it is sufficient to point out that the two can come apart.

Introducing the idea of merit allows us to associate the performance of a supererogatory act with a special kind of positive appraisal that is distinct from praiseworthiness. It is important to do so, as agents who perform acts of supererogation do seem to be worthy of positive appraisal. When someone performs an act like Autrey's we think it right that he should receive special recognition for his actions.

We are now in a position to see the importance of the Altruistic Intention Requirement. This requirement can, in Heyd's words,

⁹ The terminology 'intentional under a description' was introduced by Elizabeth Anscombe, who noted that a single action can have several different descriptions and only be intentional under some of these descriptions (2000 p.11). For example, an act may meet the descriptions 'sawing a plank' and 'making a squeaky noise with a saw,' but is only intentional under the first (Anscombe 2000 p.11).

“rule out accidental, unconscious, involuntary or self-regarding acts as cases of supererogation.”¹⁰ The Altruism Requirement prevents acts that are lacking in moral merit from counting as supererogatory. We can see why this is important by considering again an example from the previous chapter:

Accidental Donation: Louise makes a large donation that she can comfortably afford to a famine charity by bank transfer. However, Louise donated the money by accident. She had intended to transfer the money between two of her own accounts to enable her to buy an expensive car. Louise would like to retrieve the money but is unable to do so. Julie, on the other hand, also makes a large donation to a famine charity but she does so intentionally.

Although these acts are similar and are just as likely to have morally valuable consequences only one of these acts can be classed as supererogatory. Only Julie’s act can be counted as morally meritorious as only her act is intentional under a description (helping the hungry) that would make it praiseworthy if performed for the right reasons. Louise had no intention of benefitting anyone else and only did so accidentally. This example makes clear the benefits of including The Altruism Requirement in the definition of supererogation. By making it a necessary condition of acts of supererogation that the agent performing the act must intend to bring about somebody else’s good, The Altruism Requirement prevents acts like Louise’s that are lacking in moral merit from being classed as supererogatory.¹¹

¹⁰ (1982 p.116).

¹¹ A related issue is the debate between objectivists and subjectivists about moral obligation. For a defence of the subjectivist view see Smith (2010). For a defence of objectivism see Graham (2010). Objectivists hold that the rightness of performing an act in any given situation is determined by the objective facts of the situation, not the agent’s beliefs about those

3.2 Why Intentions?

We can see, then, the important role that The Altruism Requirement plays in Heyd's definition of supererogation. However, we might wonder whether Heyd's requirement goes far enough. We might think that it is an agent's motivations that are important rather than her intentions. The difference between requiring altruistic intention rather than altruistic motivation is that for an agent to have an altruistic intention, she must aim to bring about some benefit for others. To have an altruistic motive the agent must be motivated by this appreciation of the benefits the act will have for others. In this section I will consider this challenge to a requirement that looks to the agent's intentions rather than motivations. I will argue, in support of Heyd's position, that it is the intentions not the motivations of the agent that are important for an act to be considered supererogatory.

The claim that it is motivation rather than intention that is important for supererogation is made by Sheldon Peterfreund. He argues that altruistic motivation is a necessary condition of

facts. On this understanding of obligation, it does not matter whether or not the agent is aware, or even could be aware, of what the best act would be. Subjectivists, on the other hand, hold that the act that it is right for an agent to perform is determined by her beliefs about the situation. Note that on the objectivist account we can perform a moral obligation without having any awareness that doing so will be the right thing to do. Heyd's point is that whatever we think about moral obligation, the view of the agent plays an important part in assessing whether or not an act can be classed as supererogatory. Heyd's requirement, then, has a subjective component to it. In order for an act to count as supererogatory the agent must believe that it will have good consequences for others or at least be performing the act as an attempt to benefit others. Of course, The Altruism Requirement requires more from agents than subjectivists about moral obligation do. As well as performing the act that is right from a subjective point of view the act must also be performed with the intention of benefitting other people.

supererogatory action.¹² To support this claim Peterfreund gives the example of two doctors who decide to go to a plague-ridden city to help the sick residents. One doctor is motivated by a desire to help those affected by the disease while the other is motivated by a sense of adventure. Peterfreund argues that we should only count the first doctor's act as supererogatory.

The underlying point in Peterfreund's argument is that altruistic motivation is necessary for acts of supererogation because supererogatory acts are necessarily praiseworthy. However, as we saw in the previous chapter, there is no good reason to think that supererogatory acts must be praiseworthy. Peterfreund's argument, then, gives us no reason to allow only altruistically motivated acts to count as supererogatory.

Why then does Heyd think that only acts performed with altruistic intentions can count as supererogatory? The reason is that Heyd, following John Stuart Mill, thinks that intentions form part of the description of the act while motives do not.¹³ In other words, an act can be performed with different motivations and be the same act but the same cannot be said for intentions. An agent in the same situation but acting with different intentions would have performed a different act. As a result, the motivation that led to the act does not alter the moral evaluation of the act; it only alters our evaluation of the agent. As supererogation is an act evaluation this gives us good reason to think that it is merit and intentions that are necessary for supererogation rather than praiseworthiness and motives.

Heyd's approach allows for an attractive symmetry between supererogation judgements and other act evaluations. It seems

¹² (1978 p.55).

¹³ (Heyd 1982 p.137; Mill 2001 p.18 Fn.2). For an explanation of Mill's position here see Ridge (2002).

plausible to think that there is no motivation requirement for right and wrong actions. When we want to know whether someone has performed an act that is right we do not need to know what motivated her to act as she did. Someone with an obligation to pay her taxes performs her duty when she pays her taxes intentionally. It does not alter our evaluation of the act if she is motivated to do so by an appreciation of her duty or by a fear of punishment. Knowledge of the agent's motivations may alter our judgement of the agent performing a right action but it will not alter our judgement of the act itself. Given that both supererogation judgements and rightness judgements are evaluations of acts and not agents, it should be considered a benefit of Heyd's account that it fits with this plausible way of thinking about rightness judgements.

If we accept the way that Heyd and Mill choose to individuate acts then we are in a position to respond to the examples. In the doctor example, assuming that both doctors act with altruistic intentions, we can say that the doctors perform the same act. The different moral evaluations that we might think the two doctors are due is not down to the performance of different acts but down to the features of their character that led to the performance of the act. We can accept that the adventuring doctor is less praiseworthy than the other doctor but maintain that his act is supererogatory. Likewise, we can accept that the election candidate is morally blameworthy but maintain that her act is supererogatory.

Of course, the above response to Peterfreund only works if we are happy to individuate acts in the way that Heyd and Mill choose to. While a complete defence of this way of individuating acts is beyond the scope of this thesis I will briefly explain two reasons why I think this is the right way to do so.

One reason to individuate acts in this way is given by Heyd.¹⁴ He claims that it is often very difficult to determine the underlying motives that lead people to act.¹⁵ Suppose we see someone give money to a homeless person who has asked her to spare some change. It seems hard to pinpoint exactly what would motivate the person to give the money. Of course the recognition that this money will help the person is likely to play some role but there may well be other factors that influence this decision. Perhaps the agent has a desire to appear generous. Discovering what it was that motivated the agent to act in this way will prove difficult. Discovering intentions, on the other hand, is more straightforward. We can be fairly confident that an agent who gives money to charity is aware that some good shall arise from this. If we need to know that the motivations of the agent are altruistic before we can class an act as supererogatory then it will prove incredibly difficult to class any act as supererogatory.

The second reason why this seems to be the right way to individuate acts is that it allows us to explain a familiar feature of moral duties. By not including a motivation requirement, Heyd's definition allows acts of supererogation to fit easily with how we think about moral obligations and prohibitions. It is a common thought that it is possible to perform the right act for the wrong reasons.¹⁶ We can also perform a morally wrong act and be in no way blameworthy for doing so. If we individuate acts according to intentions then we can explain how this is possible. Someone can be motivated to perform acts by the wrong kind of considerations.

¹⁴ A very similar point is made by Mike Ridge (2002 p.66).

¹⁵ (1982 p.187)

¹⁶ For example, Kant distinguishes acts that are in line with duty from those that possess moral worth. In order to have moral worth, an act must be performed by an agent who is motivated in the right way (1993 p.11). Similarly, Ross claims that an act can be morally right but not morally good (2003 p.156).

If we individuate acts in terms of motivations, though, then we can no longer make sense of this, as the agent acting with a different motivation will be performing a different act.

In this section, I have given two reasons to think that Heyd is right to focus on intentions rather than motivations. First, it is much harder to determine motives than it is to determine intentions. Second, it is consistent with the possibility of performing the right act for the wrong reasons.

3.3 Why Altruism is Not Necessary

So far I have argued that The Altruism Requirement plays an important role in ruling out acts that intuitively do not seem to be supererogatory from counting as such. However, in this section I will argue that The Altruism Requirement should be rejected for two reasons. First there seems little theoretical support for restricting the claim to altruistic intentions rather than moral intentions. Second, doing so excludes acts that intuitively should count as supererogatory.

The first reason to replace The Altruism Requirement is that, given the role this requirement plays in the definition, there seems little reason to restrict the claim to altruistic intentions rather than moral intentions in general. We have seen that the reason to accept The Altruism Requirement is that it explains why acts of supererogation are morally meritorious. However, a moral intention requirement would also do this. There is nothing in what Heyd says about merit, nor in my attempt to give a more precise formulation of the idea, that restricts merit to those with altruistic intentions. There seems to be no theoretical reason to restrict the claim to altruistic intentions rather than moral intentions in general.

Nevertheless, we might think that the reason why we should restrict the claim to altruistic intentions is that this is a better fit with our intuitions. Certainly, it does seem to be true that most cases that we think of as supererogatory do involve an agent acting with altruistic intent. Many of the examples of supererogation we have looked at so far are cases where the agent is acting with altruistic intent. Perhaps, then, there is some *prima facie* support for The Altruism Requirement. Nevertheless, this support will be undermined if we can find examples of acts that are intuitively supererogatory but performed by agents with moral but not altruistic intent.

I will give two examples that I think meet this description, one from Mellema and one of my own. To be clear, what we are looking for are cases where we would happily describe the act as supererogatory even if the agent lacks any altruistic intention. If we find a case of this sort then we will have found a counterexample to The Altruism Requirement.

Mellema gives the first counterexample.¹⁷ His example is of a political prisoner being held by terrorists. The prisoner is told by the terrorists to renounce his government and pledge allegiance to their cause. No one will find out about his refusal to cooperate with the terrorists and therefore doing so will not benefit those he cares about in any way.¹⁸ The consequences for him, though, of not cooperating will be terrible, as he will face torture from the terrorists. However, the prisoner is a principled and patriotic man who believes that he should not renounce his government. Out of

¹⁷ (1991 pp.19-20).

¹⁸ In Mellema's description of this example he says that the act will bring about only 'bad consequences' (1991 p.20). I have changed this to 'indifferent consequences' because I believe it makes the example more plausible. This in no way changes the point that Mellema is making.

respect for these principles, and with knowledge of the lack of good consequences for other people that will result from his actions, the prisoner refuses to renounce his government.

Mellema argues that in this example the man performs a supererogatory act in refusing to comply with the demands of the terrorists. The prisoner heroically stands up for what he believes in and does so in a way that exceeds the requirements of duty. However, the act is not performed with altruistic intent. If we accept that the act that Mellema describes is supererogatory then we will have to accept that The Altruism Requirement is not a necessary condition of supererogation.

For the sake of those unconvinced by this example I will give one of my own. Simon Wiesenthal was a Holocaust survivor. After The Second World War, he dedicated his life to tracking down fugitive Nazis so they could be brought to justice. Most famously, Wiesenthal helped to track down Adolf Eichmann, the Nazi bureaucrat who orchestrated the killing of millions of people, leading to his capture in 1960. The reason that Wiesenthal dedicated his life to this cause was not vengeance but justice.¹⁹ By choosing to spend his life in this way Wiesenthal put himself in a position where he would have to suffer both physical attacks from neo-Nazis and the emotional pain of constantly reliving his traumatic experiences.²⁰ Wiesenthal's actions appear to be clear examples of supererogation.

¹⁹ This is why the title of one of Wiesenthal's books is *Justice Not Vengeance* (1989). In this book Wiesenthal's friend, Peter Michael Lingens, tells of the important role that the idea of justice had in shaping Wiesenthal's view of the world from an early age (1989 p.3).

²⁰ Wiesenthal talks of the emotional pain he put himself through here (1989 p.27) and the physical attacks here (1989 p.401).

It is unclear whether Wiesenthal can be said to have intended to benefit others with his act. Perhaps he viewed his acts as safeguarding people against these criminals. Perhaps he viewed his acts as in some way intending to benefit the dead by keeping their memory alive.²¹ Importantly for my purposes, we might think that his intention was not to bring about some benefit to anyone; it was just to bring about justice. I do not claim that this is how we should understand Wiesenthal's intentions, only that it is conceivable to think of his intentions, or those of someone in a similar position, in this way. For the purposes of this discussion I will be assuming that these were Wiesenthal's intentions. The important question we must ask ourselves is whether thinking of Wiesenthal's intentions in this way changes our view about whether or not his acts were supererogatory. Clearly it does not. We would still think of Wiesenthal's acts as supererogatory even if we found out that his intentions were not altruistic.

These two examples give us good reason to reject The Altruism Requirement. Before doing so though, it is worth considering a response that might be made to my argument. Someone seeking to defend The Altruism Requirement might argue that an act cannot be valuable, and so can't be better than the minimum that morality demands, unless it benefits people in some way.²² If we reject the moral value of this act then there is no reason to think that this is an act of supererogation and so no reason to see this as a counter-example to Heyd's definition. For example, a consequentialist may contend that if no positive consequences are expected to arise as a result of the act then according to this view no moral value has been produced. We might think then that we can retain Heyd's requirement by denying the moral value of the prisoner's act. Of

²¹ Wiesenthal suggests that this was the intention behind many of his acts (1989 p.411).

²² Thanks to an anonymous referee for pressing me on this point.

course this response depends upon a certain way of understanding moral value, such that it is the consequences of acts that are important for moral value.

If we accept this then in order to be counted as supererogatory an act must benefit people in some way. This leaves us with a dilemma. Either the acts in the previous examples do not benefit people and so have no moral value or they do benefit other people and so are altruistic after all.

However, I think even if we accept this claim about moral value this response does not present my argument with a real problem. Remember, we are interested in whether the intentions of the agent are altruistic. If we accept the consequentialist understanding of moral value then we can view the agent's act as having good consequences for other people, and being morally valuable as a result, without the agent intending to bring these consequences about.²³ The act can, then, be morally valuable without the intentions of the agent being altruistic.

In addition, many consequentialists claim that their account of what makes an action right or wrong is not intended as an account of what agents should be consciously aiming at when they act. As Henry Sidgwick puts the point: "It is not necessary that the end which gives the criterion of rightness should always be the end at which we consciously aim."²⁴ R. M. Hare, for example, claims that there are two levels of moral thinking.²⁵ At the 'intuitive level' we have a range of moral principles that guide our conduct in our everyday moral behaviour. The 'critical level' is where we evaluate these principles and adjudicate conflicts between them. For example, we might at the intuitive level follow the principle

²³ A similar point is made by Mellema (1991 p. 20).

²⁴ (1907 p.413).

²⁵ (1981).

‘keep one’s promises’ and this is justified, at the critical level, by the plausible thought that a system of promise keeping is generally beneficial. It seems plausible to think that on such an account an act could count as good if the agent’s intentions reflect only the intuitive level of thought. For example, if someone intends only to fulfil her promise rather than to support the institution of promise giving. Even if Wiesenthal was not acting with altruistic intentions, then, his act may still count as morally better than the minimum that morality demands on a consequentialist view if he was acting from principles at the intuitive level of moral thought.

In this section I have argued that we should reject The Altruism Requirement. However, we must not lose sight of the reason why The Altruism Requirement was seen to be an important part of the definition of supererogation. Without this requirement we would be unable to rule out cases such as that of Louise, the unwitting donator to charity, from counting as supererogatory. This may be viewed as a greater cost than having to rule that the prisoner’s act was not supererogatory.

3.4 Moral Intention Requirement

However, abandoning The Altruism Requirement does not mean we are forced to accept that Louise performs a supererogatory act. Instead we can replace The Altruism Requirement with a similar requirement that allows us to label the prisoner’s act as supererogatory but withhold the term from Louise’s act. If the requirement looked for moral intentions rather than altruistic intentions then it would be possible to describe the prisoner’s act as supererogatory but not Louise’s. We should then, replace The Altruism Requirement with a moral intention requirement. Below is a first attempt to formulate this requirement:

Moral Intention Requirement One (MIR1): For an act to be supererogatory the agent must be acting with moral intentions.

Like The Altruism Requirement, MIR1 would allow us to withhold the term ‘supererogatory’ from Louise’s act. As Louise does not intend her act to bring about any moral value it cannot be classed as supererogatory. Julie, on the other hand, does intend for her act to bring about moral value and so can be said to have performed an act of supererogation.

Unlike The Altruism Requirement though, MIR1 would allow the tortured prisoner’s act and Wiesel’s acts (on my interpretation of his intentions) to count as supererogatory. Not only would MIR1 give us the right result in these two examples it would do so in a way that explains our intuitions. The reason why it seemed implausible to say that Louise was performing an act of supererogation was that any moral value that arose from her action was accidental. Likewise, with both the prisoner and Wiesel’s acts we want to say that their intentions do not need to be altruistic in order for their acts to count as supererogatory, so long as they are moral. MIR1 allows us to do this.

However, there is a problem with this way of formulating the moral intention requirement. MIR1 could be read as requiring acts of supererogation to be performed by agents whose intention is an explicitly moral one. There are two problems that might arise from this. The first problem is in cases where an agent intends her act to bring about a state of affairs that seems to be morally valuable but which would not be recognized as such by the agent. For example, suppose Hugo is a nihilist. This does not stop Hugo from performing many morally valuable acts. One day Hugo walks past a burning building and hears shouts for help coming from inside. At great personal risk, Hugo runs into the building to save those trapped inside. If Hugo were asked whether he was

aware of a moral reason favouring his act he would deny it. We might imagine him saying the following: ‘I don’t believe in morality. I am just doing this because I feel sympathy for those trapped inside.’ This seems like an act of supererogation even if Hugo would deny it. This is a problem for any moral intention requirement, as it seems as if Hugo may be said to lack awareness of any moral reason in favour of his action. As a result Hugo’s action would fail to meet MIR1 and so could not be classed as supererogatory.

Another example of an agent who intends his act to bring about a state of affairs that seems to be morally valuable but which would not be recognized as such by the agent is the case of Huckleberry Finn.²⁶ In Mark Twain’s *Huckleberry Finn*, Huck fails to turn in his friend Jim, a runaway slave, despite judging that this is what he ought to do.²⁷ We might think that Huck performs a supererogatory act in this case. Certainly his act seems to be morally valuable and we might think that the risks he is taking would prevent it from being obligatory. Nevertheless, if we require supererogatory acts to be performed by agents with explicitly moral intentions then this act cannot be counted as supererogatory.

The second problem with requiring supererogatory acts to be performed by agents with explicitly moral intentions is that we might think that someone whose intentions are explicitly moral would be a slightly strange moral agent. We might think that such an agent has, in the words of Bernard Williams’ famous objection to utilitarianism, “one thought too many,” (1981 p.18). It would,

²⁶ Thanks to an anonymous referee for suggesting that one example might not be enough here.

²⁷ Twain (1966). For discussions of whether or not Huck Finn can be said to be morally praiseworthy see Jonathan Bennett (1974) and Nomy Arpaly (2003).

perhaps, be more appropriate for an agent to be intending to help others rather than intending to perform morally meritorious acts. The problem that could be raised against MIR1 is that it requires agents to be acting with the wrong kind of intentions.

However, these problems can be avoided if we formulate the moral intention requirement differently. These criticisms only arise if we take a moral intention requirement to require agents to have explicitly moral intentions. There is though, no need to formulate the requirement in this way. Instead, we can say that in order to be supererogatory the agent's act needs to be intentional under a description in virtue of which it is morally meritorious. The agent need not be aware that her act meets this requirement. We should then give a new version of the moral intention requirement that avoids these problems.

Moral Intention Requirement Two (MIR 2): For an act to be supererogatory the act must be intentional under a description D such that D picks out features in virtue of which the act is morally meritorious.

This formulation shares the advantages that The Altruism Requirement and MIR1 have over having no intention requirement. Like MIR1 and The Altruism Requirement, MIR2 allows us to prevent Jane's act from counting as supererogatory.

MIR2 also shares the advantages that MIR1 has over The Altruism Requirement. Like MIR1, MIR2 allows us to say that both Mellema's prisoner and Simon Wiesenthal's acts (on our interpretation) are supererogatory in a way that explains our intuitions. The reason why it seemed implausible to say that Louise was performing an act of supererogation was that any moral value that arose from her action was accidental. Likewise, in the case of the prisoner, the reason why it may be judged that

his act is supererogatory is that he has performed an act that is morally meritorious under the description 'refusing to denounce one's ideals'.

Unlike MIR1 though, MIR2 avoids the objection that arose from requiring agents to have explicitly moral intentions. The problem with both Hugo and Huck was that they were unaware that their acts were morally meritorious. Nevertheless, they both pass MIR2. Hugo's act is intentional under the description 'Saving those in the building' and to perform an act under this description certainly would be morally praiseworthy when performed for the right reasons. Huck's act is intentional under the description 'Allowing my friend Jim to remain free' and again is praiseworthy under this description, providing of course Huck is performing the act for the right reasons. Similarly, the claim that this would require agents to be acting with the wrong kind of intentions also disappears once we put the point in this way, as the agent does not need to possess an explicitly moral intention.

To sum up, including MIR1 or MIR2 in the definition of supererogation is preferable to The Altruism Requirement as these requirements allow acts such as those of Mellema's prisoner and Simon Wiesenthal to count as supererogatory. These requirements are also preferable to a definition lacking in any requirement on the intentions of the agent as it can rule out cases where the agent's act is accidentally morally valuable act. MIR2 is preferable to MIR1, as it does not require agents to possess explicitly moral intentions. As a result, we should replace The Altruism Requirement with MIR2.

One problem that might be raised against MIR2 is that it could allow self-regarding acts to count as supererogatory. If we think that self-regarding acts can be morally valuable then we might think that, unlike The Altruism Requirement, MIR2 would allow

such acts to count as supererogatory. The Altruism Requirement rules out such acts, as it requires agents to be aiming to bring about some good for other people. If we find it implausible that there can be self-regarding acts of supererogation then this might seem to be an unacceptable consequence of replacing The Altruism Requirement with MIR2. If we think that acts of supererogation cannot be self-regarding, and that The Altruism Requirement would explain this while MIR2 would not, then this puts pressure on the claim that we should replace The Altruism Requirement with MIR2.

There are two ways to respond to this criticism. The first is to accept that acts of supererogation can be self-regarding. Jason Kawall puts forward an argument in defence of this position (2003). He claims that self-regarding acts of supererogation are possible. He supports this claim by giving an example of a waiter who wants to go to university. In order to do so the waiter takes on a second job to save up the money to fund his studies. Kawall argues that this act is supererogatory despite the fact that it only provides benefits to the person performing the act.²⁸ If we accept Kawall's argument then compatibility with self-regarding acts of supererogation becomes an advantage rather than a problem. Rather than viewing MIR2 as too permissive we might now view The Altruism Requirement as too restrictive in what acts can count as supererogatory.²⁹

While this does seem like a genuine advantage for my view I do not want my argument to be dependent on Kawall's. In order to convince those who are not persuaded by Kawall's argument I must find some way of showing that those who reject the existence of self-serving acts of supererogation can nevertheless

²⁸ (2003 p.490).

²⁹ Thanks to an anonymous referee for pressing me on this point.

accept the claim that The Altruism Requirement should be replaced by MIR2. My strategy for doing so will be to show that whatever reason we give for wanting to reject Kawall's argument will prevent self-regarding acts from counting as supererogatory regardless of the intention requirement we include in our definition. To be clear, I am not attempting to show that these responses give us good reasons to reject Kawall's argument. All I am trying to show is that the reasons that might be given to reject it will show commitments that prevent self-regarding acts from counting as supererogatory regardless of our intention requirement.

Before looking at these responses though, it is worth pointing out one way of responding to Kawall's argument that cannot be brought in at this stage of the discussion. One reason that might be given to reject Kawall's argument would be because it is incompatible with The Altruism Requirement. Those convinced of the truth of The Altruism Requirement might see this as providing sufficient reason to reject Kawall's view. Note though, that while rejecting Kawall's claim for this reason is incompatible with my view it does not create a new objection to it. Remember that we are considering whether the fact that my view is compatible with the existence of self-regarding supererogatory acts is an independent problem for my view. If the only reason this is problematic is that it is incompatible with The Altruism Requirement then this is not a new problem for my view, it is simply recognition of the fact that my view is different from The Altruism Requirement.

The other way in which someone might try to respond to Kawall's argument would be to claim that the acts described in Kawall's example are not morally better than the minimum that morality demands, as only the agent gains the benefits. Those who think

that morality is concerned with our interactions with other people will not view the waiter's act as morally better than any of the alternatives that would enhance his well-being to a lesser degree. Given this, those who deny the moral value of self-regarding acts will not count such acts as supererogatory regardless of whether or not The Altruism Requirement is part of our definition of supererogation.

There are then two ways of responding to the criticism that MIR2 would make self-regarding acts of supererogation possible. First, we can embrace this consequence of this requirement. Note that if we are persuaded by Kawall's argument then the fact that my view would allow self-regarding acts to count as supererogatory is an advantage for the view rather than a disadvantage. Alternatively, we can say that whatever reasons are given for rejecting Kawall's argument will be ones that ought to prevent such acts from counting as supererogatory regardless of the intention requirement.

3.5 Do we Need Any Intention Requirement?

So far I have argued that if we are going to include an intention requirement in our definition of supererogation then Heyd's Altruistic Intention Requirement is too restrictive. We should instead include a Moral Intention Requirement according to which an act is supererogatory only if it is intentional under a description D such that D picks out features in virtue of which the act is morally meritorious. In the remainder of this chapter I will argue that there is no need to include this requirement in our definition of supererogation.

Let's begin with a reminder of why Heyd claimed there was a need to include an altruistic intention requirement in the definition of supererogation. The point of this requirement was to, "rule out

accidental, unconscious, involuntary or self-regarding acts as cases of supererogation.”³⁰ We have already seen reason to abandon the ‘self-regarding’ part of this definition. The reason Heyd gave to rule out accidental, unconscious or involuntary acts was that he, following John Stuart Mill, views intentions as forming part of the description of the act.³¹ However, if we do individuate actions according to intentions then we have no need to include this requirement as a separate component of our definition of supererogation. This is because part of what the minimum that morality demands on such a view is to act with the appropriate intentions. If the intentions are not appropriate then the act is not right. If the agent performs the act accidentally or unconsciously then this will not count as performing a right act. In order to perform an act that is morally better than the minimum on such a view the agent would need to have the appropriate intentions. Otherwise, as we have seen, the action is wrong. Such a view then presupposes that the agent is acting with the appropriate intentions. There is no need then for those who hold a Millian view of act individuation to include an intention requirement as a separate condition in their definition of supererogation.

Might there be something to be said for including such a requirement as part of the definition in the interest of full disclosure of what is required for an act to be supererogatory? Perhaps if everyone accepted a Millian view of act individuation this would be reasonable. However, not everyone does. Given that my goal is to seek to give a definition of supererogation that is

³⁰ (1982 p.116).

³¹ (Heyd 1982 p.137; Mill 2001 p. 18 Fn.2). For an explanation of Mill’s position here see Ridge (2002).

acceptable to as wide a range of differing views as possible, it would be a mistake to include this condition in my definition.

3.6 Conclusion

In conclusion, in this chapter I have investigated the claim that altruistic intentions are a requirement for acts of supererogation. I have argued that the requirement has some appeal, as it is able to make sense of the meritorious nature of acts of supererogation. Nevertheless, I have argued this requirement should be abandoned as it prevents cases like Mellema's patriotic prisoner and Simon Wiesenthal from being classified as supererogatory. I have argued that by replacing Heyd's Altruistic Intention Requirement with a Moral Intention Requirement we can achieve everything that Heyd sought to achieve in a way that avoids the problematic counter examples. However, accepting that all acts of supererogation will meet this requirement for the reasons Heyd gives does not give us any reason to include this requirement in our definition. For those who accept the Millian view of act individuation, this is presupposed by the existing definition.

Part II Supererogation and Moral Philosophy

In the first half of this thesis, I investigated how supererogation should be defined. Having provided and defended my definition, I will now examine three issues in moral philosophy that up to now have been examined without paying adequate attention either to the existence of acts of supererogation or to the nature of the supererogatory.

I will start, in Chapter 4, by looking at the issue of the relationship between moral reasons, non-moral reasons and moral obligation. The debates around this issue have involved an appreciation of the relevance of supererogation to the discussion but, as I will argue, one of the most prominent positions in this debate rests upon an implausible view about the nature of the supererogatory.

I will then, in Chapter 5, investigate the debate about motivational judgement internalism, the view that there is an internal connection between moral judgements and motivation. This debate is one that has taken place without sufficient appreciation of the need to understand the difference between moral requirements and moral goodness. I will argue that appreciating the difference between the two, as we must if we accept the existence of the supererogatory, gives motivational judgement internalists good reason to restrict their claim to moral requirements. This move might seem arbitrary but can be justified if it is accepted that internalism is explained by the truth of moral rationalism, the view that moral requirements are always in line with what there is most reason to do.

However, as I will explain in Chapter 6 this raises a problem, as moral rationalism is often thought to depend upon the view that moral reasons are overriding. This would be problematic because, as I will argue in §6.1, once we accept the existence of supererogation, moral overridingness ceases to be plausible. However, as I will then go on to explain, it is possible to accept moral rationalism without accepting overridingness. Indeed the explanation for moral rationalism that is compatible with supererogation appears to be better supported by the intuitions commonly appealed to by those seeking to argue for moral rationalism than the explanation offered by overridingness.

Having examined the impact that accepting the existence of acts of supererogation has for these three debates, I will examine the question of whether we need to make room for more deontic categories. Since Urmson's argument in support of the need to make room for the category of the supererogatory a number of philosophers have argued that there is a need to make room for even more deontic categories such as suberogation, quasi-supererogation and forced supererogation. As I will argue in Chapter 7, many of the arguments put forward in support of these claims rest upon mistaken assumptions about the nature of supererogation.

Chapter Four: Supererogation, Sacrifice and The Limits of Duty

Introduction

A noticeable feature of many acts of supererogation is that they seem to involve the agent sacrificing her own interests in order to promote the interests of others. Urmsom's soldier, for example, gives up his life in order to protect his comrades. It is often claimed that all supererogatory acts involve sacrifice.¹ Patricia McGoldrick goes further, describing sacrifice or the risk of sacrifice as, "The distinguishing feature of a supererogatory act."² Like many who associate supererogation with sacrifice, the reasoning behind McGoldrick's claim is that what prevents acts of supererogation from being obligatory is the level of sacrifice required from the agent. However, this claim is rarely defended and it is often unclear exactly what is meant by sacrifice in this context. This is unfortunate, as this claim has important consequences.

First, how we respond to this issue will have consequences for what can count as an acceptable solution to The Puzzle of The Good-Ought Tie-Up.³ The puzzle arises when attempting to reconcile the claim that acts of supererogation are morally

¹ Those who make this claim include Dancy (1993 p.138), Feinberg (1961 p.281), Fishkin (1982), Harwood (2003 p.182), Jackson (1986 p.289,1988), Jacobs (1987 p.101), McGoldrick (1984), Mellema (1991 p.179), Sikora (1979), Straumanis (1984 p.8). Dorsey (2013 p.358) and Raz (1975 p.167) claim that there might be good reason to support this view, without explicitly endorsing it. Rawls (1971 p.117) claims that most, if not all, acts of supererogation involve sacrifice.

² (1984 p.525).

³ As named by Heyd (1982 p. 4).

optional with the claim that they are morally better than the non-supererogatory alternatives. If we think that morally obligatory acts will be the acts that are best supported by moral reason then it is hard to make sense of the existence of acts that are better from the moral point of view. Second, this issue will have consequences for attempts to reconcile the possibility of supererogatory acts with normative ethical theories. While my goal here is not to provide an answer to these two questions, I will finish by examining the implications of my discussion for attempts to resolve it. Third, how we settle this issue will have implications for whether certain acts are classed as obligatory or supererogatory, which in turn may influence how we should respond to these actions.

The purpose of this chapter, then, will be to investigate whether acts of supererogation always involve sacrifice. I will start, in §4.1 by examining different ways of understanding this claim. In §4.2 I will examine some purported counter-examples that have been proposed to this view and explain their limitations. I will then, in §4.3, develop an argument in favour of the view based on brief remarks by Dale Dorsey. In §4.4, I will argue that the view, and the argument in favour of it, should be rejected. Finally, in §4.5, I will respond to objections that might be raised against my argument.

4.1 Two Kinds of Sacrifice Connection

There has been little attempt by those who claim that supererogation always involves sacrifice to explain exactly what they mean by sacrifice in this context. This is a pity, as there are two plausible ways in which we might understand this claim.

We should start by noting that those who claim that supererogation always involves sacrifice are appealing to a notion

of sacrifice that is less rich than the everyday use of the term. The use of sacrifice in this context is restricted to a cost to the agent performing the act. Jonathan Dancy, for example, claims that supererogation always involves sacrifice, which he defines as “cost to the agent”.⁴ We might think that the everyday meaning of the term ‘sacrifice’ involves some further conditions, such as being performed intentionally or voluntarily.⁵ I take it, though, that whatever else we mean by ‘sacrifice’, in order for an act to count as a sacrifice it must involve some cost to the agent’s interests. Given that my aim is to deny that acts of supererogation always involve sacrifice it will be enough for my purposes to show that acts of supererogation do not always satisfy this necessary condition of sacrifice.

The first point to make is that sacrifice is a comparative concept.⁶ When an act involves a sacrifice it makes the agent worse off in some way. The question we must now ask is what the relevant comparison is here. There are two options. One option is that the relevant cost to an agent’s interests is in comparison to her position before performing the act. The alternative is that we take the relevant comparison to be the position the agent would be in if she performed one of the other acts available to her. Clearly it is the second, counter-factual, option that picks out the relevant form of cost. Cases of ‘cutting one’s losses’, where an agent chooses the least costly option from a range of costly alternatives, should not count as cases of sacrifice.⁷ Suppose the victim of a mugging is faced with the choice of handing over his money or being beaten and having his money taken from him. While choosing the

⁴ (1993 p.118).

⁵ See, for example, Overvold (1980 pp.113-114) and Rosati (2009 p.320).

⁶ Thanks to Elinor Mason, Neil McDonnell, Lee Whittington and Alan Wilson for helpful discussion here.

⁷ Overvold (1980 p.108) makes this point.

first option will result in the victim being worse off than he was before, this is an example of minimizing one's losses rather than making a sacrifice. On the other hand, someone can make a sacrifice while making herself better off than she was before. Someone who chooses not to receive the full value of a cash prize but to take part of it and leave the rest to charity, counts as making a sacrifice despite the fact that she is better off than she was before. What makes this a sacrifice is the availability of an alternative option that would have made her better off. The relevant comparison then is to the position the agent would have been in had she chosen to perform one of the other available acts.⁸

⁸ By holding a comparative account of 'sacrifice' I do not mean to commit myself to a comparative account of 'harm', which face problems involving preemption and omission that do not arise for comparative accounts of sacrifice. See Hanser (2008) and Bradley (2012). The preemption problem is that comparative accounts seem unable to class an event as a harm if the agent would have experienced the same cost in the absence of that event. Suppose I am shot seconds before I would otherwise have had a fatal heart attack. It seems counter-intuitive, though, to say that the shooting did not harm me, or so Bradley claims (2012 p.397). This problem does not arise for a comparative notion of sacrifice. In the mugging case it would be counter-intuitive to say that the victim made a sacrifice by giving away his money, as he would have lost it anyway. The omission problem for comparative accounts of harm is that it seems counter-intuitive for some cases where someone fails to receive a benefit to describe it as a harm. For example, if I plan to buy you a birthday present and then change my mind it would seem strange to describe this as a harm. It is not strange, however, to say that someone who fails to provide themselves with some benefit makes a sacrifice. For example, suppose I plan to buy myself a holiday but change my mind and donate the money to charity instead. It seems reasonable to say that I have made a sacrifice. For a defence of the comparative account of harm against these worries see Klockslem (2012). We might also think that, unlike with sacrifice, not all costs count as harms. For example, many have attributed a restricted view to John Stuart Mill (1859). Gray (1996 p.57) argues that on Mill's view it is only costs to vital interests that count as harms while Riley (1998 p.99) argues that the costly effects our actions have on the feelings of others do not count as harms. For a defence of

The next question is whether or not the cost should be understood as an overall cost to the agent's interests. The most obvious way of understanding the sacrifice connection is as a claim about the overall cost to the agent. This view of sacrifice has the following necessary condition:

Self-Sacrifice: An act involves self-sacrifice only if it has an overall negative impact on the agent's welfare compared to some available alternative act.

However, as Vanessa Carbonell argues, often we use the term 'sacrifice' to mean something weaker.⁹ Carbonell argues that we make a sacrifice when we endure a loss that is not compensated for, where a 'compensated loss' is understood as one that is directly replaced without loss. For example, a committed child gymnast who gives up play-dates and ice cream in order to perfect her technique makes a sacrifice even if this has an overall positive impact on her well-being.¹⁰ This notion fits well with a theory of well-being that holds different sources of well-being to be incommensurable. For example, an objective list theory of well-being holds that there are different sources of well-being.¹¹ On such a view it is clear how an act that involves a net gain may involve a loss that is not directly compensated for. It may increase overall welfare but involve a loss to one source of well-being that is not directly made up for or replaced without loss. Carbonell's gymnast, for example, sacrifices one source of well-being,

the claim that Mill did not intend a restrictive view of harm see Turner (2014). Thanks to Elinor Mason, for helpful discussion here.

⁹ Carbonell (2012 p.237).

¹⁰ (2012 p.237).

¹¹ This theory is named and defended by Parfit (1984 p.499). As Fletcher (2013) notes, however, this theory is better described as a pluralist enumerative theory of well-being. That is one that specifies which things enhance well-being and identifies more than one member of this set.

friendship, for another, achievement. However, while this is a legitimate use of the term 'sacrifice', I take it that it would be inappropriate to describe this as a case of 'self-sacrifice'. We would not say that people who make sacrifices in this sense perform acts of self-sacrifice. The gymnast has not sacrificed her own interests for those of other people. Rather, she sacrifices some goods in order to obtain others.

The relevant necessary condition for this weaker version of sacrifice is the following:

Sacrifice: An act involves sacrifice only if it makes the agent worse off in some respect than she would be if she performed some available alternative act.

Although, this notion of sacrifice fits well with an objective list theory of well-being we might think that it fits less comfortably with alternative theories of well-being. A desire-fulfilment theorist, however, could hold that the fulfilment of one desire cannot directly make up or replace another being unfulfilled. Admittedly, it will be harder for hedonists, those think that pleasure is the only source of well-being, to make sense of this kind of sacrifice. Given that I will be arguing against the claim that supererogation always involves sacrifice, I will set this issue to one side. After all, it would be far from satisfactory to reject this notion of sacrifice on the basis of a seeming incompatibility with hedonism.

We can use these two kinds of sacrifice to define two ways of making the claim that supererogation always involves sacrifice.

The Self-Sacrifice View: If an act ϕ is supererogatory then ϕ -ing has an overall negative impact on the agent's welfare compared to some other available act.

The Weak Sacrifice View: If an act ϕ is supererogatory then ϕ -ing makes the agent worse off in some respect than she would be if she performed some other available act.

We might wonder whether these views are intended as conceptual truths or as claims about the extension of the concept. For the most part, it is not stated explicitly which view is being defended.¹² For my purposes, though, this issue is unimportant, as my argument against the view will be effective against both interpretations.

The Self-Sacrifice View is one that is endorsed by a number of authors writing on supererogation. For instance, James S. Fishkin explicitly endorses this in the following:

There are limits to the sacrifice which can be demanded of any individual as a matter of duty or obligation. Beyond these limits, an action is heroic [...] And the presumption is that such heroic behavior must be classified as supererogatory, not obligatory. By 'sacrifice' in this definition I mean a reversal or harm to an agent's interests.¹³

Jason Kawall endorses the weaker view. He argues that cases where an agent suffers some sacrifice to improve her own position can count as supererogatory.¹⁴ Clearly this only makes sense if we take Kawall to be using the weaker notion of sacrifice.

It is worth noting at this point that these two views do not exhaust the ways in which supererogation has been linked to sacrifice. As we have already seen, McGoldrick's view is that it is sacrifice or *the risk of* sacrifice that are distinctive of supererogation.¹⁵ A reasonable assumption to make is that we need to make room for

¹² Dorsey (2013 p.358) notes this lack of clarity. One exception to this is Feinberg (1961 p.281) who states explicitly that this is a necessary truth.

¹³ (1982 pp.14-15).

¹⁴ (2003).

¹⁵ (1984 p.525).

additional views that are concerned with the risk of sacrifice rather than or in addition to sacrifice itself.¹⁶

However, there is little reason to view these as giving independent accounts of the connection between supererogation and sacrifice. Instead, what these views point out is that it is not clear whether the loss mentioned in these definitions refer to an actual loss, the agent's beliefs about loss or what it would be rational for the agent to believe about loss. For this reason, I intend to remain silent about the issue of whether we should have an objective, subjective or prospective view of moral obligations. objectivism, at least consequentialist objectivism, is the view that it is the actual consequences of an act that determine whether or not an act is morally obligatory.¹⁷ prospectivists, on the other hand, hold that it is not the actual consequences but what it would be rational for an agent to believe the consequences will be given her epistemic limitations that are important.¹⁸ Finally, subjectivism is the view that it is the agent's beliefs that determine her obligations.¹⁹

Plausibly, if we think that it is sacrifice that separates supererogation from moral obligation then our view of the relevant kind of loss will be determined by which view of moral obligation we hold. This means that when we are assessing the two sacrifice views given above we must ensure that the arguments given are effective against all three of these views of the relevant form of loss.

4.2 The Sacrifice Views: Initial Assessments

¹⁶ Thanks to Lee Whittington and Alan Wilson for pressing me on this point.

¹⁷ For a defence of this view see Graham (2010).

¹⁸ For a defence of this view about 'moral rightness' see Mason (2013).

¹⁹ For a defence of this view see Smith (2010).

In this section I will examine some purported counter examples to The Sacrifice Views and explain their limitations. As we have already seen, Kawall's defence of self-regarding acts of supererogation involves the rejection of The Self-Sacrifice View.²⁰ Of course, we might take Kawall's claims about the possibility of self-regarding supererogation to be sufficient evidence that we should reject The Self-Sacrifice View. I take it, though, that those sympathetic to The Self-Sacrifice View are unlikely to find this persuasive. Even those who find Kawall's claims persuasive might concede that self-regarding supererogatory acts are not paradigmatic examples of supererogation that any acceptable account ought to be able to accommodate.

Similarly, Michael Ferry also seems to reject The Self-Sacrifice View.²¹ Ferry claims that it is not the case that acts of supererogation always involve *significant* sacrifice to the agent. He supports this claim by giving the following example:

Gift for Friend: You see a book on sale and decide to buy it for a friend. If you buy the book it will bring joy to your friend and the pleasure of giving an unexpected gift will also bring joy to you.²²

Ferry takes this case to show that not all acts of supererogation involve *significant sacrifice*. Unfortunately, Ferry says nothing to defend this claim beyond this appeal to intuition. We should also note that as the example stands we do not have enough information to enable us to say whether this case involves

²⁰ (2003).

²¹ (2013 p.579). Horgan and Timmons also reject The Self-Sacrifice View, in passing, by appealing to a similar example (2010 p.54).

²² This is a paraphrased version of the example given by Ferry (2013 p.580).

sacrifice or not.²³ Remember that making a sacrifice involves performing an act that has an overall negative impact, at least in some respect, on the agent's welfare compared to some other available act. All we are told about this act, though, is that it does not make you worse off than you were before performing the act. To make this a case of supererogation that does not involve sacrifice we must stipulate that there is no alternative act that you could perform that would make you better off either overall or in some respect.

Why might this case give us reason to abandon The Self-Sacrifice View? One reason is that it would seem inappropriate to blame the person who chose not to buy the book for the friend in this example. This is important because it is often claimed that people are always blameworthy if they fail to fulfil their obligations, at least in the absence of an excuse.²⁴ If we hold this view of moral obligations and think that it would be inappropriate to blame the person who failed to buy the book, even if they lacked an excuse, then we should accept that this act is not obligatory. Presumably, though, this act is morally better than whatever we are required to do in this situation. Therefore, the act is supererogatory without involving self-sacrifice.

This argument only undermines The Self-Sacrifice View. Buying the book still involves a financial cost that is not directly compensated for by the pleasure of giving the book. However, we might think that a modified example can raise problems for The Weak Sacrifice View. Consider the following case:

²³ Thanks to Mike Ridge for helpful discussion here.

²⁴ For a defence of this claim see Darwall (2006 Ch.5), Gibbard (1990 p. 40), Portmore (2011 pp.48-49) and Skorupski (1999 pp. 29, 142).

Free Gift For Friend: You see a book on sale and decide to buy it. At the counter you notice a promotional offer where buying one copy of the book allows you to receive a free second copy. You decide to pick up a second copy to give to your friend.²⁵

In this example we might think that there is no cost whatsoever involved in performing the act. Nevertheless, it does not seem like someone who failed to perform this act would be a legitimate target of blame. Perhaps, then, this gives us sufficient reason to reject The Weak Sacrifice View as well.

Of course, a supporter of The Weak Sacrifice View may claim that there are non-monetary costs involved in this case. The burden of having to carry an extra book around is a cost that presumably will not be replaced without loss. Some may wish to insist that once we remove these costs, perhaps by saying that the offer is for a free e-book and so involves no extra carrying, then we should view this act as obligatory. At the very least we might worry that it is hard to generate clear intuitions in cases where the benefits are fairly trivial.²⁶

Alternatively, others might accept that these examples are acts of supererogation but retain some version of The Sacrifice View by claiming that the view is true for a subset of supererogatory acts. Of course, if any acts of supererogation involve sacrifice then it is trivially true that some subset of supererogatory acts do involve sacrifice. However, in order for this view to be interesting the subset has to be one with independently significant features. In both of these examples the acts of supererogation secure fairly trivial moral goods. These are the kind of acts we might be unwilling to describe as ‘beyond the call of duty’ though we

²⁵ Thanks to Elinor Mason for suggesting an example of this sort.

²⁶ Thanks to Mike Ridge for helpful discussion here.

would say that there is no requirement to buy the book in these cases. This might lead some to think that there are two sorts of supererogatory act, those that are fairly trivial and those that are of great moral significance. The former may not involve sacrifice but the latter must involve sacrifice, as this is what is preventing these acts from being obligatory. This division seems to be endorsed by both Portmore, and Horgan and Timmons.²⁷ This would give us the following view:

The Restricted Sacrifice View: If an act ϕ is a *morally significant* act of supererogation then ϕ -ing involves an overall (or not directly compensated for) loss to the agent.

Of course, the phrase ‘morally significant’ is somewhat vague but the thought it is trying to capture is that there are some acts that involve costs and benefits that are fairly trivial from the moral point of view and others that are non-trivial. The question, then, is whether there is some reason to reject The Sacrifice View that does not rely on cases such as Gift For Friend where the benefits are of little moral significance. If there are then this will serve as a response both to the worry that these cases may be obligatory after all and the retreat to The Restricted Sacrifice View.

In this section we have looked at some counter examples that have been raised against The Self-Sacrifice View. We then looked at how we might modify one of these examples to be effective against The Weak Sacrifice View. We might take this to signal the end of the discussion; The Sacrifice Views get the wrong results in these cases, therefore we ought to reject them. However, we saw that there is reason to worry about the force of these

²⁷ See Portmore (2011 p.135 fn.22) and Horgan and Timmons (2010 p.62). This seems to be the case if we take morally significant acts to be ones supported by moral reasons that possess requiring force.

examples, as there appear to be resources for defenders of The Sacrifice View to respond to these counter examples. Moreover, as we shall see in the next section, there is an intuitively compelling argument that can be given in defence of either version of The Sacrifice View. It is safe to say, then, that a convincing argument against The Sacrifice View should do more than simply appeal to these examples.

4.3 How to Argue For The Sacrifice Views

In this section I will look at how a supporter of either version of The Sacrifice View might argue for this view. Unfortunately, despite the frequency with which this view is endorsed, it is seldom seen as a claim that needs to be supported by argument. The closest attempt to give an argument in support of this connection is given by Dale Dorsey. In this section I will develop Dorsey's brief remarks into an argument in support of the view that acts of supererogation always involve sacrifice.

Though Dorsey does not commit himself to the claim that all acts of supererogation must meet this connection, he suggests a reason why we might in the following:

If I am in a position to donate half my yearly salary to Oxfam International, but only at significant cost to my own well-being, doing so is supererogatory. If my donations fail to affect my well-being, or affect it only trivially, making these donations is morally required.²⁸

The point that Dorsey is making here is that it seems reasonable to think that the reason that an act is supererogatory is because it involves an overall cost to the welfare of the person performing it. Without this cost the act would be obligatory. The basic thought that Dorsey is appealing to here is that if we can help others without negatively impacting on our own well-being then it is

²⁸ Dorsey (2013 p.357).

morally required to do so. Likewise, Henry Sidgwick claims that it is part of commonsense morality that people have, “a positive duty to render, when occasion offers, such services as require either no sacrifice on our part, or at least one very much less in importance than the service rendered.”²⁹ We can formalize this thought in the following way:

The No Cost Principle: If an available act, ϕ , is morally better than what would otherwise be the minimum that morality demands and involves no cost to the agent then ϕ -ing is obligatory.

This is an intuitively appealing principle. After all, if we can perform a morally better act at no cost to ourselves then it seems reasonable to think that this is what we ought to do. This also seems like a principle that demands very little of the agent performing the act, as this principle will only apply in situations where doing the morally better act does not involve a cost to the agent.

If we accept this principle then all permissible acts that are morally better than the minimum morality requires will involve a cost to the agent’s well-being that the minimum morality requires does not. Given that acts of supererogation are those that are better than the minimum required by morality it follows that all acts of supererogation will involve sacrifice.³⁰ It is worth noting that this argument remains quiet on whether the sacrifice is the stronger, overall version of sacrifice or the weaker version. This argument,

²⁹ (1907 p.253).

³⁰ This argument has the following form: $P \leftrightarrow ((\sim Q \wedge \sim R) \wedge S)$. $S \wedge \sim T \Rightarrow Q \therefore P \Rightarrow T$. Where P: Act ϕ is supererogatory. Q: Act ϕ is morally obligatory. R: A Act ϕ is morally forbidden. S: Act ϕ is better than the minimum morality demands. T: Act ϕ involves an extra cost to the agent in comparison to the minimum that morality demands. Premise One is the definition of supererogation, Premise Two The No Cost Principle and the conclusion is The Sacrifice View.

then, can be used to support either version of The Sacrifice View, so long as we understand The No Cost Principle to involve the relevant form of cost.

4.4 Against The No Cost Principle And The Sacrifice Views

In §4.2 we looked at some problematic counter examples for The Sacrifice Views. We saw, though, that one way to respond to these counter examples is to retreat to The Restricted Sacrifice View. In the previous section, we looked at another way in which The Sacrifice Views can be defended; by appealing to The No Cost Principle. In this section I will give an argument against The No Cost Principle and all versions of The Sacrifice View.

I will begin my argument with the observation that many people who perform acts that seem like paradigmatic examples of supererogation report that it was in their self-interest to act as they did. Consider the following case:

Free Help Guy: An anonymous London man known as The Free Help Guy (henceforth ‘Guy’) uses the Internet to offer free help to those who get in touch. His acts of kindness include helping a man find his estranged father, helping a man to do DIY and helping several people find jobs. He says that he does it because he enjoys helping people saying, “I wouldn't call this altruism because I think I've got more out of this than anyone else.”³¹

In this case the agent performs what appear to be paradigmatic examples of acts of supererogation. However, if we take Guy's comments at face value then it seems that he believed that his acts would have an overall positive impact on his own well-being compared with the alternatives. Is this claim plausible? We might

³¹ Usborne (2013).

think that this claim is plausible if we hold a subjective view of well-being but less so if we accept a more objective understanding of welfare. A subjectivist about well-being holds that ϕ is intrinsically good for an agent if and only if ϕ is valued under the proper conditions by the agent. An objectivist, on the other hand, holds that there are facts about any person's well-being that are independent of her evaluative perspective.³² If we accept a subjectivist view and we think that the conditions under which moral exemplars possess their desires to perform morally valuable acts are 'proper', then we seem pushed to accept that Guy's helpful acts really do increase his well-being.

The point is perhaps less clear for objectivism.³³ Nevertheless, those who hold this view of well-being typically do not want to claim that features of the agent have no role to play in determining the agent's level of well-being. For example, in T.M. Scanlon's defence of objectivism about well-being he is careful to point out the following:

In speaking of 'objective criteria' I do not [...] mean to exclude the possibility that, according to such a criterion, the same allotment of goods and opportunities may be judged to represent different levels of well-being for two different people because of differences in their condition.³⁴

Of course, on an objective view the differences between these two agents could not be fully reducible to the agents' preferences. However, other features of the agents have an important role to play. Suppose Stuart derives great pleasure from cycling and none from watching football. Laura, on the other hand, derives great pleasure from watching her favourite football team and none from

³² These definitions are borrowed from Dorsey (2012 p.1). For a defence of objectivism see Parfit (2011 Ch.2,3).

³³ Thanks to Elinor Mason and Chris Mills for useful discussion here.

³⁴ (1975 p.658).

cycling. Even if we have an objectivist view it seems reasonable to think that if both had to choose between cycling or watching football, Laura's level of well-being would be enhanced to a greater extent if she watched football while Stuart would be better off cycling. This is consistent with their being some preferences the fulfilment of which will not improve an agent's level of well-being. Someone could have preferences that do not fit with what would actually improve her well-being. Someone might, for example, prefer to count blades of grass rather than read poetry.³⁵ We might think that in this case, the agent's well-being would actually be better promoted by reading poetry.

In fact in order to be at all plausible an objectivist view of the type of costs relevant to determining the limits of moral duty must allow that an act that involves a cost for one agent may involve no cost at all for some other agent. To see why suppose, that Laura and Stuart have both offered to help Polly move house. We might think that, as a result, both are morally obliged to help.

Unfortunately, it turns out that the day Polly needs help is the same day that Laura's favourite football team is playing in a cup final. In this case it seems reasonable to think that this would make it permissible for Laura not to help but the same would not be true for Stuart. Even those who do not think that this makes it permissible for Laura not to help would presumably accept that missing the cup final would count as a cost for Laura and not for Stuart. Of course, this view can remain fully objective, as we could, for example, claim that the reason that missing the match counts as a cost for Laura and not for Stuart is not a result of either's evaluative perspective but rather the actual level of pleasure they will each derive from this activity. Note that this

³⁵ The example of someone who strongly desires to count blades of grass comes from Rawls (1971 p.432).

particular example works even if we do not accept a purely hedonist form of objectivism, one that views well-being as determined by pleasure. After all, objective list theories typically include pleasure on the list of things that contribute to an agent's welfare.³⁶ The point I wish to make, though, is the more general one that any acceptable objectivist view of the type of costs that are relevant in determining the limits of moral duty will allow features of the agent's condition to play a role.³⁷

Of course, we could accept this and yet think that Guy's preferences do not fit with what would actually be in his best interests. This may or may not be true. What is important for our purposes, though, is not to consider whether in the actual case Guy's welfare was best promoted by acting as he did but whether it is conceivable that it could have been. Unfortunately for supporters of The Sacrifice View, it seems hard to deny that this is conceivable. After all, many people who perform acts we think of as supererogatory claim that they would have been unable to forgive themselves if they had not acted as they did. In Samuel and Pearl Oliner's study of 'righteous gentiles', those who helped Jews escape the Nazis in the Second World War, they found that many made this claim.³⁸ In addition, many reported feeling a rewarding sense of inner satisfaction as a result of their actions.³⁹ This suggests that often people who perform highly morally valuable acts are acting in line with their own self-interest. This claim is backed up by psychological research on

³⁶ See Fletcher (2013 p.214), Parfit (1984 p.502).

³⁷ This point is accepted by Hooker (2000 p.43) and Scanlon (1998 p.120).

³⁸ (1988 p.168). A similar observation is made by Munroe (1991 p.404).

³⁹ Oliner and Oliner (1988 pp. 169,177, and 220), Monroe et al. (1990 p.110). Badhwar (1993 p.107) argues that this gives us good reason to think that the rescuers were actually motivated by self-interest.

moral exemplars. In a recently published study, Frimer et al. found that exemplars were significantly more likely than the comparison group to have integrated their personal ambitions with their moral convictions. After ruling out various alternative explanations, the researchers concluded that, “These results are consistent with the claim that moral exemplars have achieved enlightened self-interest, whereby they best advance their own interests by advancing the interests of others.”⁴⁰

Accepting this allows us to say that whether we accept subjectivism or any plausible objectivist account we must accept that when someone with more developed moral sensibilities performs helpful acts it will, all else being equal, promote her own welfare to a greater extent than it would promote the welfare of someone who derives less pleasure from acting in this way. This is bad news for The No Cost Principle. When this principle is applied, those with more developed moral sensibilities will be subject to higher moral standards than the rest of the moral community. This is unacceptable. There would be something deeply unfair about holding those who experience higher levels of psychological discomfort in response to acting in less than morally optimal ways to higher moral standards than other people. To see why, imagine two people in a position to help rescue Jews from the Nazis. One has more developed moral sensibilities than the other. If we accept The No Cost Principle then we are committed to saying that the person with the more developed moral sensibilities is subject to more demanding obligations. Moreover, the reason she has these more demanding obligations is because she has more developed moral sensibilities. Effectively, then, the other’s less developed moral sensibilities get her off the

⁴⁰ (2011 p.160). For further psychological evidence for this claim see Frimer et al. (2009) and Frimer et al. (2012). See also Colby and Damon (1992).

hook from these more demanding obligations. This, though, is simply not the kind of consideration that should prevent someone from facing a moral obligation.⁴¹

Similarly, imagine Guy* is like Guy in every way, apart from the fact that, unlike Guy, Guy* finds helping people boring rather than pleasurable. Both Guy and Guy* are in a position where they could spend the afternoon helping someone. It seems implausible to think that Guy will be subject to more demanding moral obligations than Guy* simply because he has more developed moral sensibilities. Again, though, this is what we are committed to saying if we accept The No Cost Principle.

We might think that this argument only works if we are dealing with an overall sense of cost, rather than the weaker sense explored in §4.1. However, we can avoid this problem if we stipulate that the alternative act available to both Guy and Guy* is a boring afternoon spent watching daytime television. For Guy, then, performing the helpful act will not involve any loss that is uncompensated for. Guy*, on the other hand, finds helping people to be even more boring than watching TV. As a result, The No Cost Principle would generate a duty for Guy but not for Guy*.

So far I have argued that The No Cost Principle is implausible because it would generate less demanding obligations for those with less developed moral sensibilities. We might wonder whether we could accept this argument but hold on to some version of The Sacrifice View. We could, of course, accept that Guy's act satisfies Morally Optional and Morally Better but insist that it

⁴¹ This argument has the following logical form: $P \Rightarrow Q$. $Q \Rightarrow R$. $\sim R \therefore \sim P$. Where P: The No Cost Principle is accepted. Q: Those with more developed moral sensibilities will be held to higher moral standards than the rest of the moral community. R: Lacking a more developed moral conscience gets people off the hook from more demanding obligations.

cannot be supererogatory precisely because it does not involve sacrifice. In other words, while Guy's act is not morally required it is not supererogatory either, as sacrifice is a necessary condition of supererogation. However, there is no benefit to be had in stipulating that we should define supererogation in this way. After all, Guy's act seems like a paradigm example of an act that we would describe as being supererogatory. In addition, this stipulation creates the need for a new deontic category to accommodate acts that satisfy Morally Optional and Morally Better but do not involve self-sacrifice. Making this stipulation would, then, bring about a less parsimonious division of the deontic field without bringing about any obvious theoretical benefits. If we accept this argument against The No Cost Principle then we should also reject The Sacrifice View. Note that retreating to The Restricted Sacrifice View will not help here, since at least the case of helping a man find his estranged father seems like an act that brings about a non-trivial moral good. As a result it must count as a morally significant act of supererogation.

4.5. Objections and Responses

In the previous section I argued against The No Cost Principle and The Sacrifice Views. In this section I will consider a number of objections that might be raised against this argument and respond to each in turn.

The first objection that might be raised is that this account depends on a subjective view of the evidence about whether or not an act involves sacrifice. In other words, it is dependent on the agent's own beliefs about which act will involve a cost to his welfare. We might think that on an objectivist view, where the facts of the matter determine which acts count as involving sacrifice, or a prospectivist view, where what it would be rational for the agent to believe determines which acts count as involving

sacrifice, the argument will not go through. However, as we have already seen, there is good reason to think that these acts actually do improve the well-being of the agents. If this is the case then an objectivist view of the evidence about whether or not an act involves sacrifice would say that these acts do not involve sacrifice. Similarly, if an agent believes that an act involves sacrifice and that act actually does involve sacrifice then it seems reasonable to think that there will be cases where it is rational for her to believe that the act involves sacrifice. If this is right then a prospectivist view about the evidence would also say that these acts involve sacrifice.

Next, it might be objected that the fact that acting in line with the extra moral obligations that exemplars would face is in line with their self-interest means that they are not subject to more demanding obligations. It might be thought that if performing an act is in line with an agent's self-interest then there is nothing demanding about making it obligatory. However, this objection is misguided for two reasons. First, the fact that an act is in someone's self-interest is no guarantee that she will perform it. On this account, when someone with a more developed moral conscience fails to perform an act that is both in line with her self-interest and morally good then she has violated a moral obligation, even though she would not have done so had she had a less developed moral conscience. This is demanding as it makes slip-ups such as these morally blameworthy for the morally developed agent. Similarly, this objection ignores the fact that there may be two or more acts that are equally in line with an agent's self-interest. To make the morally better act morally required in cases like this is to limit the range of permissible acts that are in line with an agent's self-interest.

We might also worry that this argument commits us to the implausible view that possessing greater capabilities never results in being subject to additional moral obligations.⁴² We can see why this would be problematic when we imagine the following case: suppose someone is drowning in the sea near the shore where two people are standing. One is a poor swimmer and the other a holidaying lifeguard. In this case it seems reasonable to think that the lifeguard may have an obligation to save the drowning man but the poor swimmer does not. We might think that this case is analogous to a more conscientious person and someone with less developed moral sensibilities, as both the lifeguard and the conscientious man find it easier to perform the morally good act. We might worry then that if we claim that the more conscientious man should not face additional duties as a result of his more developed sensibilities then nor should the stronger swimmer face more demanding duties as a result of his swimming ability.

While it is plausible to think that the lifeguard does have an additional duty in this case, this causes no problem for my argument. This is because these cases are not analogous. Both the lifeguard and the conscientious man find it easier than those around them to perform the morally good act in these cases. However, there are two different uses of ‘easier’ at work here. The lifeguard finds it easier to save the drowning man because the challenge facing him is a minor one given his abilities. However, there is no reason to think that the conscientious man will find performing morally good acts easier in this sense. The sense of ‘easier’ at work when we say that the conscientious man finds morally good acts easier to perform is simply that, given his more developed moral sensibilities, they involve a lesser sacrifice for

⁴² Thanks to Sebastian Köhler, Geoff Sayre-McCord and Brian McElwee for pressing me on this point.

him. I have argued we that we should not hold people to higher standards as a result of finding morally good acts easier to perform in this latter sense. My claim is that The Sacrifice View ought to be rejected because it has implausible implications for the relationship between *conscience* and duty. This in no way commits me to any view about the relationship between *capability* and duty.

Finally, we might worry that this argument commits us to a strange position when we consider a mirror image of the person with a highly developed moral conscience.⁴³ Suppose a child is falling from a third-floor window and Selfish Steve is standing below. Suppose Steve could catch the child and save her life if he drops the cup of coffee that he is holding (he does not have time to place it on the ground). As an unusually selfish man, Steve would get little pleasure from saving the child and has little desire to do so. In fact he would get much more pleasure from drinking the coffee and desires this much more strongly. Given that I have claimed that any acceptable objectivist view of welfare would allow subjective conditions to play a role in determining what is best for the agent, we might think that I am committed to saying that in this case Steve endures a far higher cost than an ordinary person would.

However, even if we accept that Steve faces a greater cost than an ordinary person would in this case there is no need for me to accept that this will alter the obligations that he faces. My aim here has simply been to argue against The Sacrifice Views. I have not defended a positive account of what it is that makes an act supererogatory as opposed to obligatory. I am not then committed to making any assessments about whether Steve's selfish personality should make him less liable to face certain obligations.

⁴³ Thanks to Elinor Mason for raising this objection.

In fact, we could use the argument I have provided against The No Cost Principle against this assessment as well.

Concluding Remarks

I have argued that we should reject The Sacrifice View, as it is committed to the implausible view that those with more developed moral sensibilities are subject to more demanding moral duties than other people.

This discussion has important implications for how we should evaluate competing solutions to The Paradox of Supererogation. If we accept that not all acts of supererogation involve sacrifice then this is going to be problematic for solutions to the paradox that rely on the balance of the agent's self-interested reasons counting against performing the act.⁴⁴ Given that we also rejected The Restricted Sacrifice View, this discussion will also be problematic for solutions that appeal to this sort of explanation for a subset of supererogatory acts.⁴⁵

This discussion also has implications for how we might accommodate these acts within existing normative ethical theories. Strategies that seek to accommodate the possibility of supererogation by appealing to a clash between what is best for the agent and what is best overall will be unable to explain this kind of supererogatory act. For example, utilitarians who attempt to accommodate the supererogatory by appealing to a permission to increase one's own utility over overall utility in certain situations,⁴⁶ will be unable to allow for acts of supererogation that it is in the agent's interest to perform.

⁴⁴ Eg. Dorsey (2013).

⁴⁵ As both Portmore (2011 p.135 fn.22) and Horgan and Timmons (2010 p.62) do.

⁴⁶ See, for example, Harwood (2003) and Vessel (2010).

Chapter Five Supererogation and Motivational Judgement Internalism

5.1. Introduction

Motivational judgement internalism (henceforth, ‘internalism’) is the view that there is an internal connection between moral judgements and motivation. According to this theory, there is a necessary link between making a moral judgement and being motivated to act in line with this judgement. Motivational judgement externalism (henceforth, ‘externalism’), on the other hand, is the view that there is no internal connection between the two. In this chapter I will investigate a neglected question for internalists, namely which moral evaluations the internal connection holds for.

The debate between internalists and externalists is important in its own right because it has implications for who can be said to be making genuine moral judgements. If we accept internalism then we may have to conclude that someone who claims to have made a moral judgement that does not motivate her cannot be said to have made a genuine moral judgement.

The debate is also important for the role it plays in other metaethical debates. When internalism is combined with a Humean theory of motivation it lends natural support to non-cognitivism about moral judgements.¹ The Humean theory of motivation states that beliefs by themselves are incapable of motivating. If we accept internalism about moral judgements then we accept that motivation is internal to moral judgements and so according to the Humean view of motivation, they cannot be

¹ This point is made by Michael Smith (1994, p.12).

purely cognitive states.² Of course, things are not this simple and it is possible to combine internalism with a cognitivist view of moral judgements.³ An interesting development in recent years has been the attempt to reconcile the two by showing that the assertion of a cognitive state of mind can express non-cognitive attitudes via some form of implicature.⁴ Nevertheless, it is fair to say that anyone attempting to combine internalism with a Humean theory of motivation and a cognitivist view of moral judgements will have some explaining to do.

Many have found internalism attractive because it is able to explain the strong connection that appears to exist between moral language and motivation.⁵ As a number of authors have pointed out, there seems to be something puzzling about someone who claims that an act is obligatory but fails to be motivated to perform it.⁶ Suppose Jenny tells us that eating meat is wrong. It would be puzzling to then find her enjoying a hamburger. Internalists claim that the reason why this seems puzzling is that motivation is internal to moral judgements, so that it is impossible to make a genuine moral judgement without being motivated to act in line with it.

In response to arguments like this one, externalists have argued that amoralists, people who makes genuine moral judgements but

² Although, for all this argument shows, moral judgements could include both cognitive and noncognitive states. Such hybrid views are increasingly popular, see, for example, Ridge (2006) and Tresan (2006).

³ As, for example, Smith attempts to do (1994).

⁴ See Copp (2001), Finlay (2004), Strandberg (2012).

⁵ This way of characterizing the appeal of internalism comes from Strandberg (2012 p.89).

⁶ See Stevenson (1944 pp.16-17), Dreier (1990 pp.13-14), Dancy (1993 p.4), Smith (1994 p.60), Blackburn (1998 pp.48, 52-53) and Lenman (1999 pp. 443-446).

remain unmotivated by them, are conceptually possible.⁷ This would mean that there is no necessary connection between moral judgements and motivation and that externalism is true.

The most common internalist response to the challenge posed by the possibility of amoralists is to restrict the internalist claim to a subset of moral agents. Different kinds of agent restrictions have been proposed. Some restrict the claim to rational agents, others to normal agents and the claim might also be made about virtuous agents. While these agent restrictions help the internalist respond to the challenge of the amoralist, they do so at a price. The price that must be paid for using this strategy is that the further we restrict the internalist claim, the less influence the claim will have for the debate between cognitivists and non-cognitivists. Certainly for some ways of restricting internalism it seems reasonable to worry that what started as an interesting claim about moral judgements, has become a less interesting (for metaethicists at least) claim about certain kinds of moral agent.⁸ Another response that can be given is that the motivation generated by moral judgements is only *pro tanto* and is capable of being overridden by motivation deriving from other sources.⁹

There is, though, an important question that internalists have largely ignored. While there has been a great deal of discussion

⁷ See Brink (1989), Svavarsdóttir (1999), Shafer Landau (2003), Zangwill (2008).

⁸ Miller makes this point convincingly about versions of internalism that are restricted to virtuous agents. This, Miller points out, “Might be the case simply because of what it is to be a ‘virtuous agent’ in the first place,” (2008 p.252). Similar points are made by Enoch (2011 p.251) and Svavarsdóttir (1999 p.183).

⁹ David Brink first made the distinction between Strong Internalism, which holds that moral judgements provide sufficient motivation to act and Weak Internalism, which holds that moral judgements necessarily provide some motivation to act (1989 p.41).

about the kinds of agents internalism applies to and the kind of motivation under discussion there has been almost no discussion about the kinds of moral evaluations this connection holds for. One recent development in the literature has been the explicit recognition of a point that had previously been left implicit; that internalism is a theory about first-person judgements.¹⁰ However, this leaves open the question of which first-person moral judgements internalism applies to. While different definitions of internalism focus on different moral evaluations there has been little discussion about which evaluations the theory should cover. Indeed most contributions in the literature seem oblivious to the fact that different evaluations have been given and the importance that this has for the debate.¹¹ The primary aim of this chapter is to investigate whether internalism is a plausible view to hold about moral judgements in general or whether it should be understood only to be a claim about moral requirements.

I will limit myself in this discussion to focussing on the connection to motivation that internalists have focussed on up to now, that is a motivation to perform the action in question: in other words, being motivated to perform the act I have made a moral judgement about. Other forms of internalism have been defended, such as Allan Gibbard's view that moral judgements are connected to various kinds of emotions.¹² For instance, we might think that judging an act to be morally wrong is internally connected to motivation to blame the person performing the act. While this possibility is an interesting one, for the purposes of this

¹⁰ This point is made by Wedgwood (2007 p.25) and Ridge (2014).

¹¹ One notable exception is Miller (2008 p.235) who acknowledges this gap in the literature.

¹² (1990).

chapter I will focus on the standard approach, which is to link moral judgements to motivation to perform the act being judged.¹³

I will start in §5.1 with an investigation of the different evaluations that internalism has been claimed to cover. I will argue that these various evaluations can be grouped into two categories. One way to understand the internalist claim is that it covers all moral judgements. Another way in which we might understand the internalist claim is as a claim regarding only judgements about the morally required and forbidden. In §5.2 I will explain why it is important to find out which form of internalism can be defended. Unfortunately for the internalist there appear to be problems associated with both options. In §5.3 I will argue that internalists about all moral judgements face the problem of the supererogation amoralist and that this is a problem that traditional amoralist accommodating strategies will be ineffective against. In §5.4 I will examine the problem facing the form of internalism restricted to moral requirement judgements. Those who accept this form of internalism owe an explanation as to why it is only a subset of moral judgements that are internally connected to motivation. However, I will argue that there is an explanation that internalists can give in response to this problem. As a result, internalists should accept that in order to be plausible, the view must be restricted to requiring moral judgements.

5.1 A Neglected Question for Internalists

Internalists have largely ignored the question of what kinds of moral evaluation their theory applies to. Different definitions of internalism make the claim about different forms of moral evaluation. There are, at least, five different forms of moral evaluation that the internalist claim has been made about. Some

¹³ Thanks to Mike Ridge for pressing me on this point.

authors have claimed that the connection holds for judgements of moral obligation.¹⁴ Others have defined internalism as applying to ‘moral ought’ judgements.¹⁵ The claim has also been made about judgements of ‘moral rightness’¹⁶, ‘moral goodness’¹⁷ and about all moral judgements.¹⁸ In this section I will argue that these accounts can be grouped into two categories: those that hold that the internalist claim applies to all moral judgements and those who restrict the claim to requiring moral judgements.

The first way of making the internalist claim refers to moral obligation judgements. For example, Michael Cholbi describes internalism in the following way: “If an agent judges that she is morally obligated to ϕ , then, that agent is, to at least some minimal extent, motivated to ϕ .”¹⁹ This definition could be interpreted as making a claim about all moral judgements and focussing on the case of moral obligation for simplicity. Nevertheless, I think an equally acceptable interpretation would be that it is making a claim specifically about moral requirement judgements. Indeed some authors discuss internalism in such a way that this is the only way to interpret their claims.²⁰ We might think that this view of internalism restricts the claim only to moral obligation judgements. Equally, though, we might think that this claim should be understood to cover negative moral judgements as well.

¹⁴ See Frankena (1958), Korsgaard (1996 p. 81), Cholbi (2011 p.28).

¹⁵ See Kauppinen (2008).

¹⁶ Smith (1994 p.61).

¹⁷ Dreier (1990 p.11) and Blackburn (1984 p.187-8).

¹⁸ See, for example, McNaughton (1988 p.23) and Dancy (1993 p.7).

¹⁹ Cholbi (2011 p.28).

²⁰ For example, Frankena describes the debate between internalists and externalists as: “A problem with the analysis of judgements of moral obligation,” (1958 p.40).

This is how I will understand the claim in this chapter.²¹ Someone who judges that an act is morally forbidden should be motivated not to perform that act in the same way that an agent who judges an act to be morally obliged will be motivated to perform it. I will define this view as follows:

Restricted Judgement Internalism (Restricted Internalism): If an agent judges that she is morally required to ϕ then she will necessarily be motivated to ϕ . If an agent judges that she is morally required not to ϕ then she will necessarily be motivated not to ϕ .²²

Other ways of defining internalism seem to be saying something quite different. These definitions claim that all moral judgements are necessarily connected to motivation. I will define this form of internalism as follows:

Unrestricted Judgement Internalism (Unrestricted Internalism): If an agent judges that it would be morally good for her to ϕ then she will necessarily be motivated to ϕ . If an agent judges that it would be morally bad for her to ϕ then she will necessarily be motivated not to ϕ .

The claim that moral judgements are connected to motivation should be understood in the same way that Restricted Internalism has been. Positive moral judgements will motivate people to

²¹ My reason for doing so is that my goal in this chapter is to show that we must restrict internalism to requiring moral judgements. Whether or not the internalist claim needs to be restricted even further to cover only obligation judgements or only prohibition judgements is not a question I will deal with here. Thanks to Alan Wilson for pushing me to consider this issue.

²² It should be noted that I am using ‘restricted’ here to refer to restrictions on the kinds of judgement the theory covers. There are of course other ways of restricting internalism, such as to certain kinds of agent. I considered these forms of restriction in §5.3.

perform the acts whilst negative moral judgements will motivate people not to perform them. The difference between the two forms of internalism I have outlined is that Unrestricted Internalism holds the internalist claim to cover all moral judgements while for Restricted Internalism the theory only applies to judgements about the morally required and forbidden. Restricted Internalism makes no claims about acts that are given other kinds of moral evaluation.

The remaining claims about the evaluations internalism covers (those concerning moral ought judgements, moral rightness judgements and moral goodness judgements) can be divided into these two classes. Internalism about moral goodness should be placed with Unrestricted Internalism. In the same way that Restricted Internalism was taken to include judgements about the morally wrong as well as the morally required I will take this form of internalism to apply to ‘moral badness’ judgements as well as judgements relating to goodness. This form of internalism claims that there is a necessary connection between judging that an act is morally good and being motivated to do it. Likewise, there is a necessary connection between judging that an act is morally bad and being motivated not to do it. This form of internalism is making a claim about all forms of moral evaluation.

On the other hand, it seems reasonable to think that those who advocate internalism about ‘moral ought’ are referring to Restricted Internalism. When confined to moral contexts about how, morally, someone should act I take it that ‘ought’ means ‘is morally required to’.²³ More controversially perhaps, definitions

²³ Zimmerman uses the term ‘ought’ to refer to moral obligation (1996 p.1). Chrisman (2012), on the other hand, points to some problems with understanding the moral ‘ought’ in terms of obligation. More general worries about the identification of

of internalism that propose a necessary connection between ‘moral rightness’ judgements and motivation should also be included under Restricted Internalism. To say that a particular act is morally right implies that the agent is under obligation to do it.²⁴ For my purposes it does not matter whether I have grouped these forms of internalism into the appropriate categories or not. All I have sought to show is that these five forms of internalism can be reduced to two.

5.2 Why The Distinction Matters

This distinction is important because Unrestricted Internalism covers a wider range of judgements than Restricted Internalism. According to Unrestricted Internalism all moral judgements, with the exception of judgements of moral neutrality, will be necessarily connected to some form of motivation either to perform or not perform the act. According to Restricted Internalism, it is only requiring moral judgements that will motivate. Accepting Unrestricted Internalism, then, will commit us to the view that there is a necessary connection between judgements of supererogation and motivation. Accepting Restricted Internalism brings with it no such commitment. In this section I will explain what it is to judge that an act is supererogatory and why Unrestricted Internalism is committed to internalism about supererogation judgements.

As I have already discussed, a supererogatory act is one that is morally better than the minimum morality demands. When we judge an act to be supererogatory then we judge that it is morally

‘ought’ with ‘is required to’ are raised by McNamara (1996), von Fintel and Iatridou (2008), Finlay (2010 p.76) and Ridge (2014).

²⁴ This claim is made by Ross (1930 p.6). We might equally think, as Scanlon does (1998), that ‘morally right’ means ‘morally permissible’.

optional and morally better than the available non-supererogatory alternatives. As a result, if we accept Restricted Internalism then we should think that the internalist claim does not apply to these judgements. Unrestricted Internalism, on the other hand, is committed to a view of Internalism that holds that there is a necessary connection between supererogation judgements and motivation.

This, then, shows an important difference between Unrestricted Internalism and Restricted Internalism and it is one that presents problems for the defender of either view.²⁵ The problem for defenders of Unrestricted Internalism is that they are committed to the view that supererogation judgements are necessarily connected to motivation. This is problematic, as positing an internal connection to motivation for these judgements seems far less plausible than for other moral judgements. Moreover, as I will argue in the next section, the traditional internalist strategies for accommodating cases of amoralism seem incapable of saving internalism here.

However, defenders of Restricted Internalism also face a problem. Given that they wish to restrict the internalist claim to a narrow subset of our moral judgements, they owe an explanation for this. This problem is made even worse when we consider that it is commonly thought that judging an act to be supererogatory involves judging the act to be morally better than the merely required alternative. Nevertheless, as I will argue in §5.4, there is an explanation that supporters of Restricted Internalism can give

²⁵ It is possible that judgements about the offensive or suberogatory might cause similar problems for the internalist. These are judgements that an act is morally bad but not morally forbidden. Both Chisholm (1963) and Driver (1992) argue that such acts exist. For the purposes of this chapter I will look only at positive moral judgements and the problems caused by judgements of supererogation.

here. As a result, I will argue that those wishing to defend internalism should defend Restricted Internalism and not Unrestricted Internalism.

5.3 Unrestricted Internalism and The Supererogation Amoralist

The problem that Unrestricted Internalism faces is that of the supererogation amoralist, someone who makes genuine supererogation judgements but remains unmotivated by them. This is problematic, as it seems possible to judge that an act is supererogatory and remain completely unmotivated to perform it. Suppose Judy says the following: ‘It would be beyond the call of duty for me to give away all of my disposable income to charity.’ Now imagine we find out that she is not motivated to do so. This does not seem odd in the way that Jenny’s eating meat seemed odd. The reason for this is that we do not expect people to be motivated by a judgement that an act is supererogatory. Suppose we asked Judy why she isn’t acting in line with her judgement. A perfectly reasonable response for Judy to make is that she judges it to be supererogatory not obligatory.

Let me say a little about why I take this to be a particular problem for Unrestricted Internalism and not simply another battle in what David Enoch has called ‘The Amoralist Wars’.²⁶ The first reason is that, unlike traditional amoralist objections, being unmotivated by a supererogation judgement is likely to be a recognisable feature of many people’s moral experience. Certainly, when reflecting on my own experience of judging an act to be beyond the call of duty I would go so far as to say that the occasions where these judgements are accompanied by motivation are the exception rather than the norm. As I write this I judge that it

²⁶ (2011 p.250).

would be morally better for me to stop writing and go and help out at a soup kitchen but I am not motivated to do so and do not find this particularly odd. Of course, we do not need to go as far as asking whether or not someone who is never motivated by supererogation judgements exists. All we need to know is whether it is possible for someone to sometimes be unmotivated by supererogation judgements.²⁷ Nevertheless, the fact that this seems like a recognisable feature of moral experience means that internalists are going to have a difficult job persuading us that supererogation amoralists are necessarily impossible. One difference, then, between this challenge and traditional amoralist challenges is that the supererogation amoralist is more than a fantastical creature posited for the purposes of a thought experiment. It is a condition that many ordinary agents will have experienced in their own lives.

The second reason why this is a particularly important problem for internalists is that the connection between moral language and motivation that is used to provide support for internalism does not apply for talk of supererogation judgements. When Judy from the previous example utters the sentence: ‘Donating my disposable income to charity would be beyond the call of duty.’ This sentence does not seem to bring with it the presumption that Judy will be motivated to donate her money. If anything it seems that our expectation would be that she would not donate the money. Note, as we have already seen, the same is not true about obligations. If

²⁷ Zangwill says that we should not describe people who are indifferent on a particular occasion as amoralists, since amoralists are people who are never motivated by moral judgements (2008 p.101). To be clear, when I use the term supererogation amoralist I mean someone who on that particular occasion is left unmotivated by a supererogation judgement. Those who think that this is a misuse of the term can replace ‘supererogation amoralist’ with ‘someone indifferent to their supererogation judgement’.

Judy were to say instead that she viewed donating the money as morally obligatory then it would be somewhat puzzling if she lacked motivation to do so. So, while internalists can support their claim about moral obligations by an appeal to the connection that it seems odd to claim that an act is obligatory without being motivated by it they are unable to make an equivalent argument for supererogation judgements.

The challenge, then, for those seeking to defend Unrestricted Internalism is to explain the possibility of supererogation amoralists in a way that is compatible with the theory. In the remainder of this section I will consider two ways in which internalists have responded to previous amoralist challenges. I will argue that neither provides an effective defence for Unrestricted Internalism against the problem of supererogation amoralism.

The first response to consider is whether some form of agent-restriction can deal with this problem. To test this possibility we must ask whether for any of the available agent restrictions, it is true that everyone who meets the requirements of the restriction and makes a supererogation judgement will be motivated to act in line with it. There are three forms of agent-restriction that have been proposed in the literature and I will look at each of these restrictions in turn. I will argue that none of these restrictions can provide a plausible response to the problem of the supererogation amoralist.

The first agent-restricted form of Unrestricted Internalism I will consider is one restricted to practically rational agents.²⁸ For this to be an effective response to the problem we will need to be persuaded that practically rational agents will necessarily be

²⁸ This form of restriction is advocated by Smith (1994 p.61), Wallace (2006) and van Roojen (2010).

motivated to perform acts that they judge to be supererogatory. I take it that many will share the thought that this is an implausible claim to make about supererogation judgements. Suppose someone judges that donating all of her money to charity would be supererogatory. It seems absurd to think that she is irrational if she is not then motivated to act in line with this judgement. This I take to be the case whether we opt for a procedural or a substantive account of practical reasoning. Proceduralists think that we can only be rationally blameworthy for failing to have a desire we could have rationally reached from our existing set of beliefs and desires. According to substantivists we might be rationally blameworthy for failing to have some desire even if it would not have been possible for us to rationally reach it through our existing beliefs and desires.²⁹

I take it to be uncontroversial that proceduralists about practical reason should accept that supererogation amoralists might not be irrational. There seems no reason to think that everyone who judges an act to supererogatory could rationally reach the motivation to perform the act from his or her existing beliefs and desires. As a result, internalists are going to have to look to a substantive view of rationality to declare supererogation amoralists as irrational.

The problem with this, though, is that it is plausible to think that many cases where an act of supererogation is available are cases where there are rational options.³⁰ These are cases where two rationally permissible options are available. In fact this seems like a feature of supererogation that any plausible account must be

²⁹ For a discussion of the literature on proceduralism and substantivism about practical reason see Hooker and Streumer (2004). Thanks to Elinor Mason for helpful discussion here.

³⁰ This claim is made by Gert (2004 p.106), Horgan and Timmons (2010) and Portmore (2011 p.153).

capable of incorporating. If it is rationally permissible not to perform the supererogatory act then surely it will also be rationally permissible not to be motivated to perform the act. So declaring supererogation amoralists as irrational using a substantive account of irrational is not going to be an option for accounts that allow for rational options in cases where one or more available acts are supererogatory.

Another problem with attempting to declare supererogation amoralists as substantively irrational is that it seems to rule out as incoherent philosophical positions that appear to be intelligible.³¹ Take, for instance, the view that Susan Wolf defends in her influential article ‘Moral Saints’.³² One of the claims that Wolf defends here is that the life of a moral saint, someone who always acts in the morally best possible way, is an undesirable one. Presumably, those who accept this line of thought would also accept that it is not irrational to be unmotivated by a supererogation judgement. After all, those who are always motivated to perform acts they judge to be supererogatory will, according to Wolf, lead a worse life than those who are not. Of course, not everyone endorses the view Wolf defends here and we might think that the force of this point is contingent upon its acceptance.³³ However, even if we are not persuaded by Wolf’s claims, if we find her view intelligible then it is hard to see how it could be a conceptual truth that rational people will be motivated by their supererogation judgements. If this were the case then it would surely appear incoherent. For these two reasons it seems reasonable to conclude that restricting the internalist claim to rational agents will not provide supporters of Unrestricted

³¹ Thanks to Mike Ridge for suggesting this point.

³² (1982).

³³ One way to put pressure on this view is to reject this account of ‘moral sainthood’ as Carbonnell (2009) does.

Internalism with an acceptable response to the possibility of supererogation amoralists.

The next agent-restricted form of Unrestricted Internalism I will consider is one restricted to virtuous agents.³⁴ In order for this to be an effective response to the problem it must be the case that virtuous people will always be motivated by their supererogation judgements. This claim strikes me as implausible. It seems perfectly possible for a virtuous agent to be unmotivated by a judgement that an act is supererogatory.

For those who don't immediately share this thought it is important to bear in mind that we are concerned with judgements of supererogation rather than acts. I am not claiming that it is possible for a virtuous agent never to be motivated to perform a supererogatory act. All I am saying is that a virtuous agent need not be motivated to perform any acts that she judges to be supererogatory. This is a much more plausible claim, as there may be some people we consider virtuous because they have a very demanding view of what is morally required. In fact, many people who do perform acts of supererogation claim to have only performed their duty.³⁵ Take the example of John Weidner, who during the Second World War put himself at great risk while working for the Resistance helping Jews escape the Gestapo. When asked whether he had performed a particularly worthy act his response was, "Absolutely not. I did my duty. That is all."³⁶ Many might think that Weidner is wrong about this and that his

³⁴ This form of agent restriction is considered, though not defended, by Miller (2008 pp.250-252). Though I know of no one who defends this form of internalism it seems worth considering here, as it might be thought to be particularly relevant to the case of supererogation.

³⁵ This phenomenon is identified by Colby and Damon (1992), Badhwar (1993), and Carbonell (2012).

³⁶ This quote comes from Monroe (2004 p.117).

acts should be seen as supererogatory. Now let's imagine that Weidner does judge some acts to be supererogatory, perhaps he spent six days a week working for the resistance and judged that to spend a seventh would be supererogatory. If Weidner was occasionally left unmotivated by a supererogation judgement would that mean that he could not be classed as a virtuous person? Clearly not, after all he is motivated to perform acts that many people would consider to be supererogatory, he just happens to think that they are obligatory. It is possible, then, for a virtuous agent to be unmotivated by a supererogation judgement.

Those sympathetic to this form of agent-restricted internalism might respond by saying that in this context we must understand 'virtuous' to mean 'fully virtuous'.³⁷ This form of internalism avoids the problem considered above, as it is compatible with those typically regarded as saintly failing to be motivated by their supererogation judgements. However, it does so at a price. The price is that this form of internalism, insofar as it is true, no longer seems to be telling us anything interesting about our moral judgements in general. After all, if this form of internalism does not apply even to exemplars such as Weidner then it does not seem to be telling us anything about moral judgements made by human agents. Rather, it is making a trivial claim about what we mean by 'fully virtuous agent'. For these reasons, this form of agent restriction will not be able to handle the case of the supererogation amoralist.

The next form of restriction I will look at is restricting the claim to normal agents. According to this version of internalism, moral judgements are necessarily connected to motivation for normal agents. If this form of restricted internalism is plausible then we will be able to explain why it is possible for supererogation

³⁷ Thanks to Mike Ridge for pressing me on this point.

amoralists to exist. The existence of such people is compatible with this form of internalism as they are not normal agents. On this account all normal agents will be necessarily motivated by their supererogation judgements.

This account is supported by focusing not on the individual making the judgement but on the community that she is a part of. According to this form of internalism, there must be some internal connection between moral judgements and motivation even if it is just one that is generally present in the community rather than in every individual member of that community.³⁸ Even if individual amoralists are conceivable, it is argued, a community of amoralists is not. Perhaps, then, internalists can make a similar response to the conceptual possibility of supererogation amoralists. In response to such a problem it might be argued that, while it may be possible for an individual to remain unmotivated by a judgement of supererogation it would not be possible for a community of such people to do so. The important issue to consider about such a world is whether or not we could say that the inhabitants are making genuine supererogation judgements given that people tend not to be motivated to act in line with them.

The first point to make against the ability of this form of Unrestricted Internalism to accommodate the existence of supererogation amoralists is that it doesn't seem at all clear that normal people are motivated by their supererogation judgements. Of course, it isn't clear how we should understand 'normal' in this context.³⁹ However, I cannot see any way of understanding

³⁸ See Dreier (1990 p.14), Blackburn (1998 pp.61-68), Lenman (1999) and Gibbard (2003).

³⁹ We might understand it in a purely statistical way or in a more normative way. Of course, if understood in a statistical way then this seems to provide no support for the internalist claim, as internalism is supposed to be an *a priori* thesis. This point is made by Miller (2008 p.245).

normal that is related to its ordinary use for which it would be plausible to think that normal people are generally motivated by supererogation judgements. A major part of what makes it so admirable to perform supererogatory acts is that most people would not have acted in the same way if they had been in that situation.⁴⁰ This by itself seems to be good reason to reject this approach to dealing with the problem of the supererogation amoralist. It is worth noting, though, that rejecting this claim will not be enough to save this response to the supererogation amoralist. The internalist must endorse the stronger claim that it is necessarily the case that normal people are motivated by their supererogation judgements. In other words, they must show that it is inconceivable that normal people would be unmotivated by supererogation judgements. This surely gives us good reason to reject this response.

Perhaps, though, the real challenge of Communal Internalism should not be put in terms of normality. After all we might be quite happy to describe judgements as moral judgements so long as some people are motivated by these judgements. A more worrying challenge is that moral judgements are necessarily connected to motivations at a community level. In order for a judgement to be considered a moral judgement perhaps it must exist against a background of moral judgements motivating people to act.⁴¹ To relate this to the supererogation amoralist, perhaps such a person is only possible against a background of at least some people being motivated by supererogation judgements.

However, unlike with moral judgements in general, there seems to be no reason to think that a judgement of supererogation can only

⁴⁰ To be clear, I do not mean that most people never perform supererogatory acts, only that they do not perform all of the supererogatory acts available.

⁴¹ This is the point that Lenman makes (1999 p.445).

count as genuine against a background of people being motivated to act in line with these judgements. We can see why if we imagine a world, Planet Duty, where no one is ever motivated by supererogation judgements. Now imagine that on Planet Duty there are differing conceptions of what morality requires. Some people, The Maximizers, have a very demanding conception, while for others, The Satisficers, morality is thought to demand very little. Imagine that those with the less demanding conception of duty are full of praise and admiration for the acts performed by those with the more demanding view. They give them medals and write poems celebrating their acts. They just are never motivated to perform acts that they judge to be supererogatory. I can see no reason to think that these people are not making genuine supererogation judgements.

When it comes to considering judgements of supererogation, there seems no reason to think that the challenge of Communal Internalism shows that the idea of someone making a genuine judgement of supererogation is dependent on people in their community being motivated by such judgements. Appealing to Communal Internalism will not save Unrestricted Internalism from the supererogation amoralist.

The second response that might be made in defence of Unrestricted Internalism is that the motivations that moral judgements are necessarily connected to, need only be *pro tanto*. This form of internalism is often called Weak Internalism and can be formulated as follows:

Weak Unrestricted Internalism: If an agent judges that it would be morally good for her to ϕ then she will necessarily have some

motivation to ϕ . If an agent judges that it would be morally bad for her to ϕ then she will necessarily have some motivation not to ϕ .⁴²

On this account moral judgements will always provide some motivation but this motivation may be overridden by contrary motivation.⁴³ This response allows the supporter of Unrestricted Internalism to claim that judging an act to be supererogatory is necessarily connected to motivation but that this motivation may be outweighed. It is possible on this view to judge an act to be supererogatory and not be sufficiently motivated to perform it. This response is attractive as it allows us to retain an internal connection between all moral judgements and motivation without making the implausible claim that those who judge an act to be supererogatory will necessarily be motivated all the way to action.

Furthermore, we might think that two commonly identified features of supererogation judgements might provide further support for this claim. First, it is often claimed that acts of supererogation necessarily involve some degree of self-sacrifice from the agent. Of course, as I argued in the previous chapter, there is good reason to reject the claim that all acts of supererogation involve sacrifice. Nevertheless, if we accept that acts of supererogation *typically* involve self-sacrifice then we might think that judging an act to be supererogatory will typically involve a judgement that there are self-interested reasons against performing supererogatory acts.⁴⁴ Similarly, supererogatory acts

⁴² That is, unrestricted in terms of what evaluations are covered. These forms of internalism may be restricted in terms of what agents they apply to.

⁴³ Mason describes a version of this form of internalism restricted to those that are practically rational as Weakest Internalism (2008 p.144).

⁴⁴ This, presumably, will not be the case if we think that self-regarding supererogatory acts are possible, as Kawall (2003) does. This possibility is ruled out if we accept Heyd's claim (1982 p.115) that altruistic intentions are necessary for acts of

are often thought to be ones for which formal sanctions such as punishment and informal sanctions such as blame are inappropriate responses to non-performance.⁴⁵ This means that typically there are strong self-interested reasons to perform moral requirements that do not support performing acts of supererogation. It seems reasonable to think that this judgement will often provide motivation against performing the supererogatory act either because judgements of self-interest are also necessarily motivating or because they typically are. Either way, recognizing that there are often strong self-interested reasons against performing acts of supererogation and against omitting obligatory acts allows supporters to explain why the connection to motivation seems weaker for supererogation judgements than for obligation judgements.

However, this response can be dismissed for two reasons. First, even this weaker claim is implausibly strong. It seems quite possible to judge that an act is supererogatory and be completely unmotivated to perform the act. This form of internalism would be committed to saying that in order for Judy from the previous example to be said to be making a genuine moral judgement she must be motivated to some extent to donate all of her money to charity. It seems perfectly possible to judge that an act is supererogatory and to be completely unmotivated to perform the act.

Of course the supporter of Weak Unrestricted Internalism might respond by insisting that this cannot be considered a genuine moral judgement. Note, though, that this point is not supported by the usual arguments from the practicality of moral language. As

supererogation. However, as I argued in Chapter Three, there is good reason to reject this requirement.

⁴⁵ Eg. Heyd (1982 p.115).

noted in §5.1 the arguments usually used to support internalism involve an appeal to the claim that there is something odd or incoherent about someone who utters a sentence like, ‘Eating meat is wrong’, and has no motivation to stop eating meat. This argument provides no support for the claim that supererogation judgements must be accompanied by *pro tanto* motivations. In order to support this form of internalism we would need to be given reason to think that uttering a sentence like, ‘Vegetarianism is beyond the call of duty’, necessarily indicates some *pro tanto* motivation the agent has to become vegetarian. Unfortunately for the supporter of Unrestricted Internalism there does not seem to be any reason to think that it does. In fact it seems plausible to think that making an utterance of this kind often suggests that the speaker lacks motivation to perform the action. It seems natural in this exchange to think that the speaker of this sentence will not be overall motivated to become vegetarian. Moreover there does not seem to be any reason to think she would be in some way incoherent if she does not have some *pro tanto* motivation to do so. Such a situation would not make us question the sincerity of the speaker’s judgement or her competence in her use of moral terms.

This first point connects to the second more general worry that others have voiced about *pro tanto* forms of internalism. As a number of authors have pointed out it seems implausible to think that people will necessarily be motivated to some extent by a judgement that there is some reason to do something when this reason is overridden by other concerns.⁴⁶ To see why consider the following case, adapted from an example given by Caj Strandberg:

⁴⁶ Those who make this point include Dancy (2004 p.17), Wallace (2006 pp.187-188), Strandberg (2013 p.32).

Life vs. Coffee: Clara has a serious illness and knows that she must undergo a medical procedure in order to save her life but that if she gets the treatment she will be unable to drink coffee for one minute.⁴⁷

Assuming that Clara likes drinking coffee, it seems that in this case Clara has some reason not to undergo the procedure albeit one that is vastly outweighed by considerations that count in favour of doing so. According to a *pro tanto* form of internalism about normative judgements, if Clara's judgement is a genuine one then she will necessarily be motivated to some extent to avoid the procedure. Now of course it might be the case that Clara does have this motivation but there does not seem to be any reason to think that she necessarily will. There certainly seems nothing odd or incoherent about someone who made this judgement of the balance of reasons and had no motivation whatsoever to avoid the procedure. Note that this point remains even if we combine Pro Tanto Unrestricted Internalism with some form of agent restriction. There is no reason to think that Clara could not be classed as a normal, rational or virtuous agent if she lacked this motivation.

Of course, the above case concerns normative reasons in general. Perhaps a supporter of Pro Tanto Unrestricted Internalism could claim that the theory holds only for moral reasons. Again, though, this theory looks implausible. To see why consider the following case:

Children vs. Donkeys: John has five pounds he has set aside to donate to a charity. He judges that he has some moral reason to donate it to a charity that cares for donkeys but much stronger reason to donate it to a charity that feeds starving children.

⁴⁷ (2013 p.32).

In this case John judges he has some moral reason to perform one act but much greater reason to perform another. Again there does not seem to be any reason to think that he will necessarily be motivated to some extent to give the money to the donkey charity if his judgement is sincere. Nor does it prevent him from being classed as a rational, virtuous or normal agent.

To sum up this section, I have argued that the possibility of the supererogation amoralist creates an important problem for the supporter of Unrestricted Internalism. I have considered two ways in which internalists typically respond to amoralist challenges and found that they do not succeed in providing an adequate response to this objection. In the next section I will investigate the plausibility of adopting a more restricted form of internalism.

5.4 Requirement Internalism and Moral Rationalism

If we accept the implausibility of Unrestricted Internalism then those wishing to defend internalism must pin their hopes on an evaluation-restricted version of the view. However, as we saw in §5.3, in order to be considered plausible, this form of internalism must be able to explain why it is that the internal connection to motivation holds only for moral requirements. In this section I will argue that there is a plausible explanation that internalists can give here.

The solution can be found by looking at one strategy that has been used to defend internalism. Both Michael Smith and Mark van Roojen argue that a form of internalism follows from the following view:

Moral Rationalism: If an act, ϕ , is morally wrong then, from an all things considered normative perspective, ϕ -ing is against the balance of reasons.⁴⁸

This view is one that many have found plausible, as it appears to follow from our practice of blaming those who fail to act in line with their moral requirements.⁴⁹

Both Smith and Van Roojen argue that moral rationalism provides an explanation for internalism. The essence of the argument is that if we necessarily have most reason to act in line with our moral requirements then when we judge an act to be morally required we will necessarily be motivated to perform the act in so far as we are rational.⁵⁰

As Strandberg notes, in order to move from moral rationalism to a form of internalism about moral judgements we need to accept the following form of internalism about normative judgements in general:

⁴⁸ See Smith (1994 pp.61-62) and van Roojen (2010).

⁴⁹ This is a simplified version of the argument given by Portmore (2011 pp.43-44) and Darwall (2006 pp.95-99).

⁵⁰ This is roughly the argument Smith gives (1994 pp.61-62). However, as van Roojen notes this is too simplistic, even if we take moral rationalism to be true, it is possible for rational agents to be unaware of its truth. Such an agent cannot be said to be irrational if she fails to be motivated by her obligation judgement. Nevertheless, van Roojen's argument argues that it is the normal cases, those where the judgement that an action is right motivates the agent, that fix the meaning of the term 'right' and that this allows us to conclude that rational agents acting normally will be motivated by a judgement that an act is morally required (2010 pp.518-521).

Normative Judgement Internalism: If a rational agent judges that ϕ -ing is what there is most reason to do then she will necessarily be motivated to ϕ .⁵¹

For my purposes, it is important to note two implications about the form of internalism that this argument vindicates. First, if we accept this argument then internalism is not a special feature of moral judgements. Rather it is a feature shared by all conclusive normative judgements. Second, and more importantly for our purposes, this argument can only vindicate a form of internalism that is restricted both to certain kinds of agent and to certain kinds of moral judgements. As moral rationalism is a view about moral requirements rather than moral judgements in general, this argument can only vindicate a form of internalism that is restricted to judgements about what is morally required. This provides the supporter of Restricted Internalism with an explanation for restricting the claim to certain kinds of moral evaluation. The reason is that it is only all things considered normative judgements about what there is most reason to do that are necessarily connected to motivation and only moral requirements that are necessarily in line with what there is most reason to do all things considered. It is worth noting that this form of internalism allows for rational agents to be motivated to perform a morally required act or a supererogatory alternative. All that it rules out is the possibility of a rational agent being motivated to perform a morally wrong act.

Conclusion

In this chapter I have looked at what a supporter of motivational judgement internalism can say about judgements of

⁵¹ Standberg (2013). A version of this form of internalism is endorsed by Wedgwood (2007 p.25).

supererogation. I have argued that the problem of the supererogation amoralist gives us good reason to think that internalism cannot plausibly be held to cover all moral judgements. As a result, we should restrict the internalist claim to requiring moral judgements. However, we need to give some explanation for this restriction in order for this form of internalism to be plausible. I have argued moral rationalism provides us with a perfect explanation for this restriction as it allows us to say that internalism about moral requiring judgements follows from internalism about all things considered normative judgements.

Chapter Six: Supererogation, Moral Rationalism and Overridingness

Introduction

Suppose that on my way to an important job interview I find a lost child. I realize that I could help the child find her parents but that this might make me late for my interview. In this case it seems reasonable to think that, morally, I ought to help the child, despite the fact that I have a strong self-interested reason not to do so. While not everyone has faced such a dilemma, we all may find ourselves in situations where what we morally ought to do clashes with acts favoured by other kinds of reasons. When faced with such a situation we must ask ourselves what we have most reason to do, all things considered. It is commonly claimed that in cases of conflict such as this, moral reasons override other types of reason, meaning that we always have most reason to perform the act favoured by morality. This view is called the overridingness thesis (hereafter, ‘overridingness’). A closely related view is that we always have most reason, from the all things considered normative perspective, to act in line with our moral requirements. This view is called moral rationalism (hereafter, ‘rationalism’). These views may seem to be almost identical but, as I will show, there are important differences between them.

These two views are important in their own right because accepting either view will have consequences for our everyday decision-making. If we accept overridingness then we should accept that when faced with a dilemma like the one above we ought to perform the act favoured by morality. If we accept

rationalism then we accept that if we are morally required to help the child in the above case then this is what we have most reason to do. These views also have important implications for other debates in Moral Philosophy. Accepting overridingness or rationalism could provide resources for showing why we ought to act morally. If either is true, then we can say that acting in line with our moral requirements is what we have most reason, all things considered, to do. These views also have implications for ‘The Demandingness Objection’ that is commonly raised against standard forms of consequentialism.¹ It is argued that consequentialist views generate moral requirements that make unreasonable demands of moral agents. However, as a number of authors have pointed out, this objection has little force unless we accept some form of overridingness or rationalism.² Finally, as we saw in the previous chapter, rationalism is also thought to have implications for the debate about whether there is an internal connection between moral judgements and motivation. It is claimed that if rationalism is true, rational agents will be motivated by their moral judgements.³ Moreover, we can justify restricting motivational judgement internalism to obligation judgements if we accept moral rationalism, the view that if an act is morally required it is what there is most reason all things considered to do.

Despite the importance of both views there has been surprisingly little discussion of the relationship between the two theses. In fact, confusion about the differences between the two is common. Sometimes a view is described as one and labeled as the other. Sarah Stroud, for example, defines overridingness in the following

¹ See, for example Sidgwick, (1981 pp.87, 434, 492).

² See Portmore (2011 pp.26-27), Dorsey (2012 pp.1-23), and Attila Tanyi (2012).

³ See Smith (1994 pp.61-62) and van Roojen (2010).

way: “If S is morally required to ϕ , then S has most reason to ϕ .”⁴ Similarly, Samuel Schleffer defines ‘the claim of overridingness’ in the following way: “It can never be rational knowingly to do what morality forbids.”⁵ These views are versions of rationalism not overridingness, as they say nothing about whether the moral reasons have overridden other reasons. My aim in this chapter will be to clarify the difference between the two views and then to show that a form of moral rationalism is compatible with the existence of acts of supererogation. Moreover, I will also show that this form of rationalism is the version that follows from the considerations that are typically appealed to by those seeking to motivate support for the view.

This is important for our purposes because, as I will show in §6.1, if we accept the possibility of acts of supererogation then overridingness appears to be an implausible view to hold. Having explained the problems that supererogation raises for overridingness I will then explain, in §6.2, how rationalism and overridingness can come apart in a way that allows those who accept that acts of supererogation exist to endorse overridingness. However, the fact that these two views can come apart does not give us any reason to think that it is plausible to accept rationalism without overridingness. After all, the considerations that count in favour of rationalism may also count in favour of overridingness. I will investigate whether or not this is the case, starting, in §6.3, by looking at the three important intuitions that are often appealed to by those seeking to defend rationalism. In §6.4 I will show that there are two possible readings of rationalism, a *de dicto* reading and a *de re* reading and that only the former is supported by the intuitions considered in §6.2. In §6.5 I will argue that the *de dicto*

⁴ (1998 pp.170-189).

⁵ (1992 p.53).

version does not provide any support for accepting overridingness over the alternative way of understanding rationalism.

Before I begin, it is worth making clear four assumptions that I will be making in this chapter. First, in order for either rationalism or overridingness to be an interesting claim it needs to be assumed that there is a genuine distinction to be made between moral and non-moral reasons. In other words, I will be assuming that it is possible for some consideration to count in favour of performing some action without its counting in favour from the point of view of morality. Without this assumption, the claims are trivial.⁶ This rules out two kinds of justification for rationalism or overridingness. The first justification holds that all normative reasons are properly understood as moral reasons. The second holds that what we ought morally to do is fully determined by what we ought to do from a self-interested point of view. Of course, this does not mean that these views are not interesting or defensible.

In addition, I will be assuming that moral and non-moral reasons can conflict with one another; that the balance of moral reasons can support performing an act while the balance of non-moral reasons can oppose the act's performance. This assumption needs to be made in order for overridingness to be an open possibility.⁷ If moral reasons never conflict with non-moral reasons then neither will ever override the other. We do not, though, need to make this assumption in order to accept rationalism. If these two kinds of reasons never conflict then rationalism's truth is guaranteed, as what there is most moral reason to do will not

⁶ This point is made by Schleffer (1992 p.54).

⁷ Thanks to Mike Ridge for useful discussion here.

conflict with any opposing reasons.⁸ However, for the purposes of this chapter I will assume that these two kinds of reason can conflict in order to show that accepting rationalism does not give us reason to accept overridingness even if we make this assumption.

The third assumption that needs to be made is that moral and non-moral reasons are commensurable. In order to ask whether moral requirements always determine what there is most reason to do we must assume that there is some way of comparing moral requirements with other kinds of normative reasons or requirements. Otherwise it will not make sense to say that one always overrides the other. This view is not universally accepted but any version of either view that assumes that the two kinds of reason can conflict will have to make this assumption, so this is an acceptable assumption to make here.⁹

Finally, I will be assuming that rationalism is a necessity claim. In other words, the rationalist is not claiming that it is a contingent feature of moral requirements that they are always in line with what there is most reason to do. That is not to say that there are any problems with such a view. Rather, I take it that this is not what those who subscribe to rationalism for the reasons I will examine in §6.3 have in mind.

6.1 Overridingness and Supererogation

⁸ This point is made by Portmore (2011 p.39 Fn.32).

⁹ For arguments against the possibility of an all things considered point of view see Sidgwick (1981 p.508) and Copp (1997 pp.86-106). Parfit, responds to Sidgwick's argument in his (2011 Ch. 6). For a reply to Copp see McLeod (2001 pp.269-291).

In this section I will argue that if we accept the possibility of supererogation we have good reason to reject overridingness. First we must explain what is meant by ‘overridingness’. To say that one kind of reason always overrides another is to say that when the two conflict the first kind of reason will always defeat the other relative to some normative standpoint. As the standpoint we are interested in is the all things considered standpoint, we can define the relevant kind of overridingness as follows:

Rational Overridingness: One kind of reason, m, rationally overrides another, n, if and only if from an all things considered normative perspective, m reasons always defeat all n reasons in terms of importance or normative strength.¹⁰

We are interested in whether moral reasons override other normative reasons. We can define one version of this view as follows:

Strong Overridingness: From an all things considered normative perspective, moral reasons always override all non-moral reasons.¹¹

I will start by arguing that if we accept the possibility of supererogation then we have good reason to reject Strong Overridingness. This view has the implication that we always have decisive reason all things considered to do what we have most moral reason to do.¹² Accepting this view is problematic if

¹⁰ Similar definitions of ‘overridingness’ are given by Copp (1997 p.90), Portmore (2011 p.39), and Shiffrin (1999 p.773).

¹¹ Baier defends this view (using the term ‘superior’ rather than ‘overriding’), (1958 p.99). Others who define overridingness in terms of reasons include, Portmore (2011 p.40) and Tanyi (2012 p.167).

¹² As pointed out by Portmore (2011 p.40).

we also accept the existence of supererogation. As we have seen a supererogatory act is one that is morally better than the merely permissible alternatives. Presumably, then, these acts are better supported by moral reasons than the alternatives. This means that we always have decisive reason all things considered to perform the best available supererogatory act.

The problem with this is that it conflicts with the thought that supererogation is not only morally optional but also rationally optional. There are two ways of making this point. The first is by appealing to specific cases of supererogation. Suppose, for example, that Anna passes by a burning building and hears a child's cry for help coming from inside. At great personal risk she runs into the building to save the person trapped inside. This act seems like a clear case of supererogation. After all, it would seem inappropriate to blame Anna for not running into the burning building. Note though, that if we accept Strong Overridingness, then, if Anna fails to perform this act then she will fail to perform the act that she had decisive reason to perform. It appears, then, that some form of criticism is available. While we may not be able to morally criticize the agent for failing to save the child we can rationally criticize her. This is a strange result. It seems no more appropriate to criticize this agent for acting irrationally than to criticize her for acting immorally. In fact, if anything we might think that it would be more appropriate to criticize this agent for being immoral rather than criticizing her for being irrational.

We might not wish to put too much weight on our intuitive reaction to this case. However, the problem that supererogation raises for Strong Overridingness goes beyond bringing about counter-intuitive results in particular cases. Strong Overridingness also commits us to a very strange theoretical position. If we accept

that, in the previous example, Anna has most moral reason to save the child but that it is morally permissible not to then we are committed to accepting that what we have most moral reason to do does not fully determine what we are morally required to do. It would be strange to accept this view together with the view that moral reasons are rationally overriding. However, if we accept Strong Overridingness then from the all things considered normative point of view moral reasons are decisive. This is a very odd position to hold, if moral reasons fail to be decisive with regards to the moral point of view how could they be decisive from the all things considered point of view? Surely, if moral reasons are decisive in any domain then it would be in the moral domain not in the domain of all normative reasons.

A supporter of overridingness might respond by claiming that not all moral reasons are rationally overriding. Perhaps when we have a choice between a supererogatory act and a morally permissible, non-obligatory alternative then the moral reasons that support performing the supererogatory act are not rationally overriding.¹³ We can, though, accept this but still hold the following weaker version of overridingness:

Weak Overridingness: From an all things considered normative perspective, the reasons that support or are provided by moral requirements always override all non-moral reasons.

Weak Overridingness covers two possible views. We might think that it is the reasons that support moral obligations that override all non-moral reasons.¹⁴ Alternatively, we might think that moral

¹³ This point is made by Terrell (1969 pp.52-53) and also by Darwall (2006 p.286).

¹⁴ This possibility is considered by Darwall (2006b p.286) and also by Schlothfeldt and Schweitzer (2012 pp. 65-88).

requirements provide conclusive reasons to act while other moral reasons do not.¹⁵

This view, though, is at best a minimal improvement on the previous one. Accepting this view fails to provide a complete solution to the problem facing Strong Overridingness. This view was unable to plausibly accommodate cases of supererogation where the cost to the agent's self-interest prevented an act from being obligatory that would otherwise have been so. It looks initially as if Weak Overridingness is well placed to avoid this problem as, unlike Strong Overridingness, Weak Overridingness is not committed to saying that from the all things considered normative point of view there is always decisive reason to act in line with what there is most moral reason to do.

However, on closer inspection accepting Weak Overridingness fails to solve the problem. To see why, imagine that Anna could save the child at no risk to herself. Presumably, in this case it would be morally obligatory for her to do so. However, if Weak Rationalism were true then the reasons that would make saving the child morally obligatory in the case where there is no risk involved should also do so in the case where there is a risk to the agent. After all, the reasons that support the moral requirement in the first case presumably continue to do so in the second. If these reasons override all non-moral reasons then so long as the it remains morally better to attempt to save the child than not to do so this will be what there is decisive reason to do all things considered. At least, this is the case if we accept the first form of Weak Overridingness. What about the view that moral

¹⁵ This version of Weak Overridingness is endorsed by Brink who claims that moral requirements *provide* overriding reasons for action (1997 p.256) and Darwall (2006a p.26).

requirements provide reasons that always override non-moral reasons? This view is also problematic. To see why imagine that when the agent from the previous case first hears the cries for help the fire is not particularly serious. She could then save the child without much risk to herself. At this stage many would think that this act is morally required. If we accept this version of Weak Rationalism then this would provide reason to save the child that override all non-moral reasons. If this were the case then this act would remain what there is most reason to do even if the risks were to increase, at least for as long as saving the child remains the act that is best supported by moral reasons. Presumably, though, once the risks are raised sufficiently the act ceases to be morally required. Again, then, we are left with a view that holds that these moral reasons are decisive from the all things considered normative point of view but not from the moral point of view.

To sum up, if we accept the existence of supererogatory acts then overridingness appears to be implausible. First because it does not appear to be the case that we always have decisive reason from the all things considered point of view to perform the best supererogatory act available. Second because it seems theoretically odd to think that we always have decisive reason from the all things considered point of view to perform the act that is best supported by moral reasons but we do not always have decisive reason from the moral point of view to do so. Retreating to Weak Overridingness does not help avoid the problem as it cannot handle cases where an act would have been required if it were not for the presence of costs to the agent's self-interest.

This puts pressure on the claim that motivational judgement internalists can avoid the problem of supererogation by appealing

to rationalism to justify restricting their claim to obligation judgements. As we have seen many who accept rationalism do so because they accept that moral reasons override other normative reasons. Indeed, the views are often assumed to be indistinguishable. If this is the case then the internalist solution will involve endorsing a view that seems hard to reconcile with the existence of acts of supererogation. In the next section I will show that we can in fact accept that moral requirements are always in line with what there is most reason to do without accepting that moral reasons always override other forms of normative reason.

6.2 Moral Rationalism Without Overridingness

In this section I will explain the difference between rationalism and overridingness and show that it is possible to accept rationalism without accepting overridingness.

Weak Overridingness seems superficially similar to the following view:

Moral Rationalism: If an act, ϕ , is morally wrong then, from an all things considered normative perspective, ϕ -ing is against the balance of reasons.¹⁶

¹⁶ In an earlier version of this discussion (2014 p.106) I formulated moral rationalism in the following, slightly different way: “If an act, ϕ , is morally required then, from an all things considered normative perspective, ϕ -ing is what there is most reason to do.” However, this earlier formulation of moral rationalism is incompatible with there ever being most reason to perform an act that is supererogatory. By formulating moral rationalism in the way I do here I avoid this problem while retaining the basic idea that there is always most reason to act in line with our moral requirements.

However, there is an important difference between the two. Weak Overridingness says that the reasons that support or are provided by moral requirements always override all non-moral reasons. Rationalism, on the other hand, makes no mention of what explains why we have decisive reason to act in line with our moral requirements nor does it state that morality is of greater normative importance than other normative considerations.

In addition to these views being similar, the acceptance of overridingness provides an explanation for rationalism. If we accept that either version of overridingness is a necessary truth and we think that moral requirements are fully determined by what we have most moral reason to do then we should also accept rationalism. According to such a view, moral requirements will always be what agents have most reason to do.

The question I will be considering is whether the reverse applies. Does accepting rationalism give us any reason to accept some version of overridingness? One way in which it would is if overridingness provided the only possible explanation for rationalism. Stephen Darwall seems to suggest that this is the case in the following:

Reasons provided by moral oughts might be invariably supreme because they are guaranteed to *override* other normative reasons, because they invariably *defeat* (that is reduce or undermine the force of) other reasons, or through some combination of the two.¹⁷

It is not important, for my purposes, whether moral reasons defeat other normative reasons by simply exceeding them in normative

¹⁷ (2006b pp.286-287).

force or by the more complicated method of ‘defeating’ that Darwall mentions here, as both count as versions of overridingness. The important point is that Darwall seems to claim that accepting what he calls ‘supremacy’, the view that moral obligations provide conclusive reasons for action, pushes us towards thinking that some form of Weak Overridingness is true. Darwall makes a similar point elsewhere when he moves from the claim that, ‘it can never be rational to do what morality forbids,’ to the claim that, ‘moral obligations always *give* agents conclusive reasons for acting that outweigh or take priority over any potentially competing considerations.’¹⁸ This thought, that the explanation for rationalism will come from some version of overridingness, may also explain why, as mentioned previously, some philosophers have applied the label The Overridingness Thesis to views that are really forms of rationalism.

However, overridingness does not provide the only explanation for rationalism. It is possible to accept rationalism without accepting either version of overridingness. If the following constraint on what can be counted as a moral requirement is taken as a necessity claim then we have an explanation that does not appeal to either form of overridingness:

The Constraint Thesis (Constraint): If an act, ϕ , is not against the balance of reasons then ϕ -ing is not morally wrong.

This is not a version of overridingness. According to Constraint, rationalism is explained by the fact that unless an act is against what the agent has most reason to do then it cannot be morally wrong. It is important to be clear on how this view differs from

¹⁸ (2006a p.26). Emphasis added.

Weak Overridingness. It would be easy to think that Constraint tells us that in order to be counted as a reason that makes an act morally forbidden, that reason must always override all non-moral reasons. This, though, would be a mistake. Constraint does not say that any moral reasons will always take priority over non-moral reasons. It merely states that it is a necessary condition for moral prohibitions that they are against what we have most reason to do, all things considered.

Unlike overridingness, then, Constraint does not commit those who hold that acts of supererogation exist to the view that moral reasons, or a subset of moral reasons, are decisive from the all things considered normative point of view but not from the moral point of view. As we saw in §6.1, if we accept the possibility of acts of supererogation then we seem forced to accept that moral reasons are not always morally overriding. It would be odd, though, to accept that moral reasons are not morally overriding but maintain that they are rationally overriding. Constraint avoids this problem, as it does not state that any subset of moral reasons will always be decisive from the all things considered point of view. Those who accept Constraint are committed only to the view that an act cannot be morally wrong unless it is against the balance of reasons to perform. This view can allow that moral reasons are not always rationally decisive. This gives us reason to think that those who accept both moral rationalism and hold that supererogatory acts exist should accept Constraint.

Overridingness and Constraint provide two different explanations for rationalism. The view that rationalism should be viewed as a constraint is not unique to me, it is also held by Portmore. Interestingly, despite having a different view of rationalism to Darwall, Portmore's defence of the view rests on an argument

given by Darwall.¹⁹ While Portmore uses this argument to support the truth of rationalism he does not show that this argument provides equal support for both his and Darwall's versions of the view. This is important, as we might think that the reason that Darwall, and others, accept Weak Overridingness is that the arguments given to support rationalism push us to understand it in this way. If this is the case then rationalism will remain an unappealing view to hold if we accept the existence of acts of supererogation. It is this thought that I will investigate in the next two sections of this chapter.

Portmore does not address this thought. His argument for accepting Constraint is that this is the form of rationalism that we are pushed towards if we reject both Strong Overridingness and the claim that moral reasons are morally overriding.²⁰ My defence of Constraint will take a more direct approach. I will argue that it is Constraint and not overridingness that is supported by the intuitions commonly appealed to by those seeking to defend rationalism, and that both Darwall's and Portmore's arguments depend on. I will start by looking at the intuitions that are commonly appealed to in support of rationalism. I will then argue that these intuitions support Constraint and not overridingness.

6.3 Why accept Moral Rationalism?

There are three intuitions that are commonly appealed to in support of rationalism. In this section I will motivate these intuitions and show why they provide support for rationalism. I will not attempt to provide a conclusive defence of rationalism.

¹⁹ Portmore (2011 pp.42-51), Darwall (2006 pp.96-99). See also Skorupski (1999 pp.170-171).

²⁰ (2011 pp.41-42).

My aim is the more modest one of showing that those who accept it for the reasons commonly given in support of the view are not thereby given any reason to accept some form of overridingness.

The first intuition that is commonly appealed to in support of rationalism is that moral requirements place rational constraints on our actions. If someone has a moral obligation to act in a particular way then we do not think that she is free to choose how to act. Rather, we think that to violate this moral requirement would be to act against the balance of reasons. Those who think that I have a moral obligation to help the lost child rather than attend the job interview would likely also think that this settles the question of what I have most reason to do. As many have pointed out, this point seems to be presupposed by our practice of blaming those who freely and knowingly act wrongly.²¹ We think it appropriate to blame those who freely and knowingly perform a wrong act. To be blameworthy, though, it seems reasonable to think that we must judge that the agent did not have sufficient reason to act as she did. It would, after all, be odd to blame me for failing to help the child while acknowledging that this is what I had most reason to do. In order for blame to be appropriate, then, we must think that by violating a moral requirement the agent acted against the balance of reasons. If we accept these two claims then we can conclude that if an act is morally required it must be an act that the agent had most reason to perform.²²

²¹ This point is made by Darwall (2006a p.97), Gibbard (1990 p.299), Portmore (2011 p.43), Shafer-Landau (2003 p.192), Stroud (1998 p.176) and Williams (1995 pp.40-44).

²² This is a simplified version of the argument given by Portmore (2011 p.43-44) and Darwall (2006a pp.95-99).

The second intuition is that moral requirements provide a rational justification for action.²³ If there are two acts available one of which is morally required and the other morally forbidden then there does not seem to be any need to give a further rational justification for performing the required act rather than the wrong act. Imagine that Alice tells Jack that he has a moral obligation to stop eating meat. The next day Jack tells Alice that he has become vegetarian. Now imagine that in response Alice accuses Jack of acting against the balance of reasons. I think most people would agree that this would be a strange thing for Alice to say. Assuming that Alice's views have not changed, we would think that Alice is guilty of some kind of confusion, insincerity or irrationality. The reason why we would think this is that if an act is morally required then this seems to be all that is needed in order to rationally justify performing it. Rationalism provides the perfect explanation for this intuition. The reason we are justified in acting in line with our moral requirements is that it is always against the balance of reasons not to do so.

We might worry about this intuition. While it seems plausible to think that we always have most reason to perform a morally required act when we have a choice between a required act and a forbidden act, we might worry that things are less straightforward when more options are available. As we saw in Chapter 4, there are some acts of supererogation that promote the agent's self-interest to a greater degree than the minimally permissible alternatives. Surely in cases such as these there is more reason to perform the supererogatory act than the required act. After all, these acts are better supported by both moral reasons and by the

²³ This point is made by Shafer-Landau (2003 p.192) albeit for the weaker claim that moral requirements always provide some reason for action.

agent's self-interested reasons than the minimally permissible alternative. This, though, need not overly concern us. The thought behind this intuition is simply that we are rationally justified in acting in line with our moral requirements rather than not doing so. If violating a moral requirement is always against the balance of reasons then we will always have most reason to fulfil a moral requirement in some way, whether that is by performing a minimally permissible act or by performing a supererogatory alternative.

The third intuition is that demonstrating that an act was in line with the balance of reasons serves as a moral justification for action.²⁴ Showing that an act was in line with what an agent had most reason to do seems sufficient to show that the act was not morally wrong. It would be odd for someone to claim that an act was in line with what she had most reason to do but also morally impermissible. Again, rationalism is able to provide the perfect explanation for this thought; the reason that this is sufficient moral justification is that moral requirements are always in line with what there is most reason to do. As a result, showing that an act is not what an agent has most reason to do is sufficient to show that it is not morally required.

6.4 Moral Rationalism *De Dicto* and *De Re*

Before we are in a position to assess whether the intuitions used to defend rationalism support overridingness or Constraint, we must first distinguish between two different ways of understanding rationalism. This will allow us to see exactly what form of rationalism these intuitions support.

²⁴ Darwall (2006 p.98).

Rationalism says that it is a necessary truth that morally required acts are what there is most reason to do. This claim can be read in two ways, *de dicto* and *de re*. The best way to understand the difference between *de re* and *de dicto* necessity is through an example. Consider the following claims:

Claim 1: Necessarily, husbands are married.

Claim 2: Husbands are necessarily married.

These seem similar but have different truth conditions. Claim 1 is a claim about *de dicto* necessity while Claim 2 is a claim about *de re* necessity. The truth or falsity of Claim 1 depends on the truth or falsity of the following conditional proposition: if x is a husband then x is married. This claim is true; it is impossible to be an unmarried husband. Claim 2, on the other hand, is a claim about the individuals in the actual world that are husbands. This claim tells us that these individuals are necessarily married. This claim is false, for any man in this world who is married there are many possible worlds where he is unmarried.

Contrast this with the following example: the current Prime Minister of the UK is necessarily the offspring of Mary Fleur Cameron and Ian Donald Cameron. In this example the *de re* necessity seems much more plausible. When we ask whether it is possible that the actual person, David Cameron, who is Prime Minister could have been the offspring of other people it seems plausible to think that he could not. We can imagine possibilities about David Cameron, he could have died young, never had children or entered politics. Plausibly, however, he could not have had different parents. Note that the equivalent *de dicto* necessity claim is much less plausible here. It would have been possible for

someone who was the offspring of different parents to have been Prime Minister instead.

The difference, then, between *de dicto* necessity and *de re* necessity is that with the former the necessary connection is one that exists between the application of the term while with the latter the application determines what it is that the necessary connection applies to. In the husband example, it is the application of the term 'husband' that is necessarily connected to being married rather than the people who are picked out by this term. On the other hand, in the second example, it is the person picked out by the term 'Prime Minister of The UK' who is necessarily connected to being the offspring of Mary Fleur Cameron and Ian Donald Cameron.

Let's now apply these two kinds of necessity claims to rationalism:

De Dicto Moral Rationalism: Necessarily, if an act, ϕ , is morally wrong then ϕ -ing is against the balance of reasons.

De Re Rationalism: If an act, ϕ , is morally wrong then necessarily ϕ -ing is against the balance of reasons.

De Dicto Rationalism says that in every possible world, if an act is morally wrong then it is against the balance of reasons to perform it. De Re Rationalism, on the other hand, says that all of the acts that are morally wrong in this world have the essential property of being against the balance of reasons. This is a property that will be present in every possible worlds. If De Dicto Rationalism is true then it will be the application of the term 'moral wrong' that is necessarily connected to being against the balance of reasons. On

the other hand, if De Re Rationalism is true then it is the acts picked out by this term that have the essential property of being against the balance of reasons.

To see which form of rationalism these intuitions support consider how we ought, in general, to test claims of *de dicto* and *de re* necessity. To test Claim 2 above we must ask whether there is a possible world where someone is a husband without a spouse. To test Claim 3 we must consider someone who is a husband in this world and ask whether there is a possible world where he does not have a spouse. We can use the same tests to see whether rationalism is best understood as a claim of *de dicto* or *de re* necessity. To test for *de dicto* necessity we must ask whether there is a possible world where an act is both morally wrong and not against the balance of reasons for the agent to perform. To test for *de re* necessity we must ask whether any act that is morally required could fail to be what there is most reason to do in some other possible world. In the former we are testing the application of ‘morally required’ across possible worlds and in the latter we are using the term ‘morally required’ to pick out the acts that we will test across possible worlds.

Clearly, the intuitions considered in §6.3 support only a *de dicto* necessity claim. When we say that moral requirements provide rational justifications and constraints, we do not mean that the acts that are morally required in this world are ones that will provide these constraints and justifications in all possible worlds. Rather, what we mean is that in order for it to be appropriate to apply the term ‘morally required’ to an act in any given world, the act must be one that provides rational justifications and constraints in that world. Similarly, when we say that a rational justification serves as a moral justification we do not mean that there is no possible

world in which there could be most reason to perform an alternative act to an act that is morally required in this world. Rather, we mean that for any possible world, if such an alternative exists then the act is not morally required in that world.

6.5 De Dicto Moral Rationalism, Weak Overridingness and The Constraint Thesis

In §6.4 I argued that the intuitions that support rationalism support only the *de dicto* reading not the *de re* reading. I will now argue that accepting this provides greater support for Constraint than for Weak Overridingness.

Constraint says that an act cannot be morally wrong if it is not against the balance of reasons for the agent to perform. If we understand this as a necessity claim then this claim is logically equivalent to De Dicto Rationalism. This can be clearly seen when we place the two claims side by side:

De Dicto Moral Rationalism: Necessarily, if an act, ϕ , is morally wrong then ϕ -ing is against the balance of reasons.

Constraint Thesis: Necessarily, if an act, ϕ , is not against the balance of reasons then ϕ -ing is not morally wrong.

These claims are logically equivalent, accepting De Dicto Rationalism commits us to accepting Constraint and vice versa. As a result, Constraint provides the best fit for De Dicto Rationalism. It explains why, in any possible world, if an act is morally wrong then it must be the case that it is against the balance of reasons to perform in that world.

Overridingness, on the other hand, says that it is a necessary truth that moral reasons, or a subset of moral reasons, override all non-moral reasons. It is worth noting at this stage that if it were De Re Rationalism that the considerations appealed to in §6.2 supported then overridingness could explain rationalism while Constraint could not. We could not explain the claim that the moral obligations of this world will be what there is most reason to do across all possible worlds by appealing to Constraint, as this only constrains what can be classed as a moral obligation in one world to the balance of reasons in that world. A *de re* form of overridingness would be able to explain De Re Rationalism, as this form of overridingness holds that the acts that are required in this world are supported by reasons that are overriding across all possible worlds.

However, as we have seen, it is the *de dicto* reading not the *de re* reading that is supported by the intuitions examined in §6.2.

Overridingness is able to provide an explanation for De Dicto Rationalism but only if it is also understood as a *de dicto* necessity claim. We can rewrite the two forms of overridingness as *de dicto* claims in the following way:

Strong De Dicto Overridingness: Necessarily, moral reasons always override all non-moral reasons from an all things considered normative perspective.

Weak De Dicto Overridingness: Necessarily, moral requirements always provide or are supported by reasons that override all non-moral reasons from an all things considered normative perspective.

Like Constraint, these views place constraints on the correct application of moral terms. However, they are different constraints. Strong De Dicto Overridingness places a restriction on what can be counted as a moral reason; they must be reasons that override all non-moral reasons. Weak De Dicto Overridingness places a constraint on what can be counted as a moral requirement; these must either be supported by or provide reasons that override all non-moral reasons. Combining either with the claim that moral reasons fully determine an act's moral status gives us an explanation for De Dicto Rationalism.

While both Constraint and overridingness are consistent with De Dicto Rationalism it should now be clear that this form of rationalism provides more support for Constraint than for overridingness. As we have already seen Constraint and De Dicto Rationalism are logically equivalent. This means that if we accept De Dicto Rationalism then we are committed to accepting Constraint as well. In contrast, neither version of overridingness follows from De Dicto Rationalism, as they both make additional claims about the nature of moral reasons or moral requirements that De Dicto Rationalism on its own provides no support for. Strong Overridingness makes the additional claim that in order to class any reason as a moral reason it must be one that is capable of overriding all non-moral reasons. Similarly, Weak Overridingness makes the additional claim that moral requirements must provide or be supported by reasons that override all non-moral reasons. Neither follows from accepting De Dicto Rationalism, as it is silent on the reason why violating moral requirements is to act against the balance of reasons. Nor does it follow from the intuitions used to support rationalism, as these are also silent on whether moral reasons or some subset of moral reasons override other reasons. We cannot move from De Dicto Rationalism to

overridingness without giving an additional argument to convince us that this is the reason that De Dicto Rationalism is true. In the absence of such an argument, we have no reason to think that accepting rationalism for the reasons given in §6.3 gives us reason to accept some form of overridingness.

To sum up, the form of rationalism that is supported by the intuitions commonly given in favour of the view is logically equivalent to Constraint. Although both forms of De Dicto Overridingness are consistent with this view, they do not follow from it. In order to move from De Dicto Rationalism to some version of overridingness we need an additional argument that explains why the constraint should be understood in this way. We have no reason to think, then, that accepting rationalism gives us reason to think that some form of overridingness is true.

Conclusion

In this chapter I have investigated the implications of accepting the existence of supererogatory acts for both overridingness and moral rationalism. I argued first, in §6.1, that accepting that supererogatory acts exist gives us good reason to reject overridingness. The reason for this is that if we accept that supererogatory acts are better supported by moral reasons than the minimum that morality demands then existence of such acts shows that moral reasons are not morally decisive. If we accept this then it seems like we have good reason to think that moral reasons cannot be rationally decisive either. However, those who accept that supererogatory acts exist need not reject moral rationalism as a result of this. As I showed in §6.2, The Constraint Thesis offers an alternative way of understanding moral rationalism. Unlike overridingness, the combination of Constraint

and the existence of supererogatory acts does not have problematic implications. Moreover, as I showed in §6.3 to §6.5, the considerations that count in favour of accepting moral rationalism provide stronger support for Constraint than overridingness.

Chapter Seven: Do We Need Room For More Deontic Categories?

In the introduction we saw that dividing acts into the three categories of the obligatory, the indifferent and the forbidden leaves no room for acts that go beyond what is required. We need to add the category of the supererogatory to this classificatory scheme in order to make room for these acts. Some have argued that this four-fold classification is not sufficient either, as there are acts that it is unable to accommodate. Before finishing, it is worth considering whether a moral scheme that makes room for the supererogatory goes far enough. I will argue that none of the arguments used to support these additional categories succeeds in showing the need to make room for additional deontic categories in order to accommodate acts that cannot be classed as obligatory, indifferent, forbidden or supererogatory. I will start, in §7.1, by looking at the category of suberogation, before turning my attention first, in §7.2, to quasi-supererogation and then, in §7.3, to forced supererogation. I will argue that there is no need to make room for any of these deontic categories to accommodate acts that cannot be accommodated by the existing deontic categories. There may, however, be other reasons to allow these categories. It may be that although these categories are not needed to accommodate acts that lie outwith the existing scheme, they provide useful ways of labelling certain actions. I will argue that while this may be true for the category of suberogation it is not the case for quasi-supererogation or forced supererogation.

7.1 Suberogation

After supererogation, the category of suberogation or offence is by far the most popular proposed addition to the traditional tripartite

deontic scheme.¹ This category is supposed to be the mirror image of supererogation. While supererogatory acts are (roughly) optional and good, suberogatory acts are optional but bad. It seems reasonable to think that if such acts exist then it will not be enough for a moral theory to make room for the supererogatory, it must also make room for the suberogatory.

There are two ways of defending the need for this category. The first is by claiming that a moral scheme that fails to make room for a new category of the suberogatory will be incapable of accommodating all of the possible acts that might be performed. The second is by claiming that this classification is useful in helping us understand difficult moral issues. I will start this section by investigating the arguments of the first kind. I will argue that given the definition of supererogation in defended Chapter One, there is no need to make room for a new category in order to accommodate acts that could not otherwise be accommodated by a four-fold deontic scheme. However, this does not mean that giving this label to certain actions will not be useful. There might, then, be reason to use this term to pick out a class of acts that are already accommodated in the deontic scheme. This, though, is a claim that I will remain neutral on.

7.1.1 Claims From Moral Experience

Before we look at the arguments in depth, though, it is worth explaining the common starting point of all these arguments. Chisholm, Driver and Mellema have all argued that it is a familiar part of moral experience that some acts are both morally bad and

¹ Those who support making room for suberogation include Chisholm (1963), Chisholm and Sosa (1966b), Driver (1992), Ladd (1957 pp.125-130) and Mellema (1987;1991; 2005).

morally permissible. In this section I will investigate how this claim is motivated.

Chisholm supports the claim that acts can be morally bad and morally permissible by appealing to our intuitions about the following example:²

Suppose A knows concerning B, whom A dislikes, that the loss of B's employment would result in great tragedy for B and his family; that there is another man, C, who could do B's work but no more satisfactorily than B does it; and that B's employer, even if he knew the foregoing, would replace B by C if he thought that C were available. One might plausibly argue that, if A were deliberately to bring the availability of C to the attention of B's employer, his act would be permissible but at the same time heinous and inhuman.³

According to Chisholm, A's act in this example is both morally bad and morally permissible. Of course, we might disagree with Chisholm's diagnosis of the case but let's grant him that for now.⁴

Driver and Mellema's defence of this claim are a little more complex. They first defend the existence of a particular kind of moral situation and argue from this to the conclusion that some acts are both bad and permissible. Driver begins her argument by pointing to the existence of 'morally charged' situations.⁵ These are situations where we are faced with the choice of two permissible acts. Either we perform a morally good act or a morally bad act. No neutral act is available. Driver gives several examples of such situations.

The first is that of a passenger who boards a train and has a choice of seats. If he chooses his preferred seat he will prevent a couple

² Chisholm uses the term 'offence' rather than suberogation. For the sake of clarity I will stick to using the term 'suberogation'.

³ Chisholm (1963 p.5).

⁴ Heyd claims that this act is morally forbidden (1982 p.128).

⁵ Driver (1992 p.286).

behind him from sitting together. If he chooses a less preferred seat then the couple will be able to sit together. In this case there is no morally neutral option. If the passenger chooses to take the seat and prevent the couple behind him from sitting together then, according to Driver, he does something bad. On the other hand, if he decides to give up the seat to let the couple sit together then he has performed a morally good act.

The second example Driver gives is of two brothers Roger and Bob.⁶ Bob is suffering from kidney failure and will die unless he can get a transplant. Roger is the only compatible donor. Like the previous example, Driver claims that in this case there is no morally neutral option. If Roger donates the kidney then he performs a praiseworthy act. If he chooses not to then he will be blameworthy.

In both of these examples, Driver claims, the good act is supererogatory. The alternative act, on the other hand, is permissible, as follows from the fact that the alternative is supererogatory. If it were not then the alternative would be obligatory, as it would be the only permissible act. Nevertheless, Driver claims that in these examples the non-supererogatory act is “worse than the situation calls for”.⁷ These are acts that, “we ought not to do, but which would not be forbidden.”⁸ They are bad but permissible.

Mellema’s argument to support the need to make space for suberogation begins by appealing to the existence of a special kind of moral dilemma, which he calls a ‘blame dilemma’.⁹ Moral

⁶ Driver (1992 p.287).

⁷ (1992 p.290).

⁸ (1992 p.291)

⁹ Like Chisholm, Mellema names this category offence rather than suberogatory. For the sake of clarity I will continue to call this category suberogation.

dilemmas are traditionally defined as situations in which all of the acts available to an agent involve violating a moral obligation.¹⁰ However, Mellema argues that there is another form of moral dilemma, one in which the agent must choose between different blameworthy alternatives. These dilemmas are far less problematic than the previous kind, as they will not face problems of inconsistency.¹¹ Mellema gives an example of such a case in the following:

Suppose that Jones owns an autographed copy of Carnap's *Meaning and Necessity* and promises it to two different people on the occasion of his imminent retirement. To avoid the objection that one of these promises is deceitful, suppose that the two persons are identical twins whom he has taken to be one and the same person.¹²

Mellema claims that in this case Jones will be blameworthy whatever he does. Mellema then argues that, if we accept that ought implies can, then we must accept that at least one of these acts is permissible. After all, Jones cannot give the book to both people and so if we accept this principle then it cannot be the case that he ought to do so. If we accept this point then whichever act is permissible will be both permissible and blameworthy.

So far we have looked at examples given by Chisholm, Driver and Mellema that they claim give us good reason to accept the possibility of acts that are bad but not forbidden. Chisholm and Driver provide an additional reason to accept the existence of such acts. They claim that accepting the existence of the supererogatory puts pressure on us to accept the existence of acts that are bad but not forbidden. In Chisholm's words, "If there is such a thing as

¹⁰ For a defence of the claim that moral dilemmas exist see Marcus (1980), Nussbaum (1986 Ch. 2, Ch.3) and Williams (1973 Ch.11).

¹¹ Conee (1982) argues that a consistent moral theory will not generate moral dilemmas.

¹² Mellema (2005 p.293).

‘non-obligatory well-doing’ then it is plausible that there is also such a thing as ‘permissive ill-doing’.¹³ Driver makes a similar point. She claims that a familiar point amongst those writing about supererogation is that there is a distinction between acts we ought to do and acts we are obliged to do. If we accept this then there seems little reason to reject the possibility of acts we ought not to do but are not obliged not to do.¹⁴ The point both are making is that if we allow that moral goodness and moral obligation can come apart then we should accept that moral badness and moral prohibition may also come apart. This seems like a reasonable point to make. However, this does not, as yet, give us reason to think that we need to make room for a new deontic category. In the remainder of this section I will investigate whether accepting the possibility of acts that are bad but not forbidden gives us reason to introduce a new category into our deontic scheme.

7.1.2 Do We Need To Make Room for A New Category of Actions?

Chisholm claims that the possibility of acts that are morally bad but morally permissible creates the need for a new deontic category. In his words: “A system of moral concepts which provides a place for what is good but not obligatory, should also provide a place for what is bad but not forbidden.”¹⁵ According to Chisholm, in order to provide a place for such acts we must create a new deontic category. To support this claim Chisholm presents the following analysis of the deontic scheme that includes supererogation but not suberogation:

Obligatory: An act that is good to do and bad not to do

Forbidden: An act that would be bad to do and good not to do.

¹³ (1963 p.5).

¹⁴ (1992 p.291).

¹⁵ Chisholm (1963 p.5).

Supererogatory: An act that would be good to do and not good or bad not to do.

Indifferent: An act that is neither good nor bad to do and neither good nor bad not to do.¹⁶

These categories are unable to accommodate the existence of an act that is bad but permissible to perform. As a result, Chisholm claims we need to make room for the following additional category:

Suberogatory: An act that is bad to do but neither good nor bad not to do.¹⁷

A system of moral concepts that provides a place for what is good but not obligatory, should also provide a place for what is bad but not forbidden.

In brief, Chisholm's argument is that without the category of suberogation our deontic scheme will be unable to accommodate the claim that some acts are bad but permissible. These acts are simply unclassifiable on the above system. As a result, if we accept the claim that some acts are morally bad and morally permissible then we need to make our set of moral classifications compatible with this. The only way for Chisholm's

¹⁶ Summarized from Chisholm (1963 pp.10-11).

¹⁷ Chisholm (1963 pp.10-11). We might worry about the coherence of Chisholm's definitions of supererogation and suberogation. As we saw in Chapter One there is good reason to think that 'more good' is synonymous with 'better'. Presumably, 'more bad' is also synonymous with 'worse'. It will presumably be worse to omit an act of supererogation than to perform it. However, this makes it unclear how omitting an act of supererogation could be neither good nor bad. After all, if it is worse than the alternative then it must be both less good and more bad than the alternative. If this is the case then it is unclear how it could be neither good nor bad not to do. For now, though, let's set this worry to one side.

characterization of the range of deontic concepts to accommodate acts of this sort is to create a new category.

To see the appeal of this claim we can consider how Chisholm’s division of the deontic field looks in the absence of suberogation:

	Performance	Non-Performance
Obligatory	Good	Bad
Prohibited	Bad	Good
Supererogatory Performances	Good	Neither good nor bad
Supererogatory Non-Performances	Neither good nor bad	Good
Indifferent	Neither good nor bad	Neither good nor bad

Chisholm distinguishes between supererogatory performances (acts that are supererogatory to perform) and supererogatory non-performances (acts which are supererogatory to omit).¹⁸ To not demand one’s rights on certain occasions may, for instance, be classed as a supererogatory non-performance. In order to see the appeal of Chisholm’s claim we must make two stipulations. First, if the performance of an act is good then its non-performance cannot also be good and vice versa. Similarly if performing an act is bad then its non-performance cannot also be bad and vice versa. This does not rule out the possibility of situations in which more than one good act or more than one bad act are available. Rather, it is simply to say that if an act is good the non-performance of that act, by itself, can at best be morally neutral. Of course, in

¹⁸ (1963 p.10).

situations where only two bad acts, x and y , are available then by not performing act x we will be performing a bad act, y . However, what is bad in this case is the performance of y not the non-performance of x .

With these stipulations in hand we now are able to see the appeal of Chisholm's claim. In the absence of suberogation there are two possible combinations missing from this division of the deontic categories. First an act could be bad to do and neither good nor bad not to do. Second, an act could be neither good nor bad to do but bad not to do. In order to capture the full range of possibilities we must introduce a new category. This gives us the following, comprehensive division of the range of deontic possibilities:

	Performance	Non-Performance
Obligatory	Good	Bad
Prohibited	Bad	Good
Supererogatory Performances	Good	Neither good nor bad
Supererogatory Non-Performances	Neither good nor bad	Good
Indifferent	Neither good nor bad	Neither good nor bad
Suberogatory Performances	Bad	Neither good nor bad
Suberogatory Non-Performances	Neither good nor bad	Bad

This shows why Chisholm claims that it is plausible that if supererogation is possible then so is suberogation. Given

Chisholm's stipulation, his range of deontic concepts covers every other possible combination. It would indeed be odd to rule out the possibility of suberogatory performances and non-performance.

Chisholm's claim, then, is that without the concept of suberogation there will be some acts that are unclassifiable in the our deontic scheme. In the remainder of this section I will argue that given the analysis of supererogation that I have defended up to now, we can reject Chisholm's claim that we need to create a new deontic category in order to classify acts that are morally bad but morally permissible.

The first step of this argument is to point out that The Symmetry Argument will not work with the analysis of supererogation I defended in Chapter One. To recap, I argued that the best analysis of supererogation is the following:

Supererogation: An act, ϕ , is supererogatory for an agent A, at time t , if and only if ϕ -ing is better than the minimum that morality demands of A at t and there is no other obligation that forbids A from ϕ -ing or for which ϕ -ing is the minimum that morality demands of A.

The important point for our purposes here, though, is simply that supererogatory acts are better than the minimum required in a given situation and the above analysis should be seen as a way of spelling out what this means in more detail.

There are two ways in which we might attempt to mirror this analysis in order to give us an account of suberogation.

Worse than Minimum: A suberogatory act is morally optional and morally worse than the minimum morality demands in that situation.

Worse than Maximum: A suberogatory act is morally optional and morally worse than the morally best act available.

Unfortunately for those who claim we must make room for the suberogatory, both of these attempts to mirror supererogation fail to capture the claims about moral experience given in §7.1.1.

The problem facing *Worse Than Minimum* is the more severe as this looks like an incoherent concept. We can see why by noting that an act that is morally worse than the minimum that morality requires must be morally forbidden. After all if some act is the minimum that is needed in order for a duty to be fulfilled then to do less than this is to violate duty and act impermissibly.

However, an act that satisfies *Worse Than Minimum* is morally optional and so cannot be morally impermissible.¹⁹

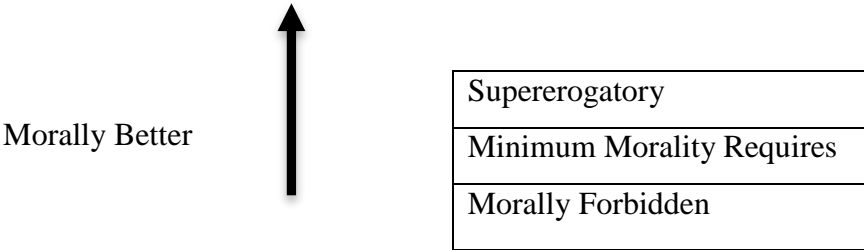
While *Worse Than Maximum* is a coherent concept there are two problems with defining suberogation in this way. First, this account will not pick out any acts that cannot be accommodated by the existing categories. The minimum that morality demands will be an act that is worse than the morally best act available and morally optional. In fact it seems plausible to think that an act could be supererogatory and also be worse than the best available act. For example, suppose I have promised to spend an hour helping my friend move house. We might think that given this promise, it is now morally obligatory to do so, though previously it was not. Now suppose that the morally best thing to do in this situation would be to spend the whole day helping. If I decide to help for two hours then this act will be both morally better than the minimum morality demands (an hour of helping) and morally worse than the morally best act available (a whole days helping). If we take *Worse Than Maximum* to be our account of

¹⁹ McNamara (2011 pp.227-8) makes a similar point.

suberogation then all of the acts it picks out will already be classifiable as either supererogatory or obligatory.

Moreover, *Worse Than Maximum* also fails to capture the claims about moral experience made in §7.1.1. In the previous example helping for two hours is worse than the maximum but we would be unlikely to say that it is a morally bad. Of course, this does not mean that there will never be situations in which an act that meets the conditions given by *Worse Than Maximum* are both morally bad and morally permissible. The point, though, is that *Worse Than Maximum* will not help us in identifying or classifying these acts.

So far I have argued that the account of supererogation that I defended in Chapter One provides no resources for a mirror image of supererogation that is able to capture the claims from moral experience made in §7.1.1. However, the problems for those who argue that there is a need to make room for suberogation go further than this. Not only does the account of supererogation I have given present no resources for defenders of suberogation to offer a version of The Symmetry Argument, it also leaves no space for a deontic category with distinct parameters that is able to capture the claims from moral experience made in §7.1.1. To see why, consider the following illustration of how the acts are divided according to the definition of supererogation that I have defended:



According to the definition of supererogation that I have defended, when an act is not morally indifferent, these three deontic categories exhaust the range of possibilities. Any further deontic categories we add in will be ones that overlap with existing categories. For example, if we were to introduce *Worse Than Maximum* then this category would include all acts from the minimum morality requires to the second best supererogatory act. To sum up, given the most plausible analysis of supererogation, there is no need to create a new deontic category to accommodate acts that cannot be categorized in the above scheme.

7.1.3 Suberogation As A Useful Classificatory Tool

In the last section we looked at Chisholm's argument that unless we make room for the concept of the suberogatory, there will be some acts that our deontic scheme will be unable to classify. We saw that this is not the case if we accept the definition of supererogation defended in Chapter One. We might think that this is all there is to be said on the matter. There is though, another reason that has been put forward by those wishing to defend the need for the supererogatory. Driver has argued that introducing the concept of the suberogatory allows us to, "explain and clarify some problems in ethics".²⁰

Driver claims that this category also allows us to solve a problem facing those who think that abortion is morally permissible but that having multiple abortions is bad. Driver claims that even those committed to a woman's right to an abortion might find something morally suspect about someone who has nine abortions without having an excuse. This seems odd, if abortion is permissible, as the pro-choice activist thinks it is, then why are multiple abortions bad? If we accept the possibility of acts that are

²⁰ (1992 pp.294-295).

bad but permissible then we can say that although getting an abortion is permissible, to do so without an excuse is bad. While a single abortion might not be that bad, by the time someone has nine abortions without an excuse then perhaps this is a significant moral failing.²¹ Accepting the existence of suberogation, then, allows us to say that there is nothing contradictory about this combination of beliefs.

Of course, we might think that we can say all this without the need for a new concept to describe these acts. We could, after all, simply say that these acts are bad but permissible. However, Driver claims that there is reason not only to accept that such acts exist but to introduce a new term to pick out these acts. She defends this claim in the following way:

In embracing the suberogatory, we allow for another useful distinction in our moral language. It provides a way of evaluating actions which accurately reflects many moral judgements we do make - judgements lying in the dark corners between right and wrong.²²

By introducing this term, then, we will not only be able to describe our moral judgements more precisely, we will also be kept alert to the fact that permissible acts may, nevertheless be bad.

Liberto challenges this defence of the suberogatory. Liberto claims that the examples appealed to by supporters of suberogation fail to show the need to make room for a new deontic concept. Liberto argues that these are difficult cases in applied ethics. The question of whether Roger has an obligation to donate his kidney to Bob or the train passenger to give up his preferred seat are questions that raise issues concerning, amongst

²¹ (1992 p.293).

²² (1992 p.295).

other issues, how much self-partiality is morally permissible.²³ Liberto claims that the reason that such cases are problematic is, “because categorizing them takes a lot of work, not necessarily because the categories are insufficient.”²⁴ According to Liberto making space for the suberogatory, “is neither useful nor necessary in handling the problem cases.”²⁵ While it may be tempting to posit a new category to avoid this applied ethical work, this is a temptation that should be avoided. Liberto’s point is that we should avoid positing the need for this ethical category until this work is complete.

Liberto is right to point out that we need both more information and the answers to some difficult applied ethical questions before we can say whether we need the category of the suberogatory to handle Driver’s cases. However, this is not enough to show that the concept is not a useful one. After all, the possibility remains that once we have the information that Liberto claims is needed we will discover that Driver’s assessment of these situations is correct.

Moreover, even if we accept Liberto’s claim that we do not yet have good reason to think that some acts are suberogatory, it does not follow that there is no use for the concept of suberogation. We might think that allowing for the category of suberogation allows us to capture a feature of moral phenomenology that is not reflected by the existing deontic categories. Even if we think that further work in applied ethics might show Driver’s diagnosis of the cases she gives to be in error, we might think that the very fact that we cannot immediately dismiss this diagnosis as being wildly counter-intuitive is enough to show that we can conceive of an

²³ (2012 p.400).

²⁴ (2012 p.400).

²⁵ (2012 p.402).

act's being morally bad without being morally wrong. Perhaps the fact that we can at least conceive of its being the case that an act could be bad without being forbidden is sufficient reason to make room for this concept in our moral thinking, even if it turns out that there are no acts that fall into this category. Even if Driver's assessment turns out to be wrong, we might think that making room for this category will have been useful in making us aware of possible answers to applied ethical questions which we might otherwise have overlooked.

Perhaps, then, there is a use for the concept. Of course, if we accept Driver's argument we must give some account of what it is for an act to be suberogatory. As we saw in the previous section, there is no mirror image of the supererogatory that captures this concept. Perhaps, though, rather than attempting to mirror the most plausible analysis of supererogation, we could accept the original definition of these acts as morally bad and morally permissible. Perhaps it could be claimed that the fact that suberogation picks out a recognizable feature of moral phenomenology justifies making room for such a category regardless of whether acts that meet this description will also meet the descriptions given by the other deontic categories.

For our purposes though, the important lesson to take from this is that if there is a need for the concept of suberogation, it is not as a fully separate deontic category but as a useful label.

7.1.4 Summary

To sum up, I have argued that given the account of supererogation I defended in Chapter One, we can accept all of the claims from moral experience that supporters of suberogation appeal to without needing to accept the existence of a new deontic category. In addition, this account of supererogation leaves no room for The

Suberogatory as a distinct deontic category. While it would be possible to accept the existence of such a category regardless of this, doing so will pose no threat to the adequacy of the four-fold deontic scheme.

7.2 Quasi-Supererogation

Let's now turn our attention to another proposed addition to the range of deontic categories, The Quasi-Supererogatory. Gregory Mellema has argued that we need to make room for this category to accommodate acts that are praiseworthy to perform and blameworthy to omit.²⁶ In this section, I will argue that Mellema's argument is unsuccessful.

7.2.1 Mellema's Argument

As we have seen, Chisholm argued that there are five deontic categories: the obligatory, the indifferent, the forbidden, the supererogatory and the suberogatory. Mellema argues that this division omits two important combinations. Mellema illustrates this by listing all the possible combinations of praiseworthiness and blameworthiness for morally optional acts (those that are permissible to perform or omit). Mellema characterizes Chisholm as allowing for the possible combinations amongst the morally optional acts:²⁷

²⁶ (1987). Horgan and Timmons (2010 p.32) also use the term 'quasi-supererogation' to refer to acts that go beyond duty but the non-performance of which requires justification. My use of the term will follow Mellema's not that of Horgan and Timmons.

²⁷ It is worth noting that although Chisholm (1963 pp.10-11) analyzes these acts in terms of 'good' and 'bad' (and again in Chisholm and Sosa (1966b p.327)), Mellema uses the terms 'praiseworthy' and 'blameworthy'. The root of this terminological mistake seems to be that Mellema sees 'praiseworthy' as roughly equivalent to 'morally good' (1991 p.24).

	Performance	Non-Performance
Supererogatory Performances	Praiseworthy	Neither praiseworthy nor blameworthy
Supererogatory Non-Performances	Neither praiseworthy nor blameworthy	Praiseworthy
Indifferent	Neither praiseworthy nor blameworthy	Neither praiseworthy nor blameworthy
Suberogatory Performances	Blameworthy	Neither praiseworthy nor Blameworthy
Suberogatory Non-Performances	Neither praiseworthy nor blameworthy	Blameworthy

When we look closely at this way of dividing up the deontic categories we can see that there are four combinations that are missing. The first is acts that are praiseworthy to perform but blameworthy not to perform. Mellema uses the term quasi-supererogation to refer to these acts. The second are acts that are blameworthy to perform and praiseworthy not to perform. Mellema uses the term quasi-suberogation to refer to these acts.²⁸

²⁸ (1987 p.144).

The final two combinations are acts that are praiseworthy both to perform and omit and acts that are blameworthy both to perform and omit. Mellema does not focus on these combinations, as he views such acts as unlikely to be possible for human agents.²⁹

Having demonstrated that there is space for these two new deontic categories, Mellema then argues that there are acts that are plausibly seen as members of these categories. Mellema's starting point is to imagine a case where an agent refrains from performing an offence. Mellema gives the following as an example of such a case. Suppose S is in a restaurant and is sitting next to an obnoxious man who is making mocking gestures and loud jokes about the physical disabilities of S's wife. As a result, S is strongly tempted to walk over to the table and empty the contents of the obnoxious man's plate onto his lap. Mellema claims that if S can resist this temptation then he does something praiseworthy.³⁰ On the other hand, if he succumbs to temptation then he acts in a blameworthy way. In both cases these acts are morally optional. In the first case he will have performed a quasi-supererogatory act and in the second case he will have performed a quasi-offence.

Mellema's claim, then, is that there are some acts that do not fit into Chisholm's categories. In order to accommodate these acts we need to add the following categories onto Chisholm's scheme:

Quasi-Supererogation: An act that is praiseworthy to perform and blameworthy to omit.

Quasi-Suberogation: An act that is blameworthy to perform and praiseworthy to omit.

²⁹ (1987 p.150 Fn.3).

³⁰ (1987 p.146).

Note that the force of this argument comes not from the example that Mellema gives but from its structure. If acts of offence exist then, claims Mellema, there will be cases where it is praiseworthy to refrain from performing the offence. Once we have accepted that it is possible for an act to be blameworthy and morally optional it seems reasonable to think that there will be occasions where refraining from performing the blameworthy act will be praiseworthy. Even those who do not find Mellema's example particularly persuasive have reason to take the structure of his argument seriously. Once we have accepted that refraining from performing a morally optional act can be praiseworthy then we have accepted the existence of two kinds of act that cannot fit into the standard deontic scheme. As a result, if we want to give a deontic scheme that can capture the full range of possibilities we must make room for two new categories. Nevertheless, in the next section I will show that Mellema's argument is unsuccessful.

7.2.2 Mellema's Dilemma

As with my argument against the need for the category of suberogation, I do not intend to say anything against Mellema's claims from moral experience. As with the suberogatory, we can accept everything Mellema says about his example without accepting that it gives us any reason to make room for a new deontic category. The reason for this, as I will demonstrate in this section, is that Mellema's argument faces a dilemma.

The first step in identifying this dilemma is to note that the way in which 'The Standard Position' is outlined is somewhat misleading. It is true that Chisholm's division of the deontic categories is unable to accommodate the existence of acts that are morally optional and that meet the criteria given by quasi-supererogation and quasi-suberogatory. However, the reason for this is that the definition Mellema gives for acts of quasi-

supererogation and quasi-suberogatory are already taken by the categories of the obligatory and the forbidden. For Chisholm an obligatory act is one that is praiseworthy to perform but blameworthy not to perform, while a forbidden act is one that is blameworthy to perform and praiseworthy not to perform.³¹ By leaving out these categories from the way he lays out The Standard Position, Mellema creates the misleading impression that we need to create two new deontic categories to accommodate acts that meet these criteria.

We might think that although the way Mellema sets out The Standard Position is misleading, this does not create any major problems for his thesis. After all, the claim he is defending is that acts of quasi-supererogation are not only praiseworthy to perform and blameworthy to omit but that they are also neither obligatory nor forbidden to perform or omit. While the categories of the quasi-supererogatory and the quasi-suberogatory have similar definitions to the categories of the obligatory and The Forbidden, the former are morally optional while the latter are not. In fact, Mellema makes it clear that this is what distinguishes quasi-supererogation from the obligatory and quasi-offence from the forbidden.³²

However, while this response would allow Mellema to maintain that quasi-supererogation and quasi-suberogation are distinct categories, it does so at a cost. The problem with this response is that it means that the definitions Mellema gives for these categories are incomplete. The complete definitions should read as follows:

³¹ (1963 p.10).

³² (1987 p.147).

Quasi-Supererogation (Revised): An act that is morally permissible and praiseworthy to perform and morally permissible and blameworthy to omit.

Quasi-Suberogation (Revised): An act that is morally permissible and blameworthy to perform and morally permissible and praiseworthy to omit.

These definitions capture the claims that Mellema makes about moral experience while marking these categories out from the obligatory and the forbidden. Unfortunately for Mellema's argument, though, once we allow moral permissibility to feature in our definition of deontic terms then we are able to define the five categories that exist in The Standard Position in a way that is able to capture Mellema's claims from moral experience without creating any new categories.

Obligation (Revised): An act that is morally permissible to perform and morally impermissible to omit.

Prohibition (Revised): An act that is morally impermissible to perform and morally permissible to omit.

Supererogation (Revised): An act that is morally optional and praiseworthy to perform.

Suberogation (Revised): An act that is morally optional and blameworthy to perform.

Morally Indifferent (Revised): An act that is moral optional and neither praiseworthy nor blameworthy to perform or to omit.

These revised definitions retain what is distinctive about supererogation and suberogation. Although the performance of supererogatory acts is praiseworthy and suberogatory acts blameworthy, unlike obligations and prohibitions these acts are

morally optional. The most important point for our purposes, though, is that these definitions are able to accommodate Mellema's claim that there are acts that meet the criteria given by his definitions of quasi-supererogation and quasi-suberogation without the need to create new deontic categories.

This, then, is a dilemma for Mellema's argument. Mellema must either accept that we need to introduce moral permissibility into our definitions of deontic categories or reject this. If he rejects this then the categories of quasi-supererogation and quasi-suberogation will be indistinguishable from those of obligation and prohibition. If he accepts this then we can revise our definitions of supererogation and suberogation in a way that allows these categories to remain distinct from obligation and prohibition and accommodate acts that meet the criteria given by Mellema's definitions for quasi-supererogation and quasi-suberogation. While the first option will leave Mellema unable to define his new categories in a way that distinguishes them from existing categories, the second option removes the need to create any new categories.

Someone sympathetic to Mellema's position might claim that these definitions fail to capture all of the relevant necessary conditions for supererogation and suberogation. After all, Mellema began his argument by claiming that it is a necessary condition for supererogatory acts that they are not blameworthy to omit. Similarly, Mellema also claimed that it is a necessary condition for offences that they are not praiseworthy to omit. If we accept these two necessary conditions then the revised definitions are inadequate and there is a need for two new deontic categories after all.

However, while it is true that accepting these necessary conditions would present the need for two new deontic concepts, Mellema

does not provide any reason to suggest that these are rightly seen as necessary conditions of supererogation and suberogation. In effect, then, accepting Mellema's argument that there exist morally optional acts that are praiseworthy to perform and blameworthy to omit as well as morally optional acts that are blameworthy to perform and praiseworthy to omit presents us with a choice. We either take Mellema's argument to show that the conditions that Mellema stipulates as necessary for supererogation and offence are overly restrictive or we accept the need for new deontic categories. As Mellema fails to give us any reason reject the first option his argument fails to support his thesis that we need to make room for these two deontic categories. At best, his argument in fact supports the disjunctive thesis that we must either make room for two new deontic categories or revise the necessary conditions for supererogation and offence that Mellema presents at the start of his paper.

To make matters worse for Mellema's argument, if we accept that parsimony is a virtue of any classificatory scheme then we seem to have good methodological reason to opt for the latter of these disjuncts. Of course this is at best a defeasible reason to favour this option and if good reason can be found to favour instead creating new deontic categories then this will likely outweigh any concerns for parsimony. However, in the absence of such reasons the right conclusion to draw from Mellema's argument for the existence of acts that meet the criteria for quasi-supererogation and quasi-suberogation is not to create these new deontic categories. Rather we should reject the claim that no supererogatory acts are blameworthy to omit and no offences praiseworthy to omit.

7.2.3 Summary

In this section we have looked at Mellema's argument in support of the claim that we need to make room for two new deontic categories, quasi-supererogation and quasi-suberogation. I have shown that Mellema's argument in support of his claim that acts meet the criteria he offers for these two categories at best provides support for a disjunctive conclusion. Either we make room for these additional categories or we reject the claim that no supererogatory acts are blameworthy to omit and no suberogatory are praiseworthy to omit. Moreover, if we accept that parsimony is a virtue of any classificatory scheme then we appear to have good reason to opt for the latter option. In the final section of this chapter we will look at another attempt to show the need for a new deontic category and see that this also fails to give us any reason to expand our range of deontic categories.

7.3 Forced Supererogation³³

So far I have argued that accepting the need to accommodate the category of supererogation does not give us any reason to think that we should also make room for the categories of suberogatory or quasi-supererogation. In the final section of this chapter I will argue that there is no good reason to accept the category of forced supererogation either. In a recent paper, Cohen has offered a defence of the view that we should make room for this category.³⁴ Cohen argues that there is a need to make room for a new deontic category between the obligatory and the supererogatory. Acts of forced supererogation are acts that are supererogatory to perform but wrong not to perform. In this section I will argue that Cohen's argument does not show the need for this new deontic category. At best, Cohen's argument shows that there is not a uniform

³³ Section 7.3 is largely based on Archer (2014b).

³⁴ Cohen (Forthcoming).

relationship between moral obligations, praiseworthiness and blameworthiness.

Cohen argues that we need to make room for the following additional form of deontic classification:

Forced Supererogation: Performance is praiseworthy and nonperformance is wrong but not blameworthy.³⁵

As Cohen accepts, to justify allowing this new concept into our explanatory scheme we must show that there are good reasons to do so that outweigh the considerations of parsimony that count against doing so.³⁶

7.3.1 The Argument From Moral Experience

Cohen argues that without Forced Supererogation we cannot do justice to our moral experience. Cohen gives three examples to support this claim. For the sake of brevity I will look only at the first:

Case 1: Your brother is suffering from a rapidly progressive lung disease, which is fatal without treatment. The only treatment is lung transplantation, which has a significant negative impact on one's prospective quality of life. You are the sole matching donor.³⁷

To show that we need to make room for Forced Supererogation Cohen needs to persuade us that the act of donating a lung in this case cannot be accommodated by any of the existing deontic concepts and that it meets the three conditions listed in the definition.

I am willing to accept Cohen's assessment that performing this act is especially praiseworthy and that nonperformance of this act

³⁵ Cohen (Forthcoming p.4).

³⁶ (Forthcoming p.15).

³⁷ (Forthcoming pp.1-2).

would not be blameworthy. It is the task of showing that the act is wrong to omit that causes problems.

Cohen defends this claim in the following:

Although one is not wrong not to donate one's own lung, even if there are no other matching donors, things are different when it is for one's own brother. (...) This personal factor creates particular obligations, which are notoriously difficult to account for in classical deontic classifications.³⁸

Cohen is arguing that the personal request creates a special obligation, which in turn makes the non-performance of the act morally wrong. Importantly for Cohen these special obligations are often more demanding than ordinary obligations.

The problem with this argument is in the final step. We are told to accept that these acts are wrong because there is a special obligation to perform them. However, this argument only works if these special obligations generate moral obligations, which would make their non-performance morally wrong. However, if this is the case then we can no longer say that these acts go beyond duty. If, on the other hand, they do not generate moral requirements then Cohen's argument in support of the claim that the non-performance of these acts is morally wrong is unsound.

Cohen makes two responses to the worry that these acts should be viewed as obligatory. First, he argues that classing these acts as obligatory ignores the special praise an agent deserves for performing such an act. The cases he mentions are all praiseworthy in a way that moral obligations normally are not. This Cohen claims, gives us reason to reject the claim that these acts are obligatory.³⁹

³⁸ (Forthcoming p.6).

³⁹ (Forthcoming p.7).

However, if Cohen is right to say that these special obligations do not generate moral obligations then we have no reason to think that non-performance is morally wrong. Non-performance will of course violate a *special* obligation but if we accept Cohen's claim then there will be no violation of a *moral* obligation. The claim that these acts are worthy of special praise does nothing to help Cohen avoid this dilemma. We must either accept that the performance of some moral obligations is worthy of special praise or reject the claim that the non-performance of such acts is morally wrong.

Cohen does not acknowledge this dilemma but there is good reason to think that he would opt for the first horn. This certainly seems to be the natural way to interpret the following, "acts of FSE (Forced Supererogation) are especially praiseworthy in a way that Obligations *normally* are not."⁴⁰ This is also the most charitable interpretation. Consider a firefighter who, in the course of her job, runs into a burning building to save someone's life. This act is praiseworthy but given her job it is also obligatory. If Cohen is conceding that sometimes performing an obligatory act is especially praiseworthy then he is committed, at pains of inconsistency, to accepting the following revised definition of obligation:

Obligation 2: Performance is not *normally* especially praiseworthy and non-performance is both blameworthy and wrong.

However, revising the definition of obligation in this way means that the praiseworthiness of the act in Case 1 no longer gives us any reason to create a new deontic category. Instead we can accept

⁴⁰ (Forthcoming p.7).

that this is one of the unusual cases where fulfilling an obligation is especially praiseworthy.

Cohen's next response is to claim that we would not regard the nonperformance of the act in Case 1 to be blameworthy.⁴¹ Cohen gives two reasons to accept this. First, a failure to perform the act would not be violating anyone's rights and so no one could legitimately blame the agent for a failure to perform the act. Second, the obligations generated by friendship are a special case, as a failure to act in line with such obligations can change the nature of the relationship in such a way that there are no longer any grounds for blame.

However, in order for these acts to count as wrong to omit Cohen is going to have to accept another revision to the definition of obligation:

Obligation 3: Performance is not *normally* especially praiseworthy and non-performance is wrong and is *normally* blameworthy.

If this is our definition of obligation then the fact that it would not be blameworthy to fail to perform the act in Case 1 gives us no reason to think that it is not obligatory.

In summary, Cohen's argument faces a dilemma. To defend the claim that the non-performance of these acts is morally wrong he needs to show that the agent has a moral obligation to perform the act. However, in defending this claim Cohen commits himself to revising his definition of moral obligation. This revised definition removes the need for a new deontic category to accommodate acts that meet the criteria given by forced supererogation. As a result, Cohen must either accept that such

⁴¹ (Forthcoming p.8).

acts are not conceptually possible or accept that there is no need to make room for a new deontic category to accommodate them.

7.3.2 The Missing Combination Argument

Cohen also argues that we need this new category in order to capture the full range of deontic possibilities. Cohen lays out the positive half of the deontic field in the following way:⁴²

	Is Failure to Act Blameworthy?					
	Yes	No				
Is Failure to act Wrong?	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; padding: 5px;">Obligation</td> <td style="width: 50%; padding: 5px;">Forced Supererogation</td> </tr> <tr> <td style="width: 50%; padding: 5px;">Quasi- Supererogation</td> <td style="width: 50%; padding: 5px;">Supererogation</td> </tr> </table>		Obligation	Forced Supererogation	Quasi- Supererogation	Supererogation
Obligation	Forced Supererogation					
Quasi- Supererogation	Supererogation					
Yes						
No						

Cohen claims that unless we accept the existence of Forced Supererogation we will be left with a missing box in this diagram, meaning that we have failed to capture all of the possible combinations.

However, this table ignores the question of whether or not these acts are praiseworthy to perform. As such it is a misleading representation of the factors Cohen takes to determine an act’s deontic status. Given Cohen’s taxonomy, the following is a more accurate representation of the possibilities:

⁴² (Forthcoming p.4).

	Is performance praiseworthy?	
	Yes	No
Failure to perform blameworthy and Wrong	?	Obligation
Failure to perform wrong but not blameworthy	Forced Supererogation	?
Failure to perform neither wrong nor blameworthy	Supererogation	?
Failure to perform blameworthy but not wrong	Quasi-Supererogation	?

As the empty spaces in this grid demonstrate, accepting the new deontic category of forced supererogation does not capture the full range of combinations of the components Cohen takes to determine an act's deontic status. Clearly, then, allowing forced supererogation into our conceptual scheme will not provide a full answer to the question of how to represent the range of options. The problem for this argument is that obligation is going to have to take up more than one space on this matrix. As we have already seen with the firefighter example, it is possible to be worthy of praise for performing an act that is morally required. This means that obligation will also have to fill the gap on the top left of this matrix. Once we have accepted that obligation occupies more than one space in the diagram Cohen's argument loses its force. The need to account for acts that are praiseworthy to perform and wrong but not blameworthy to omit no longer appears to require a new deontic category. We could instead say that obligation fills this box as well. Moreover, as we have seen, Cohen's defence of the claim that there are acts of this sort relies on the definition

given by Obligation 3. However, Obligation 3 can fill all of the boxes in the first two rows of the diagram, including that occupied by Forced Supererogation. We should not, then, accept the need to allow forced supererogation into our conceptual scheme on the basis of Cohen's claim that we must do so to capture the full range of deontic possibilities.⁴³

Conclusion

In this chapter we have looked at three attempts to show that making room for supererogation is not enough and that in order to accommodate every kind of action we must expand our range of deontic categories even further. I have argued that none of these attempts are successful. The set of categories that includes the obligatory, the forbidden, the indifferent and the supererogatory is able to accommodate all possible acts. Of course, this does not mean that at some point in the future a persuasive argument might not be given to show that there is good reason to divide these categories into subcategories. In the absence of such an argument, though, there is no reason to think that this set of categories is insufficient.

⁴³ Thanks to Mike Ridge and an anonymous referee for helpful comments here.

Concluding Remarks

In the first half of this thesis I investigated how supererogation should be defined. I argued that it should be defined in the following way:

Supererogation: An act, ϕ , is supererogatory for an agent A, at time t , if and only if ϕ -ing is better than the minimum that morality demands of A at t and there is no other obligation that forbids A from ϕ -ing or for which ϕ -ing is the minimum that morality demands of A.

I defended this definition both against similar definitions and against Dorsey's very different account. I then defended this account against two ways in which it might be argued that it is insufficient. First, against the claim that only praiseworthy acts can be supererogatory. Second, against the claim that only acts performed with altruistic intent can be supererogatory.

In the second half of this thesis I investigated three issues in moral philosophy that up to now have either ignored or not paid adequate attention to either the existence of acts of supererogation or the nature of the supererogatory.

First, I looked at the issue of the relationship between moral reasons, non-moral reasons and moral obligation. I argued that one of the most prominent positions in this debate, the view that what prevents the morally best act from being morally obligatory are the costs to the agent, rests upon an implausible view about the nature of the supererogatory.

I then investigated the debate about motivational judgement internalism, the view that there is an internal connection between moral judgements and motivation. I argued that appreciating the difference between moral goodness and moral requirements gives

motivational judgement internalists good reason to restrict their claim to moral requirements. One way to justify this move is to claim that internalism is explained by the truth of moral rationalism, the view that moral requirements are always in line with what there is most reason to do.

However, as I explained, this raises a problem as moral rationalism is often thought to be depend upon the truth of moral overridingness. However, I argued that it is possible to accept moral rationalism without accepting overridingness. Indeed the explanation for moral rationalism that is compatible with supererogation appears to be better supported by the intuitions commonly appealed to by those seeking to argue for rationalism than the explanation offered by overridingness.

I finished by looking at the question of whether we need to make room for more deontic categories than the forbidden, the indifferent, the required and the supererogatory. I showed that many of the arguments put forward in support of these categories rest upon mistaken assumptions about the nature of supererogation.

I hope to have shown that the existence of acts of supererogation has important implications for a number of debates within moral philosophy. There are a number of relevant issues that I have not sought to explore in this thesis. Most obviously, I have not sought to provide a solution to the paradox of supererogation, the puzzle of trying to explain the connection between moral obligations and reasons for action that leaves room for the supererogatory. I have, though, sought to show some problems for existing solutions. The question of how this puzzle ought to be resolved is one that I intend to investigate in later research. Similarly, in normative ethics it has long been recognized that both consequentialists and deontologists should have something to say about the

supererogatory. In recent years those seeking to give a virtue ethical account of right action have also acknowledged the importance of supererogation.¹ However, it is far from clear that any of the attempts to provide a virtue ethical account of right action that leaves room for a distinction between supererogatory and obligatory acts are successful.

Of course, it is not only those investigating theoretical moral philosophy who should pay attention to the importance of supererogation. There remains a wide range of debates in applied ethics for which it is of the utmost importance that the distinction between moral goodness and moral obligation is given sufficient attention. To give just one example, Ben Saunders has recently pointed out how a prominent argument in support of The Principle of Procreative Beneficence, which states that prospective parents have a *prima facie* obligation to select the child who will have the best expected life, is unsuccessful because it fails to pay attention to this distinction.²

¹ For example, Brännmark (2006) and Kawall (2009).

² Saunders (Forthcoming). For the argument Saunders criticizes see Savulescu (2001).

Bibliography

- Anscombe, G.E. (2000) *Intention* (Cambridge MA: Harvard University Press).
- Åqvist, L. (1968). 'Chisholm-Sosa Logics of Intrinsic Betterness and Value'. *Noûs* 2 (3):253-270.
- Archer, A. (2013) 'Supererogation and Intentions of the Agent'. *Philosophia* Vol. 41 (2):447-462.
- Archer, A. (2014a.) 'Moral Rationalism Without Overridingness'. *Ratio* Vol. 27 (1):100-114.
- Archer, A. (2014b.) 'Forcing Cohen to Abandon Forced Supererogation'. *Journal of Ethics and Social Philosophy* (Online).
- Arpaly, N. (2003). *Unprincipled Virtue: An Inquiry Into Moral Agency*. (Oxford: Oxford University Press).
- Ashford, E. (2003) 'The Demandingness of Scanlon's Contractualism.' *Ethics* Vol. 113 (2):273-302.
- Attfield, Robin (1979) 'Supererogation and Double Standards'. *Mind* 88 (352):481-499.
- Badhwar, N. K. (1993) 'Altruism Versus Self-Interest: Sometimes a False Dichotomy'. *Social Philosophy and Policy* 10 (1):90-117.
- Baier, K. (1958) *The Moral Point of View* (Ithaca: Cornell University Press).
- Baron, M. (1995) *Kantian Ethics Almost Without Apology* (New York: Cornell University Press).

BBC News (2012) 'Newark Mayor Cory Booker Saves Woman from Fire.' Accessed on 31/1/13. Available at:

<http://www.bbc.co.uk/news/world-us-canada-17700783>

Benn, Claire (2014) 'What is Wrong with Promising to Supererogate'. *Philosophia* Vol. 42 (1):55-61.

Bennett, J. (1974) 'The Conscience of Huckleberry Finn'. *Philosophy* 49 (188):123-134.

Blackburn, S. (1984) *Spreading the Word* (Oxford: Clarendon Press).

Blackburn, S. (1998) *Ruling Passions* (Oxford: Clarendon Press).

Bradley, B. (2012) 'Doing Away with Harm'. *Philosophy and Phenomenological Research* 85 (2):390-412.

Brännmark, J. (2006) 'From Virtue to Decency'. *Metaphilosophy* 37 (5):589-604.

Brogan, A. P. (1919) 'The Fundamental Value Universal'. *The Journal of Philosophy, Psychology and Scientific Methods*, 16(4), 96-104.

Broome, J. (2004) *Weighing Lives* (Oxford: Oxford University Press).

Brink, D. (1989) *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press).

Brink, D. (1997) 'Kantian Rationalism: Inescapability, Authority, and Supremacy', in G. Cullity and B. Gaut (Eds.) *Ethics and Practical Reason* (Oxford: Oxford University Press, 1997)

Carbonell, V. (2009) 'What Moral Saints look Like'. *Canadian Journal of Philosophy* 39 (3): 371-398.

Carbonnell, V. (2012) 'The Ratcheting-up Effect'. *Pacific Philosophical Quarterly* Vol. 93 (2): 228-254.

Carney, F. S. (1983) 'Some Aspects of Islamic Ethics'. *The Journal of Religion*, 159-174.

Chisholm, R. M. (1963) 'Supererogation and Offence: A Conceptual Scheme for Ethics'. *Ratio* Vol. 5 (1): pp.1-14

Chisholm, R. M., & Sosa, E. (1966a) 'On the Logic of "Intrinsically Better"'. *Arizona Philosophical Quarterly*, 3(3):244-249.

Chisholm, R. M., and Sosa, E. (1966b) 'Intrinsic Preferability and The Problem of Supererogation'. *Synthese* 16 (3):321-331.

Cholbi, M. (2011) 'Depression, Listlessness, and Moral Motivation.' *Ratio* Vol. 24 (1): 28-45

Chrisman, M. (2012) "'Ought" and Control'. *Australasian Journal of Philosophy* 90 (3):433-451.

Cohen, S. (Forthcoming) 'Forced Supererogation'. *European Journal of Philosophy*

Colby, A. and Damon, W. (1992) *Some Do Care: Contemporary Lives of Moral Commitment* (New York: The Free Press)

Copp, D. (1997) 'The Ring of Gyges: Overridingness and the Unity of Reason'. *Social Philosophy and Policy* 14:86-106.

Copp, D. (2001) 'Realist-Expressivism: A Neglected Option for Moral Realism.' *Social Philosophy and Policy* 18 (02):1-43.

Dancy, J. (1993) *Moral Reasons* (Oxford: Blackwell).

Darwall, S. (2006a) *The Second Person Standpoint* (Cambridge MA: Harvard University Press).

- Darwall, S. (2006b) 'Morality and Practical Reason: A Kantian Approach,' In D. Copp (Ed.) *The Oxford Handbook of Ethical Theory*, (Oxford: Oxford University Press), pp.282-320.
- Darwall, S. (2013) 'Bipolar Obligation', in *Morality, Authority, and Law: Essays in Second-Personal Ethics I* (Oxford: Oxford University Press).
- Davis, W. (1998) *Conversational Implicature: Intention, Convention and Principle in the Failure of Gricean Theory* (Cambridge: Cambridge University Press).
- Dorsey, D. (2012) 'Subjectivism without Desire'. *Philosophical Review* 121 (3):407-442.
- Dorsey, D. (2013) 'The Supererogatory, and How to Accommodate It'. *Utilitas* 25 (3):355-382.
- Dreier, J. (1990) 'Internalism and Speaker Relativism'. *Ethics* Vol. 101 (1):6-26.
- Driver, J. (1990) 'The Suberogatory'. *Australasian Journal of Philosophy* Vol.70 (3): 286-295.
- Enoch, D. (2011) *Taking Morality Seriously* (Oxford: Oxford University Press).
- Feldman, F. (1986) *Doing the Best We Can* (Dordrecht: D. Reidel Publishing).
- Feinberg, J. (1961) 'Supererogation and Rules'. *Ethics* 71 (4):276-288.
- Ferry, M. (2013) 'Does Morality Demand Our Very Best? Moral Prescriptions and the Line of Duty'. *Philosophical Studies* 165 (2):573-589.

- Finlay, S. (2010) 'What Ought Probably Means, and Why You Can't Detach It.' *Synthese* Vol.177 (1): 67-89.
- Fishkin, J. S. (1982) *The Limits of Obligation* (Binghamton NY: Yale University Press).
- Fletcher, G. (2013) 'A Fresh Start for the Objective-List Theory of Well-Being'. *Utilitas* 25 (2):206-220.
- Frankena, W. K. (1958) 'Obligation and Motivation in Recent Moral Philosophy', in A. I. Melden (Ed.), *Essays in Moral Philosophy*. (University of Washington Press).
- Fried, C. (1978) *Right and Wrong* (Cambridge, Mass: Harvard University Press).
- Frimer, J. A., and Walker, L. J. (2009) 'Reconciling the self and morality: An empirical model of moral centrality development'. *Developmental Psychology* 45 (6).
- Frimer, J. A., Walker, L. J., Dunlop, W. L., Lee, B. H., & Riches, A. (2011). 'The Integration of Agency and Communion In Moral Personality: Evidence of Enlightened Self-interest'. *Journal of Personality and Social Psychology* 101(1).
- Frimer, J. A., Walker, L. J., Lee, B. H., Riches, A., & Dunlop, W. L. (2012). 'Hierarchical integration of agency and communion: A study of influential moral figures'. *Journal of Personality*, 80(4):1117-1145.
- Gert, J. (2004) *Brute Rationality: Normativity and Human Action* (Cambridge: Cambridge University Press)
- Gibbard, A. (1990) *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. (Cambridge MA: Harvard University Press).

- Gibbard, A. (2003) *Thinking How To Live* (Cambridge, Mass: Harvard University Press)
- Graham, P. (2010) 'In Defense of Objectivism about Moral Obligation'. *Ethics* 121:88-115.
- Gray, J. (1996) *Mill On Liberty: A Defence* (New York: Routledge).
- Grice. H. P. (1989) *Studies in the Way With Words* (Cambridge MA: Harvard University Press).
- Guevara, D. (1999) 'The Impossibility of Supererogation in Kant's Moral Theory'. *Philosophy and Phenomenological Research*, 59: 593–624.
- Gustafsson, J. E. (2014) 'Neither "Good" in Terms of "Better" Nor "Better" Terms of "Good"'. *Noûs* 48 (3): 466-473.
- Hale, S. (1991) 'Against Supererogation'. *American Philosophical Quarterly* 28 (4): 273-285
- Hampton, J. (1993) 'Selflessness and the Loss of Self.' *Social Philosophy and Policy* 10 (1):135-16.
- Hanser, M. (2008) 'The Metaphysics of Harm.' *Philosophy and Phenomenological Research* 77 (2):421-450.
- Hansson, S. O. (1990) 'Defining Good and Bad in Terms of Better.' *Notre Dame Journal of Formal Logic*, 31(1), 136-149.
- Hansson, S. O. (Forthcoming) 'Representing Supererogation.' *Journal of Logic and Computation*.
- Hare, R. M. (1981) *Moral Thinking: Its Levels, Method and Point* (Oxford: Clarendon Press).

Harwood, S. (2003) 'Eleven Objections to Utilitarianism.' In Louis P. Pojman, ed., *Moral Philosophy: A reader*, 3rd edn. (Indianapolis) pp.179 - 92.

Hazlett, A. (2007) 'Grice's Razor.' *Metaphilosophy* 38 (5):669-690.

Heyd, D. (1982) *Supererogation: Its Status in Ethical Theory* (Cambridge: Cambridge University Press).

Heyd, D. (2011) 'Supererogation', in *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/supererogation/>.

Hill, T. (1971) 'Kant on Imperfect Duty and Supererogation.' *Kant-Studien*, 62: 55-76.

Hooker, B. and Streumer, B. (2004) 'Procedural and Substantive Practical Rationality', in A. Mele and P. Rawling (eds.), *The Oxford Handbook of Rationality* (Oxford: Oxford University Press).

Horgan, T. and Timmons, M. (2010) 'Untying a Knot from the Inside Out: Reflections on the "Paradox" of Supererogation'. *Social Philosophy and Policy* 27:29–63.

Jacobs, R.A. (1987) 'Obligation, Supererogation and Self-sacrifice'. *Philosophy* (Vol. 62 No. 239 pp. 96-101).

Jackson, M. W. (1986) 'The Nature of Supererogation'. *Journal of Value Inquiry* 20 (4):289-296.

Jackson, M. W. (1988) 'Above and beyond the call of duty.' *Journal of Social Philosophy* 19 (2): 3-12.

- Joerden, J. C. (2012) 'Deontological Square, Hexagon, and Decagon: A Deontic Framework For Supererogation'. *Logica Universalis* 6 (1-2):201-216.
- Kamm, F. M. (1985) 'Supererogation and Obligation'. *Journal of Philosophy*, 82 (3): 118-138.
- Kant, I. (1993), *Grounding for the Metaphysics of Morals*.
Translated by J.W. Ellington (Indianapolis: Hackett Publishing)
- Kauppinen, A. (2008) 'Moral Internalism and the Brain'. *Social Theory and Practice* Vol.34 (1): 1-24
- Kawall, J. (2003) 'Self-Regarding Supererogatory Acts'. *Journal of Social Philosophy* 34 (3): 487-498.
- Kawall, J. (2005) 'Promising and Supererogation'. *Philosophia* 32 (1-4):389-398.
- Klocksien, J. (2012) 'A Defense of the Counterfactual Comparative Account of Harm'. *American Philosophical Quarterly* 49 (4):285 – 300.
- Kotarbinski, T. (1914) 'Utilitarianism and the Ethics of Pity.'
Nowe Tory (i-ii). Translated by W. Rabinowicz (2000) and
reprinted in *Utilitas* 12 (01): 80-84.
- Korsgaard, C. M. (1996) *The Sources of Normativity* (Cambridge: Cambridge University Press).
- Ladd, J. (1957) *The Structure of a Moral Code* (Cambridge MA. Harvard University Press).
- Lenman, J. (1999) 'The Externalist and the Amoralist'.
Philosophia Vol. 27 (3-4): 441-457).

- Levinson, S. C. (2000) *Presumptive Meanings: The Theory of Generalized Conversational Implicature* (Cambridge MA: MIT Press).
- Liberto, H. R. (2012). 'Denying the Suberogatory'. *Philosophia* 40 (2):395-402.
- Lockhart, T. (2000) *Moral Uncertainty and Its Consequences* (Oxford: Oxford University Press)
- Marcus, R. B. (1980) 'Moral Dilemmas and Consistency.' *Journal of Philosophy* 77 (3):121-136.
- Markovitis, J. (2010) 'Acting for the Right Reasons'. *Philosophical Review* Vol. 119 (2).
- Mason, E. (2008) 'An Argument Against Motivational Internalism'. *Proceedings of the Aristotelian Society* Vol. 108 (2): 135-154.
- Mason, E. (2013) 'Objectivism and Prospectivism About Rightness'. *Journal of Ethics and Social Philosophy* Vol.7 (2).
- McLeod, O. (2001) 'Just Plain Ought'. *Journal of Ethics* 5 pp. 269 – 291.
- McElwee, B. (2010) 'The Rights and Wrongs of Consequentialism'. *Philosophical Studies* 151 (3):393 - 412.
- McGoldrick, P. (1984) 'Saints and Heroes: A Plea for the Supererogatory'. *Philosophy* Vol. 59 (230): 523- 528.
- McNamara, P. (1996a) 'Making Room For Going Beyond The Call'. *Mind* Vol. 105 (419): 415-450.
- McNamara, P. (1996b) 'Doing well enough: Toward a Logic For Common-sense Morality.' *Studia Logica* 57 (1):167-192.

- McNamara, P. (1996c) 'Must I Do What I Ought? (Or Will the Least I Can Do Do?)', In M. A. Brown and J. Carmo (Eds.) *Deontic Logic, Agency and Normative Systems*. (Berlin: Springer-Verlag pp.154-173).
- McNamara, P. (2011) 'Supererogation, Inside and Out: Toward an Adequate Scheme for Common-Sense Morality', in *Oxford Studies in Normative Ethics Volume 1* (Oxford: Oxford University Press).
- McNaughton, D. (1988) *Moral Vision* (Oxford: Blackwell).
- Mellema, G. (1987) 'Quasi-supererogation'. *Philosophical Studies* 52 (1):141 - 150.
- Mellema, G. (1991) *Beyond the Call of Duty: Supererogation, Obligation and Offence* (New York: State University of New York Press).
- Mellema, G. (2005) 'Moral Dilemmas and Offence'. *Ethical Theory and Moral Practice* 8 (3):291 - 298.
- Mill, J. S. (1859) *On Liberty* Reprinted (2008) *On Liberty and Other Essays* (Oxford: Oxford University Press).
- Mill, J. S. (1865; 1985) *Auguste Comte and Positivism* In J. M Robson (ed.) *The Collected Works of John Stuart Mill Volume X* (London: Routledge).
- Mill, J. S. (2001) *Utilitarianism*. (Indianapolis: Hackett Publishing).
- Miller, C. B. (2008) 'Motivational Internalism'. *Philosophical Studies* (Vol.139 No.2 pp.233-255).
- Mitchell, E. T. (1950) *A System of Ethics* (New York: Scribner).

- Moghaddam, F.M., Novoa, C. and Warren, Z. (2012) 'Duties and Rights.' In J. Valsiner (Ed.) *Oxford Handbook of Culture and Psychology* (Oxford: Oxford University Press pp.796-816).
- Monroe, K. R., Barton, M. C., & Klingemann, U. (1990) 'Altruism And The Theory of Rational Action: Rescuers of Jews in Nazi Europe'. *Ethics*, 101(1): 103-122.
- Monroe, K. R. (1991) 'John Donne's people: Explaining differences between rational actors and altruists through cognitive frameworks.' *Journal of Politics*, 53(2), 394-433.
- Monroe, K. (2004) *The Hand of Compassion: Portraits of Moral Choice During the Holocaust*. (Princeton NJ: Princeton University Press).
- Montague, P. (1989) 'Acts, Agents and Supererogation.' *American Philosophical Quarterly* 26(2): 100-111.
- New, C. (1974) 'Saints, Heroes and Utilitarians'. *Philosophy* 49 (188): 179-189.
- Nussbaum, M.C. (2001) *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. (Cambridge: Cambridge University Press).
- Oliner, S. P., & Oliner, P. M. (1988) *The Altruistic Personality: Rescuers of Jews In Nazi Germany* (New York: Free Press).
- Overvold, M. C. (1980) 'Self-Interest and the Concept of Self-Sacrifice'. *Canadian Journal of Philosophy* Vol.10 (1):105-118).
- Oxford English Dictionary Online (2013) 'supererogatory, adj. and n.' September 2013. (Oxford University Press). [Accessed on 23 September 2013]
<<http://www.oed.com/view/Entry/194278?redirectedFrom=supererogatory>>.

- Parfit, D. (1984) *Reasons and Persons* (Oxford: Oxford University Press).
- Parfit, D. (2011) *On What Matters* (Oxford: Oxford University Press).
- Peterfreund, S.P. (1975) 'Supererogation and Obligation'. *The Personalist* 56:151-4
- Peterfreund, S. P. (1978) 'On the Relationship Between Supererogation and Basic Duty'. *The Personalist* 59 (1):53-57.
- Portmore, D. (2003) 'Position-Relative Consequentialism, Agent-centered options and Supererogation'. *Ethics* 113(2):303-332.
- Portmore, D. (2008) 'Are Moral Reasons Morally Overriding?' *Ethical Theory and Moral Practice* 11: 369-388.
- Portmore, D. (2011) *Commonsense Consequentialism: Wherein Morality Meets Rationality* (Oxford: Oxford University Press).
- Pybus, E. (1982) 'Saints and Heroes'. *Philosophy* 57 (220):193-199.
- Quinn, P. L. (1978) *Divine Commands and Moral Requirements*. (Oxford: Clarendon Press).
- Rabinowicz, W. (2000). 'Kotarbinski's Early Criticism of Utilitarianism'. *Utilitas* 12 (1):79-80.
- Rawls, J. (1971) *A Theory of Justice* (Cambridge MA: Harvard University Press).
- Raz, J. (1975) 'Permissions and Supererogation.' *American Philosophical Quarterly* 12 (2):161-168.
- Reinhart, A. K. (1983). 'Islamic Law as Islamic Ethics'. *The Journal of Religious Ethics* Vol.11:186-203.

- Richards, D. A. J. (1971) *A Theory of Reasons for Action* (Oxford: Clarendon Press).
- Ridge, M. (2002) 'Mill's Intentions and Motives'. *Utilitas* 14 (1): 54-70.
- Ridge, M. (2006) 'Ecumenical Expressivism: Finessing Frege'. *Ethics* 116 (2): 302-336.
- Ridge, M. (2014) *Impassioned Belief* (Oxford: Oxford University Press).
- Riley, J. (1998) *Mill On Liberty* (New York: Routledge).
- Rosati, C. S. (2009) 'Self-Interest and Self-Sacrifice.' *Proceedings of the Aristotelian Society* (Vol.109 Part 3 pp. 311-325).
- Ross, A.(1941) 'Imperatives and Logic.' *Theoria*, 7: 53–71.
- Ross, W. D. (2003) *The Right and the Good* (Oxford: Oxford University Press)
- Saddock, J. M. (1991) 'On Testing for Conversational Implicature.' In Steven Davis (Ed.) *Pragmatics: A Reader* (Oxford: Oxford University Press).
- Savulescu, J. (2001) 'Procreative Beneficence: Why We Should Select The Best Children.' *Bioethics*, 15(5-6), 413-426.
- Saunders, B. (Forthcoming) 'Is Procreative Beneficence Obligatory?' *Journal of Medical Ethics*.
- Scanlon, T. M. (1975) 'Preference and Urgency'. *Journal of Philosophy* 72 (19):655-669.
- Scanlon, T. M. (1998) *What We Owe To Each Other* (Cambridge MA: Harvard University Press).

Scanlon, T. M. (2008) *Moral Dimensions* (Harvard University Press).

Schleffer, S. (1992) *Human Morality* (New York: Oxford University Press).

Schlothfeldt, S. and Stephen, S. (2012) 'Is Morality Overriding?' in S. Schleidgen (Ed.) *Should We Always Act Morally? Essays on Overridingness* (Marburg: Tectum Verlag) pp.65-88.

Shafer-Landau, R. (2003) *Moral Realism: A Defence* (Oxford: Oxford University Press).

Shiffrin, S. V. (1999). 'Moral Overridingness and Moral Subjectivism.' *Ethics* 109 (4): 772-794.

Sidgwick, H. (1981;1907) *The Methods of Ethics* Seventh Edition (Indianapolis: Hackett Publishing).

Sikora, R. I. (1979) 'Utilitarianism, Supererogation and Future Generations.' *Canadian Journal of Philosophy* (Vol. 9 No.3 pp.461-466).

Skorupski, J. (1999) *Ethical Explorations* (Oxford: Oxford University Press).

Slote, M. (1984). 'Satisficing Consequentialism.' *Proceedings of the Aristotelian Society, Supplementary Volumes*, 58, 139-176.

Slote, M. A. (2007). *The Ethics of Care and Empathy*. (London: Routledge).

Smith, H. M. (2010). 'Subjective Rightness.' *Social Philosophy and Policy*, 27(2), 64–110.

Smith, M. (1994) *The Moral Problem* (Oxford: Blackwell)

- Stanlick, N. A. (1999). 'The Nature and Value of Supererogatory Actions.' *Journal of Social Philosophy* 30 (1):209-222.
- Stevenson, C. (1944) *Ethics and Language* (New Haven: Yale University Press)
- Stocker, Michael (1967). 'Acts, Perfect Duties, and Imperfect Duties'. *The Review of Metaphysics* 20 (3):507 - 517.
- Strandberg, C. (2011). 'The Pragmatics of Moral Motivation'. *Journal of Ethics* 15 (4):341-369.
- Strandberg, C. (2012) 'A Dual Aspect Account of Moral Language.' *Philosophy and Phenomenological Research* 84: 87–122.
- Straumanis, J. (1984). 'Duties to Oneself: An Ethical Basis for Self-Liberation?' *Journal of Social Philosophy* 15 (2):1-13.
- Strawson, P. F. (1968). 'Freedom and Resentment,' In *Studies in the Philosophy of Thought and Action*. (London: Oxford University Press).
- Stroud, S. (1998) 'Moral Overridingness and Moral Theory'. *Pacific Philosophical Quarterly* 79 (2):170-189.
- Svavarsdóttir, S. (1999) 'Moral Cognitivism and Motivation.' *Philosophical Review* 108 (29):161-219.
- Tanyi, A. (2012) 'The Case for Authority', in S. Schleidgen (Ed.) *Should We Always Act Morally? Essays on Overridingness* (Marburg: Tectum Verlag):159-189.
- Terrell, H. (1969) 'Are Moral Considerations Morally Overriding?' *Australasian Journal of Philosophy* 47:51-60.

- Timmerman (2005) ‘Good but Not Required? – Assessing the Demands of Kantian Ethics.’ *Journal of Moral Philosophy*, 2: 9–27.
- Tresan, J. (2006) ‘De Dicto Internalist Cognitivism’. *Noûs* 40 (1):143-165.
- Trianosky, G. W. (1986). ‘Supererogation, Wrongdoing and Vice: On the Autonomy of Virtue’. *Journal of Philosophy* 83(1).
- Turner, P. N. (2014). “‘Harm’ and Mill’s Harm Principle.’ *Ethics*, 124(2), 299-326.
- Twain, M. (1966) *The Adventures of Huckleberry Finn* (London: Penguin)
- Urmson, J. O. (1958) 'Saints and Heroes' Reprinted in *Moral Concepts* Joel Feinberg (ed.) (1969) (Oxford: Oxford University Press)
- Usborne, S. (2013) ‘Happy to Help: Meet the Free Help Guy.’ *The Independent* (20/11/2013) Available at: <http://www.independent.co.uk/life-style/happy-to-help-meet-the-free-help-guy-8909408.html> [Accessed on 20/11/2013].
- van Benthem, J. (1982). ‘Later than late: On The Logical Origin Of The Temporal Order’. *Pacific Philosophical Quarterly* 63(2), 193-203.
- von Fintel, Kai and Iatridou, Sabine (2008) ‘How to Say *Ought* In Foreign: The Composition of Weak Necessity Modals.’ In J. Guéron and J. Lecarme (eds.) *Time and Modality* (New York: Springer).
- van Roojen, M. (2010) ‘Moral Rationalism and Rationalist Amoralism’. *Ethics* 120 (3): 495-525).

- von Wright, G. H. (1963) *The Logic of Preference* (Edinburgh: Edinburgh University Press).
- Vessel, J.P (2010) 'Supererogation for Utilitarianism.' *American Philosophical Quarterly*, 47: 299–317.
- Wallace, R. J. (2006) 'Moral Motivation', in *Contemporary Debates in Moral Theory* J. Dreier (Ed.) (pp. 182-195) (Oxford: Blackwell).
- Wedgwood, R. (2007) *The Nature of Normativity* (Oxford: Oxford University Press)
- Weinberg, J. (2011) 'Is Government Supererogation Possible?' *Pacific Philosophical Quarterly* 92 (2): 263-281.
- Wiesenthal, S. (1989) *Justice Not Vengeance* (London: Mandarin).
- Williams, B. (1973a) *Problems Of The Self: Philosophical Papers 1956-1972* (Cambridge: Cambridge University Press).
- Williams, B. (1973b) 'A Critique of Utilitarianism', in J.J. C. Smart, and B. Williams. *Utilitarianism; For and Against*. (Cambridge: Cambridge University Press).
- Williams, B. (1981) *Moral Luck* (Cambridge: Cambridge University Press).
- Williams, B. (1985) *Ethics and the Limits of Philosophy* (London: Fontana).
- Williams, B. (1995) *Making Sense of Humanity and Other Philosophical Papers 1982-1993* (Cambridge: Cambridge University Press).
- Wolf, S. (1982) 'Moral Saints'. *Journal of Philosophy* 79 (8):419-439.

Zangwill, N. (2008) 'The Indifference Argument'. *Philosophical Studies* 138 (1):91-124.

Zimmerman, M. (1996) *The Concept of Moral Obligation* (Cambridge: Cambridge University Press).

Zimmerman, M. (2008) *Living With Uncertainty: The Moral Significance of Ignorance* (Cambridge: Cambridge University Press).