

Judgement Internalism and Supererogation

B000250

Taught Msc in Philosophy

The University of Edinburgh

2011

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 indicated by proper use of quotes and references.

Abstract: A curious feature of the literature on motivational judgement internalism is the absence of a discussion of which moral judgements are expected to motivate and how. This dissertation aims to address this issue by investigating what account an internalist can give of judgements of supererogation. This investigation will proceed in three stages. First I will investigate the difference between judging that something is a moral obligation and judging that it is supererogatory. I will argue that, unlike judgements of obligation, there is no reliable connection between judgements of supererogation and motivation. Next I will look at what account a judgement internalist can give of the necessary connection between moral judgements and motivation that is compatible with existence of moral judgements that we do not expect people to be motivated by. I will argue that an all things considered reasons internalism is able to give such an account but cannot explain the motivational power of judgements of supererogation. I will then argue that positing an internal connection between judgements of supererogation and feelings of admiration allows for an internalist account that is capable of explaining the ability of judgements of supererogation to motivate.

Word Count:14938

Chapter 1 – Is there a reliable connection between moral judgements and motivation?

Motivational judgement internalism is a theory about the nature of moral judgements that claims that there is a “necessary” connection between moral judgements and motivation¹ (Darwall 1983 p.54). In other words, part of the nature of moral judgements is that they are motivating. Motivational externalism on the other hand is the view that any connection that exists between moral judgements and motivation is contingent and dependent on the dispositions of the person making the judgement. Internalists claim that their view is more appealing because only internalism is able to provide a plausible account of “the reliable connection” between moral judgements and motivation (Smith 1994 p.76). However, these claims lack an account of which moral judgements are being referred to. The aim of this chapter will be to investigate whether this claim can plausibly be made about two different types of moral judgement. The first type of moral judgement I will look at is a judgement that an act, or potential act, is morally obligatory. The second type of moral judgement I will look at is a judgement that an act is supererogatory. I will start by outlining why internalists argue that this claim adds to the appeal of their view. I will then investigate whether it is plausible to make this claim about judgements of obligation and judgements of supererogation. I will argue that this claim can plausibly be made about judgements of moral obligation but not judgements of supererogation.

First we should consider why internalists claim to be better placed to explain the reliable connection that they claim exists between moral judgements and motivation. Smith argues that the motivation of strong willed people tracks their moral judgements (1994 p.71). In other words, moral judgements are connected to motivation in such a way that we expect strong willed people to generally become motivated to do what they believe is right and to lose motivation for what they believe to be wrong. For example, if a strong willed meat eater becomes convinced that eating meat is wrong then we would expect that person to

¹ Darwall distinguishes this from what he calls “existence internalism” which holds that in order for an act to be rational or right it must be capable of motivating (1983 p.54).

stop eating meat. If someone made this judgement and continued to eat meat then we would not describe them as strong willed.

An acceptable account of moral judgements must be able to explain this connection. The connection can be explained as being internal to the moral judgement or as being external to it (Smith 1994 p.72). Smith argues that an explanation that holds that the connection is external to the judgement is committed to an implausible view of moral motivation (1994 p.74). Such a view would have to posit the existence of a motivation that is explicitly moral such as the desire “to do the right thing where this is read de dicto and not de re” (1994 p.74). In other words, externalist explanations of the connection between moral judgement and motivation will need to invoke a further desire to do the right thing, whatever that happens to be. Smith argues that this explanation depends upon the existence of a step in the mental process of the moral agent that does not exist. We expect people to be motivated by the features of an act that make it right not by its 'rightness'. Smith argues that an agent with this extra thought would be regarded as a “moral fetishist” (1994 p.74)². The internalist explanation is preferable as it does not make a fetish of morality.

Let's accept for now that this is a genuine advantage that counts in favour of the internalist position if their claim about the reliable connection between judgement and motivation is accepted. The question we must now ask is: Of the moral judgements that we make, which, if any, does this apply to? I will start by investigating whether they apply to judgements of moral obligation before going on to look at how they apply to judgements of supererogation. I will do this by looking at what our expectations would be for someone who made these judgements.

Let's start by looking at moral obligation. There are many competing accounts of moral obligation and it is outwith the scope of this paper to adjudicate between them. I intend instead to outline a view of moral obligation that I find attractive and explain briefly why I

²For an argument to support an externalist explanation for this connection see Sigrun Svavarsdóttir (1999)

find it attractive.

A plausible conception of moral obligation can be found in the work of John Stuart Mill. Mill argues that, "Duty is a thing which may be exacted from a person as one exacts a debt." (2004 Chapter 5). On Mill's account, if an act is morally obligatory then we can demand that people perform that act. Darwall builds on Mill's account of moral obligation by introducing the idea of "the second person standpoint." (2006 p.1) To evaluate something from this standpoint is to evaluate from the perspective we take when we make and accept claims on each others conduct (2006 p.1). Darwall's account of obligation is linked to accountability. When we evaluate from the second person standpoint we think about the attitudes we would hold towards an agent and their behaviour if we were confronted by it. To hold someone morally obliged is to address them in a way that gives them reasons for acting. Someone has a moral obligation to act in a certain way if that way of acting could be legitimately demanded of them (Darwall 2006 p.27). If they do not act in this way then they are the legitimate target of punishment. The authority to make such an address comes from shared membership of a moral community (2006 p.7). As members of a moral community we are accountable to one another and have the authority to make demands of each other's conduct in relation to others. If someone is thought to have a moral obligation to act in a certain way then they are judged to have most reason to act in this way. As Darwall points out, "it makes no sense to blame someone for doing something and then add that he had, nonetheless, sufficient reason to do it, all things considered. (2006 p.28). Someone who is able to show that they did not have most reason to act in the way that is being demanded of them will have shown that they had no moral obligation to act in that way.

I do not have space here to engage in a thorough defence of this account and intend only to show why I find it appealing. As Watson points out, one of the attractive features of Darwall's account of moral obligation is that it offers a response to Anscombe's challenge (2007 p.37). This challenge asks where the authority of moral demands come from in the absence of a divine law giver (Anscombe 1958 p.2). Darwall's account of moral obligation

is able to meet this challenge by claiming that the authority to make moral demands comes from members of a shared moral community holding each other to account (Watson 2007 p.38). Of course much more can be said about whether this theory will stand up to intense scrutiny but that is the task for another project³.

Let's think now about what it means for someone to judge that they have a moral obligation and how we would expect someone who made a judgement of this sort to act. If someone judges that they have a moral obligation to act in a certain way then they are judging that other members of the moral community can legitimately demand that they act in a particular way. Someone who judges that he does not have most reason to act in this way does not judge that he is under obligation to act in that way. We expect strong willed people who judge that they have most reason to act in a particular way to be motivated to do so. Someone who judges that they have an obligation but is unmotivated by it will not be thought to be strong willed. The reason for this is that we expect people to be motivated by their judgements about what there is most reason to do. We expect this because there is a reliable connection between judging that you have most reason to act in a particular way and being motivated to act in that way. The internalist claim about the reliable connection between moral judgements and motivation is convincing when we consider judgements of moral obligation.

This is only the case if we assume that moral obligations are always what we have most reason to do. In his book *Brute Rationality*, Joshua Gert argues that we should not accept this. Gert distinguishes between reasons that constitute requirements of rationality and those which provide rational justification (2004 p.23). Justifying reasons can explain action and give an account of why it was reasonable for an agent to have acted in a particular way (2004 p.24). Rational requirements, on the other hand, must be acted upon in order for an agent to maintain rationality (Gert 2004 p.23). Gert argues that the two should be seen separately. Justifying reasons can explain action, can give an account of why an agent acts in a certain way but they need not lead to requirements (2004 p.24). Gert then goes on to

³ For a discussion of Darwall's account of moral obligation see *Ethics* Vol.118 No.1 Symposium on Darwall's Second Person Standpoint.

claim that moral reasons only ever give justifying reasons and never provide rational requirements. In support of his claim Gert gives an example of someone who is deciding whether to spend their money on a bottle of wine or give it to charity (2004 p.26). Gert argues that while this person would be justified in giving their money to charity we would not call them irrational if they spent the money on the wine. If we accept Gert's point then we will no longer have reason to think that judgements of moral obligation are what we have most reason to do.

Gert makes an interesting point in distinguishing between requirements of rationality and justifying reasons. However, Gert is wrong to claim that moral reasons can only ever provide justifying reasons. Returning to Darwall's account of moral obligation will allow us to see why this is the case. Darwall understood moral obligations as the legitimate demands that members of the moral community can address to one another. The ability to make demands of one another comes from a "second personal competence" (2006 p.29). This is the ability to make and accept legitimate demands to and from other members of the moral community. In order to make legitimate demands of people you must be a member of a shared moral community. If this is the case then the demands will also apply to you. To judge that you have a moral obligation is to demand a certain course of action from yourself. It certainly seems reasonable to claim that when someone demands of herself that they act in a certain way, they are rationally required to act in that way. Of course we might think that there are no moral obligations. Nevertheless, if people judge that something is obligatory then this does seem to provide a rational requirement. If we accept this interpretation of moral obligation then we will expect people to be motivated to act in line with what they judge to be a moral obligation. We should therefore accept the internalist's claim about the reliable connection between moral judgements and motivation when applied to judgements of obligation.

Gert would reject this account of the link between morality and rational requirements because accepting such a link means that people who act immorally can no longer be held to be morally responsible (2004 p.82). If we claim that people who act immorally act

irrationally and that rationality is a requirement for moral responsibility then we are forced into the unattractive position of being unable to hold anyone responsible for their immoral actions (2004 p.82). Gert argues that this gives us reason to reject the claim that moral reasons provide rational requirements (2004 p.83)

However, I think this criticism rests on a simplistic account of moral responsibility. As Darwall argues, making a moral demand of someone involves an assumption that the person being addressed is competent of understand and responding to such demands (2006 p.75). If someone is capable of understanding and responding to moral reasons then it is appropriate to make moral demands of them. This involves more than simply accepting that a demand has been made of them and responding to it. Darwall argues that in order to be held morally responsible it is important that the person being addressed possesses “second personal competence” (2006 p.75). This means that they must be capable of recognising the legitimacy of such demands and holding herself and others to these types of demand (2006 p.78). This is what moral responsibility consists of from Darwall's view. If someone is capable of recognising the legitimacy of moral demands and responding to them then they are morally responsible. The fact that such a person may on occasion fail to act in line with the moral demands that are made of them does not remove them of this responsibility. There is no reason to think then that judging that someone has acted irrationally makes us unable to hold them morally responsible for their actions. Only if someone is incapable of recognising and responding to moral demands would we no longer think them to be fitting subjects of blame.

We can clarify this discussion by distinguishng between irrational and arational agents. An irrational agent is one who is capable of acting in line with reasons but happens not to on a particular occasion. An arational agent on the other hand is one that is incapable of recognising and responding to reason. Darwall's account of moral responsibility allows us to hold irrational agents responsible but not arational agents. This means that this account is capable of holding immoral agents morally responsible, so long as they are capable of responding to moral demands. We should therefore accept the internalist's claim about the

reliable connection between moral judgements and motivation when applied to judgements of obligation as Gert's objections give us no reason to think that moral reasons are incapable of providing rational requirements.

Let's look now at judgements of supererogation. In 2007 a man collapsed onto the tracks at a Subway station in New York as a train was approaching. Seconds later Wesley Autrey jumped onto the tracks and held the man down as the train came to a halt above their heads (Carwell 2007). Risking your own life to save that of another is an act that deserves high moral praise. What is interesting about Autrey's case is that it is not only morally good but it also seems reasonable to think that acting in the way that Autrey did is not morally required. After all, it seems almost inconceivable that anyone would have criticised Autrey had he not acted in this way. Furthermore, the fact that this act is not one we would expect most people to perform partly explains the level of praise received by Autrey. Acts like Autrey's that are morally good without being morally required are called supererogatory acts.

The term 'supererogation' originates from the idea of giving more than is asked (Heyd 1982 p.1). J. O. Urmson opened the contemporary discussion of the concept. He argued that the traditional view of moral action which classifies actions as either moral duties, morally indifferent or forbidden by morality is insufficient (1958 p.60). Urmson argued that this account leaves no room for acts that are good but not moral duties such as the actions of saints and heroes (1958 p.61). Such actions go beyond what is demanded by our moral duties (Urmson 1958 p.65). Because supererogatory acts are not obligatory we do not expect people to perform these actions nor do we blame or punish those who fail to perform them (Mellema 1991 p.5).

Heyd argues that "continuity" is an important aspect of the relationship between duty and supererogation (1982 p.5). In order for an act to be considered supererogatory Heyd argues that the act must have the same sort of value as that of the moral obligation that it goes

beyond (1982 p.5). The performance of a supererogatory act has the same type of value as the performance of the related moral obligation but to a greater extent (Heyd 1982 p.5). Mellema argues that this means that overall value is increased to a greater degree when an act does not fulfil a moral obligation (1991 p.28). The continuity requirement allows us to see how supererogation and obligation relate to each other. Fulfilling a moral obligation and performing an act that goes beyond that obligation are both acts which have the same sort of value. However, the supererogatory act realises a greater amount of that value than the obligatory act.

Our discussion of supererogation up to now has focussed entirely on actions so let's now look at what it is to judge that something is supererogatory. To make a judgement of this sort is to judge that an act is one that goes beyond what is morally obligatory, that the person performing the act is a fitting subject for moral praise and that those who pass on the opportunity to act in such a way do not deserve to be the subject of blame or criticism. As a result, making a supererogatory judgement does not bring about any expectation of action⁴. If someone tells me that they think giving money to charity is supererogatory then I would not be surprised to find out that this person gives money to charity nor would I be surprised to find out that they don't.

It would be a significant drawback of the account of obligation and supererogation that I have given here if it was unable to accommodate important features of the relationship between these two kinds of action. Kamm argues that one of the features of this relationship is that it can sometimes be permissible to forego what is morally obligatory in favour of an act of supererogation (2001 p.313). Kamm backs this up with the following example. Suppose I have made an arrangement to meet a friend for lunch. We might think that this means that I have a moral obligation to do so. However, on the way to lunch I pass a car crash and volunteer to give up my kidney in order to save the life of one of the

⁴ At least not in the abstract. We might think that if we know the person making the judgement then we might be able to form a reasonable expectation. For example, if we know that the person is a moral saint then we might expect this person to be motivated by judgements of supererogation. Whereas, if the judger is someone who only ever does what they judge to be the minimum requirements of morality then we would expect this judgement not to motivate.

victims of the crash causing me to miss lunch. This seems like it is an act of supererogation. It is not only a morally permissible act but one deserving high moral praise and intuitively I had no obligation to donate my kidney (2001 p.314). It seems morally acceptable to choose this act over my obligation to keep my lunch date. Kamm concludes that it may sometimes be morally permissible or even praiseworthy to choose an act of supererogation over an obligation (2001 p.314). This causes a problem for the account that I have given of moral obligation and supererogation. If it is permissible to choose an act of supererogation over an obligation then it seems that it is rational to do so. If we accept this then we must reject my account of moral obligation which claimed that it is irrational to fail to be motivated to do what you judge you have a moral obligation to do.

However, Kamm's argument does not work against the way that I have understood moral obligation. Earlier in the chapter I said that if someone shows that they did not have most reason to act a certain way then we cannot call that way of acting morally obligatory. Kamm is right to say that there are occasions in which we can choose to do something supererogatory instead of a moral obligation. However, in choosing to do the supererogatory act the alternative stops being obligatory. To apply this to Kamm's example, once the decision has been made to donate the kidney there is no longer an obligation to meet the friend for lunch. No one would be blamed for doing this and no one could legitimately demand that their friend choose to keep their lunch date rather than save a life.

This raises an interesting question as to what it means to say that something is obligatory if there remains an option to act differently. A tempting way to respond to this would be to say that moral obligations are disjunctive. Perhaps there will always be a range of ways in which we can fulfil our obligations. A plausible way of thinking about this is that to have a moral obligation to perform an act is to be required to perform that act or one which brings about greater value. If someone has an obligation to keep their lunch date then this should really be seen as an obligation to keep the lunch date or do something which brings about greater value. By donating the kidney we fulfil our obligation to either keep the lunch date

or do something that brings about more value. This seems like a plausible thing to say for an account of obligation that is linked to accountability. If we choose to perform an act of greater value than an act we have an obligation to perform then it seems reasonable to think that no one would think us blameworthy for this. We can now see why it is permissible to miss the lunch date in order to donate a kidney. Donating a kidney is an act of far greater value than attending a lunch date and so this act fulfils the obligation to either keep the lunch date or perform an act that brings about greater value.

If donating the kidney in the above example can be thought of as a way of fulfilling a moral obligation then we might think that it can no longer be described as an act of supererogation. However, I think that even though this act fulfils an obligation we can still think of this as a supererogatory act. If we have a range of ways in which we can fulfil the obligation to keep the lunch date then those that go beyond the minimum level of sacrifice and also bring more than the minimum level of value can be thought of supererogatory as they go beyond the basic requirements of moral obligation. This does not mean that all acts of supererogation fulfil an obligation rather that this can be seen as one type of supererogatory act. We might perform an act of supererogation when we had no moral obligations to act a certain way.

Given what we have said so far about supererogatory acts this account appears problematic. We said earlier that an act of supererogation has greater value than the corresponding moral obligation. If we accept this and we accept that moral obligations always allow us to perform an act of more value than what we are obliged to do. Suppose Jane comes home to find her house on fire with her son and his friend trapped inside. Suppose that it is only possible for Jane to rescue one and attempting to do so will put her own life at significant risk. We might think that parents have special obligations of care to their offspring and because of this Jane has a moral obligation to save her son. Let's also accept that Jane has no obligation to risk her life in order to save her son's friend and that to do so would be supererogatory. If we accept the above account of supererogation then we are forced to say that in this situation it is preferable for Jane to save her son's friend

rather than her son as this is a supererogatory act and therefore of more value than the moral obligation.

We can respond to this problem by distinguishing between agent-relative and agent-neutral value. We can think of this distinction in terms of what Schroeder calls “better than relations” (2007 p.268). This is a way of expressing preferences for one state of affairs over another. A value is agent-relative if for someone with a particular involvement in a situation, what is better for them differs to what is better for an impartial observer (Suikkanen 2009 p.6). If we accept that the relevant value in a disjunctive obligation may be agent-relative then the next issue we must face is deciding which agent is the value relative to. A natural response to this would be to say that it is the person performing the act whose preferences are taken into account. However, this is problematic for the following reason. Suppose I have a moral obligation to visit a friend in hospital. I think about doing this but decide that going out for dinner with healthy friends would bring about more that I find valuable than going to the hospital. It would be an odd account of moral obligation that accepted that this is a satisfactory way of fulfilling my obligation. A more plausible account would be to take the preferences of the person who is owed the obligation into account. On this account if I have a moral obligation towards a certain person then I must either perform the act I am obligated to do or perform an act which is held to be more valuable by that person. This allows us to explain why it is not permissible for Jane to save the friend rather than the son as for the son a world in which he is saved but his friend is not is better than a world in which only his friend is saved.

Another objection that could be made is that it is a mistake to think that we need to accommodate supererogatory judgements. Given that we can deny the existence of acts of supererogation perhaps we can also deny the existence of judgements of supererogation. Urmson argued that supererogation caused problems for traditional ethical theories as they could not accommodate the concept. However, it could be argued that the fact that a normative ethical theory is incompatible with supererogation does not in itself show that we should reject the moral theory. We might think that this gives us reason to abandon the

concept of supererogation instead. As Zimmerman points out, the fact the most significant moral theories seem to have difficulty accommodating supererogation might give us reason to think that supererogation is impossible (1996 p.237). Perhaps then we should treat the problems that ethical theories have had explaining supererogation not as evidence against those theories but as reason to abandon the concept. We might also think that if a theory about the nature of moral judgements was found to be incompatible with judgements of supererogation then this gives us reason to think such judgements don't exist⁵.

However, such an argument is unlikely to convince anybody. While it seems reasonable to think that a plausible argument could be made against the existence of supererogatory acts it does not seem plausible to use such an argument to deny the existence of judgements of supererogation⁶. That such judgements exist strikes me as an intuition that cannot be denied. As Attfield argues, acts of supererogation are established features of common sense morality in most societies (Attfield 1995 p.116). Heroes and saints receive special recognition in the form of medals, ceremonies and public adoration because they are judged to have gone beyond what is morally obligatory in a particularly impressive way. This recognition goes beyond the acknowledgement accorded to those who are judged to have acted in line with moral duty throughout their lives. Such people may receive recognition but this recognition is likely to be muted in comparison and typically will be accorded only upon death or retirement. This special recognition given to those who are judged to have done more than is required of them is firm evidence that make judgements of supererogation. Someone might deny that they make such judgements but how could they convince anyone that no one else does? In the same way, we could deny the existence of moral facts, as error theorists do⁷, but any argument against the existence of moral judgements is unlikely to be taken seriously. We can debate whether or not these

⁵ Elizabeth M. Pybus argues for a related position (1981). She argues from the fact that moral commendation involves a desire to emulate that acts of supererogation do not exist. While I think Pybus was right to point to an incompatibility between judgement internalism (without actually using this term) and supererogation, I think that given the debate that exists between internalism and externalism we should not assume that if such an incompatibility exists that this shows that supererogatory acts do not exist.

⁶ One way in which it might be denied is by appealing to a global eliminative materialist theory that holds that beliefs and desires do not exist. I do not have space to explore such a response here. Arguments against this theory can be found in Kitcher (1984) and Stich (1996)

⁷ For example, Richard Joyce (2001)

judgements are correct and whether acts of supererogation are possible. What seems impossible to deny is that these judgements exist.

Let's briefly sum up what has been said so far. Judgement internalists claim that there is a necessary connection between moral judgements and motivation. Internalists have claimed that their theory is preferable to externalism as it is able to explain the fact that a change in motivation generally follows from a change in moral judgement. In this chapter I have investigated which positive moral judgements we would expect people to be motivated by. I have argued that there is a reliable connection between judgements of obligation and motivation. However, the same cannot be said for judgements of supererogation. There is no reliable connection between judging that an act is supererogatory and being motivated to do it.

Accepting this poses a challenge to judgement internalists. As it stands the claim that there is a necessary connection between moral judgements and motivation is under threat. The fact that judgements of supererogation will not always result in motivation suggests that there is no necessary connection between the two. The internalist must provide an account of the necessary connection between moral judgements and motivation that is supported by the reliable connection between judgements of obligation and motivation but is also able to account for the lack of this connection for judgements of supererogation. The next chapter of this paper will look at how an internalist could meet this challenge.

Chapter 2 – Can internalism accommodate judgements of supererogation?

In the last chapter I argued that a reliable connection between motivation and judgements of moral obligation exists but that there is no such connection for judgements of supererogation. This creates a challenge for internalists. They must account for the necessary connection between moral judgements and motivation in a way that is compatible with the existence of a reliable connection between some moral judgements but not others. In this chapter I will investigate what kind of connection is capable of achieving this. In answering this question we will examine whether the connection is unbreakable, what level of motivation comes from making a moral judgement and whether the connection is between any moral judgement or an 'all things considered' normative judgement to which a particular moral judgement may play a contributory role. In this chapter I will examine these issues and then propose an alternative necessary connection between judgements of supererogation and feelings of admiration. I will defend this connection against various ways in which it could be argued that these feelings are not always compatible with such judgements.

Let's deal with the first issue. Internalists have said that there is a necessary connection between moral judgement and motivation. This could be interpreted to mean that whenever someone makes a moral judgement they will always be motivated by it. However, as Stocker has points out, people suffering from depression can find themselves completely lacking in motivation to do what they think is right (1978 p.744). An internalist that is committed to an unbreakable connection between moral judgement and motivation will be forced to say that people in this situation are not making sincere moral judgements. This seems an implausible way of characterising people with depression. As Stocker makes clear, when someone is suffering from depression they are not prevented from classing things as right or wrong in the way they did previously. Their problem is in transforming these judgements into action (1978 p.744). As a result it seems reasonable to think that proposing a connection that claims that everyone who makes a moral judgement will

always be motivated to act in line with that judgement is too strong.

Weaker⁸ versions of internalism have sought to propose a necessary connection between moral judgements and motivation that is able to accommodate cases where the connection breaks down. These cases are typically dealt with by a rationality clause⁹. Moral judgements are said to motivate necessarily when the agent is practically rational. This approach passes the explanatory work onto the concept of practical rationality and so raises the question of what is meant by this. We can distinguish between two forms of rationality: practical rationality, concerned with what we should do and theoretical rationality, concerned with what we should believe (Holton 2009 p.175). Practical rationality governs our intentions or plans about what to do (Harman 1999 p.13). To be practically rational is to have intentions that are compatible with the agent's desires and evaluative judgements (Wallace 2009). The rationality clause advocated by internalists holds that to fail to transform moral judgements into a motivation to act involves a failure of practical rationality. One form of practical irrationality involves possessing intentions which are not coherent with one's desires and value judgements. Another form of practical irrationality is failing to intend to do what you judge to be the best means to your ends. By invoking a rationality clause this form of internalism is able to accommodate the cases where people fail to be motivated by their moral judgements while maintaining the existence of a necessary connection. On this view, being motivated by your moral judgements is a necessary condition for practical rationality.

The next issue to investigate about the internalist thesis concerns the level of motivation being proposed. Internalists claim that moral judgements motivate. We can understand this to mean that moral judgements will motivate all the way to action. We can also understand it to mean that someone who makes a moral judgement will be motivated but not necessarily motivated enough to act in line with the judgement¹⁰. Suppose Dave has just

⁸David Brink (1989 p.40-41) was the first to make the distinction between 'strong' internalism and 'weak' internalism.

⁹Wedgewood (2007 p.25) and Smith (1994 p. 61) deal with the problem in this way.

¹⁰Mason characterises this distinction as being between "Weak internalism" and "Weakest internalism" (2008 p.143-144).

enough money to pay his rent. Dave could sincerely judge that it would be good to give his money to Oxfam to help the hungry and also judge that he should pay his rent. Someone in this situation may be motivated to give money to Oxfam but find that this motivation is greatly outweighed by his motivation to pay his rent to avoid all the unpleasant consequences that might occur should he fail to do so. The necessary connection is maintained as the agent in this case is motivated to act in line with his judgement, its just that this motivation is not strong enough to lead to action. We might think that this form of internalism offers a way of meeting the challenge of supererogation. When someone judges that a possible way of acting is supererogatory they may have some motivation to act in line with this judgement however this motivation may well be overruled by other reasons that recommend against acting in such a way. In other words, judging that something is supererogatory will lead to some motivation to act in line with this judgement, however small that motivation may be. This form of internalism allows us to explain why we would not expect the connection between moral judgements and motivation to hold for supererogatory judgements.

To evaluate whether this form of internalism is plausible we must distinguish between two levels of reason and how these two types of reason relate to motivation. Dancy distinguishes between all things considered reasons¹¹ and “contributory reasons”¹²(2004 p.17). A contributory reason for action is one that counts in favour of acting a certain way while an all things considered reason concerns what we should do after we have weighed up all the relevant reasons. The question that this raises for internalism is whether the theory applies to judgements concerning contributory reasons, all things considered reasons or both. If the theory applies to contributory reasons then every reason that an agent judges to count in favour of action will be expected to motivate to some extent. The above solution to the challenge supererogation poses to internalism is a contributory reasons internalism. An all things considered reasons internalism would only apply to

¹¹ Dancy uses the term 'overall ought'. However, In a forthcoming paper Ridge and Chrisman (in progress) raise a number of concerns about the use of the word 'ought' when what is meant is 'must'. In light of these worries it seems that Dancy's use of 'ought' may be misplaced. In order to remain neutral on this topic I refer to 'all things considered reasons' rather than 'overall oughts'.

¹² As Dancy notes the term “pro tanto” (meaning at “as far as that goes”) is often used in relation to this concept (2004 p.17). I will follow Dancy in referring to this concept in terms of 'contributory' rather than 'pro tanto' reasons.

judgements that agents make about the way they ought to act once everything has been taken into account¹³. We can make this distinction clearer by looking at how it applies to the example of Dave. Dave judges that he does have good reason to give to Oxfam as to do so may well result in a life being saved. This is a contributory reason. However, he judges that when he takes all the reasons he has into account he ought to pay his rent. The question we must ask of internalism is whether a complete lack of motivation on Dave's part to give money to Oxfam would show that Dave is practically irrational.

I see no reason to think that someone in the above situation who lacks any desire to give money to Oxfam is practically irrational. As Dancy argues, someone who has no motivation to act in line with a contributory reason that is greatly outweighed by other reasons does not seem to be guilty of any failure in practical rationality (2004 p.22). It seems odd indeed to think that lacking any motivational pull towards giving money to Oxfam would display a lack of rationality. As a result, the necessary connection between moral judgements and motivation should be restricted to judgements about all things considered reasons. The necessary connection can be formulated in the following way: If someone judges that an act, ϕ , is what they have most reason to do then they will either be motivated to ϕ or be practically irrational.

Restricting the necessary connection to judgements about all things considered reasons allows the internalist to explain why there is a reliable connection between judgements of moral obligation and motivation. In the last chapter we said that if someone judges that they have a moral obligation to perform a particular act then they are judging that this act is what they have most reason to do. This explains why there is a reliable connection between judgements of moral obligation and motivation.

This form of internalism is also able to explain why there is no reliable connection between judgements of supererogation and motivation. Judging that an act is supererogatory

¹³ Wedgewood (2007) defends an internalism about "all things considered" .

involves judging that there is a moral reason to perform that act. Knowing that someone has judged that an act is supererogatory will not by itself give us any indication of whether that person judges that this way of acting is the way they have most reason to act once everything has been taken into account. This will depend upon the costs and benefits involved in performing this act over another and the level of importance the agent places in performing morally praiseworthy acts. Someone could judge that an act is supererogatory and judge that it is what they have most reason to do. Equally people can judge that something is supererogatory and judge that it is not what they have most reason to do. This form of internalism is well placed to explain the lack of a reliable connection between judgements of supererogation and motivation. If there is a necessary connection between judgements about what there is most reason to do and motivation then a group of judgements that will sometimes be what there is most reason to do and not at other times will not be reliably connected to motivation. This form of internalism is able to account for the existence of a reliable connection between judgements of moral obligation and motivation and the lack of such a connection for judgements of supererogation.

As it stands this account gives us an unsatisfying explanation of how a judgement of supererogation may motivate someone to action. So far we have said that if someone judges that acting a certain way is what they have most reason to do then this will motivate them to action. This allows us to say the following about supererogatory judgements: if someone judges that a possible act is supererogatory and what they have most reason to do then this judgement should motivate. As it stands, the judgement of supererogation adds nothing to our understanding of why someone would choose to act this way. Unless we can explain how judgements of supererogation are capable of influencing our judgement of what there is most reason to do then we will lack an account of how these judgements could lead to action. We might think that one way to explain this is to argue that because judging something is supererogatory involves a judgement that that way of acting is morally valuable this gives us reason to act in this way. This would explain how judgements of supererogation can alter what we judge to be the balance of reasons. However, this account assumes the truth of internalism the internalist claim that judging an act to be morally valuable involves judging that there is reason to perform it. Given that

this paper is concerned with evaluating internalism about different sorts of moral judgement in the hope of shedding new light on the debate between internalists and externalists we cannot defend internalism for judgements of supererogation by appealing to internalism. Another approach would be to posit a necessary connection between judgements of supererogation and a pro-attitude that falls short of motivation. This would allow for an internal but defeasible connection between judgements of supererogation and motivation. Internalists need to give such an account as the alternatives are either an explanation that is external to the judgement or one that gives no role to the judgement. To give such an account we must first find a pro-attitude that is able to play this role.

In order to consider what the most plausible candidate for this attitude could be we should consider the attitude we would usually expect people to have if they have made this type of judgement. The obvious candidate for this attitude is 'admiration'¹⁴. Jollimore argues that admiration is a pro-attitude we have towards people who we feel have achieved something that is “both worthy and difficult” (2006 p.159). This makes an important point about the concept of admiration. We only feel admiration towards people if they have acted in a way that we find valuable. A military general may admire the ingenuity of a scientist who creates a new weapon but a pacifist will not. In addition we will only admire people if their achievements are not easy to obtain. This may be because it takes hard work or particular skills to achieve what they did (Jollimore 2006 p.160).

When people describe someone as a saint or a hero we would expect them to feel admiration towards that person. It would seem strange for someone to say that an action is beyond the call of duty and feel no admiration toward people who would perform the act. We would question the sincerity of the judgement in such a case. We can formulate the necessary connection between judgements of supererogation and admiration in the following way: If someone judges that an act, ϕ , is supererogatory then they will either feel admiration towards those who ϕ or they are irrational¹⁵. We can explain this connection by

¹⁴ Robert Audi also links supererogation with admiration. He claims that supererogatory ideals are “admirable to fulfil even though we are not criticizable for not doing so.” (2005 p.135)

¹⁵ This assumes that emotions can be assessed in terms of rationality. For an argument to support such a view see Provis (1981).

saying that these feelings are part of the content of supererogatory judgements. Part of what it is to sincerely judge that someone has acted in a supererogatory way is to feel admiration for that person's moral commitment. For this connection to be plausible it must be the case that feelings of admiration are always compatible with judgements of supererogation. In the remainder of this chapter I will respond to some problems that might be raised against this compatibility. In order for this connection to enable internalism to explain how judgements of supererogation can motivate an explanation is needed of how feelings of admiration are capable of motivating. This account will be given in the next chapter.

One problem that could be raised against this account is that it doesn't fit with the testimony of people who perform acts of supererogation. The reaction of Autrey and two other men who have performed similar heroic acts on the New York subway in recent years supports this claim. As the New York Times reported in 2010, Autrey, Chad Lindsay and an unidentified man all decided to walk straight out of the subway station after jumping on the tracks to save a life (Daly 2010). That all of these men decided to leave promptly after their acts suggests that they did not consider themselves worthy of admiration. If this had been the reason for their actions then it is reasonable to think that they would have stayed to receive the acclaim that they would have received. The fact that those who perform acts of supererogation often seem not to be interested in admiration might be thought to be evidence against the claim that there is a necessary connection between judgements of supererogation and admiration.

However, this line of criticism does not highlight a genuine problem for the claim that there is a necessary connection between moral judgements and admiration. We can accept that many people who perform acts of supererogation have no interest in gaining the admiration of others and continue to hold that such a connection exists. As Heyd points out, we can think that an agent has performed an act of supererogation even if the agent denies this (2006). After all people who perform acts that many people would consider supererogatory often claim to have only done their duty. In their study of moral exemplars

Colby and Damon observed that people who dedicate their lives to moral causes often feel that they are obliged to do so (1992 p.70). For example, Suzie Valdez, who has dedicated her life to feeding the poor of Ciudad Juarez, made it clear to Colby and Damon that she felt she had to help in this way (1992 p.45). Similarly the civil rights activist Virginia Durr claimed that when it came to considering whether or not to dedicate her life to this cause “there were no choices to make” (1992 p.70). The agent could simply be mistaken in their judgement. They may also simply be being modest in claiming that they were only doing their duty. (Heyd 2006). If the agent is not interested in admiration then this may simply be because they do not judge that they have acted in a supererogatory way.

There are two further problems that may be raised against the connection between judgements of supererogation and admiration. The first problem relates to the balance of reasons. As we said earlier, someone who acts in a way that we judge to be supererogatory may have been acting against what we judge that they have most reason to do. It seems odd to feel admiration for someone who has acted against the balance of reasons. To defend the claim of a necessary connection between judgements of supererogation and feelings of admiration, the account of admiration must explain how people can admire those they judge to be acting against their best interests.

The second problem concerns the relationship between admiration and emulation. Feeling admiration for someone is often claimed to be capable of motivating people to emulate those they admire. However, not all judgements of supererogation will motivate the agent to emulate the act. This by itself is not too great a problem as this could be explained in terms of conflicting motivation. What is more of a problem is that it also seems possible to judge that someone has acted in a supererogatory way and feel no desire at all to emulate that person. To solve this problem we require an account of admiration that explains the existence of a defeasible connection between admiration and emulation.

In order to solve both these problems then we will need to examine how feelings of

admiration can lead to action. These two problems provide constraints on the account that can be given of admiration in order for it to be compatible with judgements of supererogation. The account of admiration will need to explain how it is possible to judge that someone has acted against their best interests and admire them. It will also need to explain the defeasible connection between feelings of admiration and emulation. The next chapter of this thesis will attempt to provide such an account.

In the first half of this chapter I looked at ways in which we could formulate the necessary connection between moral judgements and motivation in such a way that is compatible with the existence of a reliable connection between judgements of moral obligation but not judgements of supererogation. I argued that this can be achieved by restricting the necessary connection to overall judgements, judgements about what to do once everything has been taken into account. This successfully explains why there is a reliable connection between judgements of moral obligation and motivation but not for judgements of supererogation. However, as it stands this account is unable to explain the motivational power of judgements of supererogation. In order to do this an explanation needs to be given of how judgements of supererogation can influence our judgements concerning what there is most reason to do. I have argued that this can be achieved by positing the existence of a necessary connection between judgements of supererogation and a pro-attitude that falls short of motivation. I proposed 'admiration' as a plausible candidate for this pro-attitude. In the remainder of this chapter I have defended the existence of a necessary connection between judgements of supererogation and feelings of admiration against problems that could be raised against the compatibility of the two. In the next chapter I will investigate how feelings of admiration are capable of motivating.

Chapter 3 – How do feelings of admiration lead to action?

In the last chapter I argued that there is a necessary connection between judgements of supererogation and feelings of admiration. I then responded to various ways in which it could be argued that feelings of admiration are not always compatible with such judgements. In the first chapter we found that judgements of supererogation are capable of motivating. An internalist account of supererogation must explain how feelings of admiration are capable of motivating. In the last chapter I noted two constraints for such an account. First, the concept of admiration must be able to explain the existence of a defeasible connection between feelings of admiration for someone and a desire to emulate them. The concept of admiration must also allow us to find someone admirable even if they have acted against what they have most reason to do. This chapter will investigate how these feelings are capable of motivating someone to action and respond to problems that could be raised against this account. I will argue that feelings of admiration are able to bring to the fore the right-making features of an action and that this is capable of changing our judgement of what there is most reason to do.

It is claimed that admiration is linked to emulation. In his *Theory of the Moral Sentiments* Adam Smith argues that, “the love and admiration that we naturally conceive for those whose character and conduct we approve of, naturally disposes us to desire to become ourselves the objects of like agreeable sentiments” (2007 p.114). Admiring someone often brings about a desire to be more like them. Admiring the discipline of a hard working colleague may lead me to a desire to work as hard as them. We might think that a disposition to emulate is a necessary part of feelings of admiration. On this account, to feel admiration for someone who has performed a supererogatory act would involve a wish to emulate them. We can have this desire to emulate someone's behaviour even if we think that we are unlikely to get the chance to perform a similar act. Most people do not get the chance to act heroically but this does not prevent them from hoping that if given the chance they would act in a similar way to those they find admirable. This desire is also compatible

with a complete lack of confidence in our ability to actually perform heroic acts.

This account of admiration is able to explain why feelings of admiration may lead someone to action. If admiring someone involves a disposition to emulate the person who performed the act then we can see why these judgements may motivate someone to action. The account also manages to make the connection defeasible. By connecting admiration to a disposition to emulate the account leaves open the possibility that for any individual case of admiration we may not in fact have a desire to emulate the person we admire. This is important as judging something is supererogatory will not always motivate. Our account of admiration then, should not claim that people will always be motivated to act on these feelings.

However, I think there is good reason to think that this account posits too strong a connection between admiration and emulation. We may find someone admirable without being disposed in any way to emulate their behaviour. For example, someone who agrees with Susan Wolf's argument (1982 p.419-439) that we should not try to emulate moral saints may feel no desire to act like a saint. This does not rule out the possibility that this person feels admiration towards those who are saints. In this case there could be admiration without any disposition to emulate. This problem ties in with the second condition that the account of admiration must meet, that admiration is compatible with judging that someone has acted against what they have most reason to do. If we judge that someone has acted against what they have most reason to do then presumably we would not be disposed to emulate that person. Linking admiration with a disposition to emulate seems unable to provide an account of admiration capable of meeting the conditions we set out at the start of the chapter.

By itself this does not give us reason to reject this account of how feelings of admiration can motivate. At the start of the chapter I argued that in order for the claim that there exists a necessary connection between judgements of supererogation and admiration to fit with

our account of supererogation a plausible account of admiration must be given that meets two conditions. The above account may fail to meet these conditions but that does not mean that the account is wrong. Rather it means that it is incompatible with the existence of a necessary connection between judgements of supererogation and feelings of admiration. In order to defend the claim that such a connection exists we will need to find an independent reason to think that a plausible account of admiration should be capable of meeting these conditions.

To do so, let's consider an example of non-moral admiration. Mark and Peter are brothers who have run marathons together since they were teenagers. Mark runs marathons for a hobby. He always manages to complete the race but never comes close to winning. Mark has to make some sacrifices for his hobby. He has to get up early to train before work. Sometimes he has to miss out on social occasions if they are the night before a race. However, these sacrifices do not prevent Mark from leading a well-rounded life. Mark has a family and close friends who he sees regularly. He likes reading, watching films and going to art galleries. Mark's brother, Peter, is a competitive marathon runner and often manages to win. The sacrifices that Peter makes often lead him to miss important family and social occasions. As a result, Peter struggles to form lasting relationships with people. He is always too tired from training to enjoy books or films and never has time to go to an art gallery. Whenever Peter wins a race Mark is full of admiration for him. He knows how hard Peter has to work to achieve what he does and finds this dedication admirable.

Nevertheless, Mark has no desire to emulate Peter. He would not be willing to give up his enjoyment of books and films nor would he wish to sacrifice his trips to art galleries. Indeed, Mark thinks that, on the balance of reasons, Peter should not prioritise marathon running over anything else. He thinks that Mark would be happier if he were able to commit to a long term relationship, develop more meaningful friendships and appreciate culture, yet he still finds his dedication admirable. If we accept that Mark does feel admiration for Peter in the above example then we have reason to think that it is possible to admire someone without feeling any desire to emulate them. We may also admire someone

who we think has acted against what we judge they had most reason to do.

To defend the claim that admiration involves a disposition to emulate we would need to deny that Mark genuinely admires his brother in this example. In the same way that internalists claimed that a moral judgement that is not connected to motivation is not genuine, perhaps we could argue that without a disposition to emulate there is no genuine feeling of admiration. This strikes me as an unreasonable claim to make. In the same way that Mark can admire Peter without being disposed to emulate him, a football supporter who attends the occasional game may admire the dedication of the fan who goes every week without in any way being disposed to emulate him. Further evidence of this claim can be found in a recent social psychology study that found that while feelings of admiration are capable of motivating, they often fail to do so. Feelings of admiration were found to be much less likely to motivate than feelings of benign envy (van de Ven et al 2011). If a necessary connection between admiration and emulation existed then feelings of admiration would motivate more reliably than the results of the study suggest they do. There does not seem to be any good reason to think that for someone to have a genuine feeling of admiration they must be disposed to emulate the person they admire.

Nevertheless, we should not lose sight of the fact that admiration is capable of leading to motivation. If we reject the existence of a necessary connection between the two concepts we should give an alternative account that can explain how feelings of admiration can lead to motivation. Such an explanation can be found by looking again at Jollimore's claim that finding someone admirable involves perceiving her achievements to be valuable in some way (2006 p.159). Valuing someone's achievements could motivate us to try to emulate them, assuming of course that we do not judge that they are beyond our abilities. As van der Burg and Taekema point out, recognising that someone embodies ideals we hold to be valuable can inspire us to try to be like them with respect to that ideal (2004 p.94). Feeling admiration for someone who embodies a particular ideal strengthens our attachment to an ideal. These ideals may or may not be ones that were already recognised as valuable by the agent. Admiring someone for a particular character trait may change what we judge the

balance of reasons to be. They can do this by bringing to the fore the right-making features of an action. This can increase the importance the agent gives to them or recognise the importance of these features for the first time. Someone who values marathon running may want to devote their time to training for marathons. This desire may well increase if they were to meet someone like Peter who embodies the ideal of the devoted athlete. Equally someone who had no previous interest in marathon running may recognise the value of this ideal and be moved to embody it. Admiring someone can lead to a new or strengthened desire to act like that person in respect to that ideal.

However, we are only likely to have this desire if it does not clash with achieving other things that we value. While Mark admires Peter's achievements at marathons he does not want to emulate him as this would mean sacrificing too much of what he holds valuable. If Mark's priorities change and he decides that what he wants most is to win a marathon then Mark's feeling of admiration for his brother may well lead to him having a desire to emulate him. Feelings of admiration are only likely to lead to a desire to emulate if we judge that this would bring about more value than what we would need to sacrifice in order to secure it. This may seem to leave the concept of admiration out of the motivational picture. If we already had most reason to act in a certain way then our admiration for someone plays no part in motivating us to act in a similar way.

We must remember, though, that feelings of admiration are capable of providing us with a new or strengthened desire to emulate the person we admire in respect to certain ideals. This will provide an additional or strengthened reason to emulate the person we admire. Feelings of admiration can change our judgements concerning what there is most reason to do. Someone with a strengthened desire to pursue athletic achievement may judge that this is what they have most reason to do when previously they judged that they had more reason to pursue other interests. Equally though the strengthened desire may not be strong enough to alter judgements about the balance of reasons. This will depend on our previous judgement of the balance of reasons and the strength of this new desire. A strengthened desire to dedicate oneself to marathon running may still lack the strength required to make

it appear to be what there is most reason to do.

Let's sum up what has been said so far about admiration. Feelings of admiration may lead to a new or strengthened desire to emulate. This desire is capable of changing our judgement of what is favoured by the balance of reasons. While there is no necessary connection between feelings of admiration and a desire to emulate, feelings of admiration are capable of changing our judgements about what there is most reason to do. So far we have focussed on situations where an appraiser judges someone else to have acted in a supererogatory way. In order to explain the motivational power of these judgements we must also explain how someone who judges that a possible way of acting open to them is supererogatory could be motivated by this judgement. To relate what we have said so far to first person supererogation judgements, when we judge that someone has performed a supererogatory act we feel admiration towards the agent. However, when the act has not yet been performed it is less clear who is being admired. In first person judgements of supererogation it seems reasonable to think that our admiration is not toward any particular person but towards the type of person who would perform such an act. Judging that a way of acting open to us is supererogatory involves feelings of admiration to people who would perform that act or who have performed an equivalent or similar act in the past. This admiration may lead us to reassess our judgement concerning what is supported by the balance of reasons. If our new assessment of the balance of reasons supports performing the supererogatory act then we will be motivated to perform it so long as we are rational.

On this account we can admire someone who dedicates themselves to fulfilling a particular ideal and also judge that in doing so they sacrifice more value than they gain. This means that we can admire someone who we judge to have acted against what we believe they have most reason to do. This may strike some as problematic, it is one thing to admire someone who is acting in a way you judge to be against what you have reason to do and another to judge that someone is admirable when they have acted against what they have most reason to do. To act against what there is most reason to do leaves one open to the charge of irrationality. This account needs either to explain how we can admire those we

think are irrational or to deny that acting against the balance of reasons is irrational. There does seem to be something strange in admiring people as a result of actions that you deem irrational. For example, suppose Peter were to claim that he gets no pleasure or satisfaction out of running marathons and saw no reason to continue to devote his time to them. Despite this Peter continues to run marathons. In this case the irrationality of Peter's dedication would make it very hard to continue to admire him. Denying that it is always irrational to act against the balance of reasons seem like a more promising solution to this problem. Clearly many cases where we act against the balance of reasons will be irrational. To mount a defence of this sort we will need to outline the situations in which it can be acceptable to act against the balance of reasons.

While it may at first seem contradictory to think that someone who has acted against the balance of reasons is not irrational, this point becomes persuasive when we distinguish between the reasons an agent has for acting and the reasons they believe they have. As Broome points out, someone may have no knowledge of reasons that count in favour of them acting a certain way (2007 p.352). For example, I may judge that what I have most reason to do is to take the train as I know that this is the quickest way for me to get to work. Unknown to me though, the train drivers have gone on strike and choosing to take the train rather than the bus will make me late. In this case although I have most reason to take the bus, it is not irrational for me to choose to take the train. From the point of view of an informed observer assessing my rationality, I have acted against what I have most reason to do although I have not acted against what I believe to be favoured by the balance of reasons. The fact that we judge someone to have acted against what we believe they have most reason to do does not give us reason to call her irrational. One way in which it can be rationally blameless to act against what there is most reason to do is if the agent had no knowledge of the reasons counting against her action.

Although Broome's point gives us good reason to think that acting against the balance of reasons is not irrational, we might still think that it would be odd to admire someone in this situation. We might think them to be rationally blameless but there would seem to be no

reason to feel admiration for such a person. This defence will fail if we cannot explain why it is possible to admire someone who you judge to have acted against the balance of reasons.

Such an explanation can be found in Owen Flanagan's paper 'Admirable Immorality and Admirable Imperfection'. Flanagan argues that we can admire a particular character trait possessed by an individual without admiring the person as a whole (1986 p.43). Flanagan supports this claim by giving the example of Gauguin whose commitment to his art saw him abandon his family to travel to the South Pacific¹⁶ (1986 p.43). Flanagan argues that among the different responses people have to this example is a feeling that Gauguin acted in a way that was both admirable and morally imperfect (1986 p.44). Our admiration for Gauguin's commitment to art is compatible with judging that the abandonment of his family was morally imperfect. Flanagan uses his claim that we can admire an individual trait without admiring the person as a whole to support the idea that admirable moral imperfection is possible (1986 p.52). We can apply this argument to the idea that we can admire those who act against what we judge they have most reason to do. We can judge that someone has acted against the balance of reason and still admire certain character traits that are necessary for that person to act in that way. This does not mean that we admire them for their decision-making powers. Mark can admire the dedication to sporting achievement that allows Peter to win marathons while holding that Peter's judgement of what he has most reason to do is flawed. Just as we can consider Gauguin's act to be admirable but morally imperfect, we might think that Peter's commitment is admirable yet goes against what he has most reason to do. It is possible then to feel admiration for some aspects of a person's character while not admiring every part of that character. In cases of supererogation we may admire someone's moral commitment while feeling that their judgement of what is supported by the balance of reasons leaves room for improvement.

This might give us reason to think that admiration is entirely frame dependent. In other words, we might think that it is always possible to admire a particular character trait regardless of how we feel about the character as a whole. We should note though that

¹⁶ This example was first used by Bernard Williams (1981 pp.20-39).

Flanagan's point is not designed to make admiration compatible with all levels of immorality. Flanagan argues that: "Admiration for particular character traits is invariably conditional," (1986 p.43) and that our admiration depends on, "the assumption that they are moderated by other devotions and sensitivities within the psychological economy of a whole character." (1986 p.43). In other words, our admiration for a particular character trait does depend to some extent upon the person's overall character. In some cases the repellent nature of someone's overall character may prevent us from admiring any individual trait. So, we can admire Gauguin's commitment to art while we judge that abandoning his family was morally imperfect but we would not admire Gauguin's commitment to art if he had murdered his family rather than abandon them. Admiration is not entirely frame dependent then, as there may be cases where the distaste we feel for the character as a whole prevents us from admiring an individual trait.

We have said that it is possible to admire those we judge to have acted against what they have most reason to do providing they have acted in line with what they believe to be the balance of reasons. This allows us to maintain that feelings of admiration are a part of judgements of supererogation and that supererogation is compatible with acting against the balance of reasons. Perhaps though we can admire people even if they act against what they judge they have most reason to do. As Arpaly argues, there are some cases where acting against one's judgement of what is supported by the balance of reasons is less irrational than acting in accordance with this judgement (2000 p.491). Arpaly gives a number of examples to support this claim. One of these examples concerns a student who has always been convinced that she wants to complete a chemistry Ph.D. (2000 p.504). This conviction is maintained despite the fact that a year into the degree she has found the process frustrating and beyond her abilities. The student judges that staying on the course is what she has most reason to do, ignoring the reasons that speak against doing so and the fact that like many people she is reluctant to give up long held goals. However, one day her frustration becomes too much for her and she quits the course, cursing her irrationality as she does so. Arpaly argues that while the student is not acting in an ideally rational way it is more rational for her to act as she did than to continue her studies (2000 p.504). Her decision was made in response to good reasons. The student acted in response to these

reasons even if she was not fully aware that she was doing so (Arpaly 2000 p.505).

Looking back on her decision many years later, after a successful career, the student may judge that her instincts were better at tracking her reasons than her judgement.

I think that we would be able to admire someone who has acted against their judgement of the balance of reasons in a similar way. We would only be able to do so if we judge that the way they have acted is in line with what we judge the agent has most reason to do. Suppose Rex is a young and impressionable philosophy student who becomes convinced by rational egoism, the claim that the most rational thing for an agent to do in any situation is that which maximises his self-interest, after attending a lecture on the subject. This conviction does not represent a lasting commitment on Rex's part, he will abandon it when he attends the next week's lecture on problems associated with the position. Now let's imagine that Rex finds himself in a situation where he can save the life of his friend, Fraser. In doing so Rex would break his own leg. Rex has many friends and his friendship with Fraser is not a particularly close one. Although Rex would miss Fraser were he to die, he judges that this pain is far outweighed by the pain of a broken leg. Rex judges that the rational thing for him to do here is not to save Fraser, as this is unlikely to maximise his self-interest. Suppose though that Rex acts against this judgement and opts to save his friend's life at the expense of suffering a broken leg, he then finds himself troubled by this irrationality. However, after attending his next lecture (after a remarkable recovery) and changing his opinion on rational egoism he decides that the way he acted was rationally preferable to acting in line with his judgement. I think in this case if we judge that this action is supererogatory we would feel admiration for Rex, even though he has acted against what he judges to be the balance of reasons.

There is though, one constraint on admiring those who have acted against what they judge to be their best interests. We would not be able to do so if they acted against what we judged them to have most reason to do. In the case of Rex, we are able to admire him so long as we are not ourselves rational egoists. If we were convinced rational egoists then we would not think that Rex's instincts tracked his reasons better than his judgements. In

certain situations then we can admire someone if they have acted against what we judge to be the balance of reasons or even what they judge to be the balance of reasons. However, we would not admire someone who acted against both our judgement and their judgement of the balance of reasons, at least when this seems clear to both us and the agent.

However, even if we accept that acting against one's judgement of what there is most reason to do may be more rational than acting in line with it this might still be seen to be irrational. After all, a more rational agent would accept that their judgement about what there is most reason to do is flawed before acting against it. While this may be true it only gives us reason to think that the act is irrational if we accept a view of rationality that holds that it is always rationally blameworthy to fail to act in an ideally rational way. However, as Slote argues, a view that holds that rational imperfection is not always blameworthy is more in keeping with our pre-theoretical intuitions about the concept (1989 p.4). To avoid the charge of irrationality it is often enough to satisfy some basic requirements of rationality rather than to act in the most rational way possible (1989 p.30). Accepting this view of rationality allows us to say that although those who act against their best interest are not completely rational they do not deserve to be called irrational. We can reserve this criticism for those who make a clear breach of what is demanded by rationality. In situations where someone is irrational in this sense then, as I pointed out earlier, we will find it difficult to admire her. This allows us to say that admiration is compatible with imperfect rationality up to the point at which it becomes blameworthy. While it may seem odd to admire someone who is rationally imperfect we should bear in mind that we are not admiring her for this imperfection but despite it.

This clashes slightly with what has so far been said about supererogation. If we think that admiration is not possible when someone has acted irrationally then by the account of supererogation we currently have we will not be able to say that such a person is performing a supererogatory act. It could be argued that this creates a problem as there may exist cases where someone has acted in a way that goes completely against the balance of reason but also in a morally praiseworthy way that goes beyond what could

reasonably be demanded by other members of the moral community. For example, suppose someone jumped in front of a train not to save a life but to save a child's favourite toy from being destroyed. Saving a child's toy certainly seems like a morally praiseworthy act. It is also an act that no one could be blamed for not performing. It seems then like it meets all the criteria necessary for it to be called a supererogatory act. However, given the balance between what is gained from such an act and what is risked it seems like this act is irrational. While the loss of a toy to a child is preferably avoided it is not worth risking your life for. If the account of admiration I have given is correct then this will be a very difficult act to admire given how irrational it seems to be. This would raise a problem for the connection I have proposed between acts of supererogation and feelings of admiration.

I do not think that this example causes too many problems for the connection between supererogation and feelings of admiration that I have suggested. As Curtis argues, someone who performs a morally valuable act when the moral value is far outweighed by the risks or costs incurred by the agent, performs a foolish act not a supererogatory one (1981 p.314). This strikes me as a reasonable thing to say, as a natural reaction to someone who acted in such a way would be one of disbelief not praise. As with admiration then, judgements of supererogation are compatible with judging that someone has acted against what they have most reason to do but there comes a point where the balance of reasons is weighed so heavily against the action that we can no longer think of it as supererogatory. This requires a slight amendment to the account of supererogation given in the first chapter. For an act to be supererogatory the costs or risks to the agent must not outweigh the benefits likely to arise from the act to such a great extent that the act becomes foolish.

We might think this account of how feelings of admiration can lead to motivation is incompatible with what was said about supererogation and obligation in chapter one for a different reason. My account of the two types of judgement claimed that to judge that something is morally obligatory involves judging that it is what you have most reason to do while this is not the case for judgements of supererogation. Given this, it seems odd to account for the ability of judgements of supererogation to motivate by claiming that they

will do so if the agent judges that this is what she has most reason to do. If judging that something is supererogatory does not involve judging that it is what we have most reason to do then it seems that this account rules out the possibility that judgements of supererogation may motivate.

However, this objection misrepresents what was said about judgements of moral obligation and supererogation in an important way. The account of judgements of supererogation in the first chapter claimed that, unlike judgements of moral obligation, judging an act to be supererogatory does not necessarily involve a judgement that this act is what there is most reason to do. This does not rule out the possibility that what we judge to be supererogatory may also be what we judge we have most reason to do. The account of moral obligation that I endorsed linked the concept to moral accountability. Judging an act to be morally obligatory involves judging that it would be blameworthy not to perform it. This means that in order for an act to be morally obligatory it must be what we judge we have most reason to do. It does not mean that if an act is judged to be what we have most reason to do then that act is morally obligatory. That depends on whether or not performing the act would make us appropriate subjects of blame. It is possible to judge that an act is what you have most reason to do without judging that you would be a fitting subject of blame if you failed to perform it. There is then, no incompatibility between judging that something is supererogatory and judging that it is what you have most reason to do.

Nevertheless, it could be argued that this account of how feelings of admiration leaves judgements of supererogation playing a very similar motivational role to judgements of obligation. In first person judgements of supererogation (where the agent is also the appraiser) if the agent judges that a supererogatory act is clearly and decisively against the balance of reasons then she wouldn't feel admiration for someone in her position who acted in that way. Alternatively, if the agent judged that the act was clearly what there was most reason to do then she would judge someone in her situation irrational if they failed to perform it. The surprising conclusion is that in cases of first person supererogatory judgements where the agent is convinced that the balance of reasons is clear in what it

supports, judgements of supererogation will either be irrational or they will be motivating. This leaves judgements of supererogation playing a very similar role to judgements of obligation which we said will motivate when the agent is rational. We might think that in trying to explain how judgements of supererogation can motivate we have ended up making the concept redundant.

This objection makes an interesting point about the similarity between these two kinds of judgements in cases where the agent judges the balance of reasons to clearly favour a particular course of action. It is important to note the limits of this similarity though. The first limit is an obvious one but is worth mentioning for the sake of clarity. While there might be similarities in rational blameworthiness in situations where the balance of reasons are clear, this similarity does not extend to moral blameworthiness. We may be rationally blameworthy for failing to perform an act of supererogation that we judge was supported by the balance of reasons but we would not be morally blameworthy. A more important limit is that there will be cases in which the balance of reasons may support two acts to an equal degree. In such cases we can judge an act to be supererogatory, supported by the balance of reasons and not irrational to remain unmotivated by. A further limit is imposed by our status as fallible agents without access to all the relevant information about the balance of reasons. This means that we will often find ourselves in situations where we are unsure about the balance of reasons. Indeed, Kawall argues that even fully informed, ideal observers may disagree about whether or not an act is supererogatory (2009 p.189). This means that there will be many cases where it isn't clear what the balance of reasons supports. It would be justifiable, in such cases, to judge that an act is supererogatory and supported by the balance of reasons but not rationally blameworthy if not performed. Note that the same is not true of judgements of moral obligation, judging that something is morally obligatory commits us to the view that it is supported by the balance of reasons. If we were to discover that the balance of reasons did not support acting in this way we would no longer judge it to be obligatory. The same discovery about a judgement of supererogation would only lead us to withdraw the judgement if the balance of reasons were clear and decisive enough to make us view the act as foolish rather than supererogatory. One interesting consequence of this is that someone who excuses a

decision not to act in a morally praiseworthy way by claiming that the act was supererogatory will still be rationally blameworthy unless they show that the balance of reasons was not clear or decisive.

I have argued that feelings of admiration are able to lead to motivation by changing our judgement concerning what there is most reason to do. They can do this by bringing to the fore the right-making features of an action, leading us to reconsider our judgements about what there is most reason to do. This account is compatible with the two constraints I mentioned at the beginning of the chapter as admiration will not always lead to motivation on this account and they are compatible with a judgement that someone has acted against the balance of reason. I defended this account against various ways in which it might be argued that feelings of admiration are incompatible with judging that someone has acted against the balance of reason. I argued that it is also possible to admire the rationally imperfect but not the rationally blameworthy. An interesting consequence of the account I have given of how judgements of supererogation can motivate was that in cases where the balance of reasons are clear and decisive a judgement of supererogation will either be motivating or irrational. While I accepted that this may be seen to be a problematic feature of my account it is worth noting that the cases in which the balance of reasons are clear and decisive are likely to be rare.

Conclusion

In conclusion, after raising the problem of judgements of supererogation for motivational judgement internalism I have proposed a way in which internalism can handle this problem. In the first chapter of this paper I argued that the internalist claim that there is a reliable connection between moral judgements and motivation can be made plausibly about judgements of moral obligation but not judgements of supererogation. As a result of this I argued that a plausible form of internalism must explain the necessary connection that is claimed to exist between moral judgements and motivation in a way that is compatible with the existence of a reliable connection between some types of moral judgements but not others. In the second chapter I argued that this can be achieved by restricting internalism to 'all things considered' normative judgements. This account is able to explain why we would expect someone to be motivated by a judgement that an act is morally obligatory but not by a judgement that an act is supererogatory. This would suggest that internalism has nothing to say about judgements of supererogation. However, while an internal connection between judgements of supererogation and motivation may not be plausible, I argued that an internal connection between these judgements and a pro-attitude that falls short of motivation would allow internalists to propose an internal connection between judgements of supererogation and feelings of admiration. I proposed 'admiration' as the obvious candidate for this pro-attitude and defended the existence of a necessary connection between feelings of admiration and supererogation. In the last chapter I explained how feelings of admiration may alter an agents judgement of what is supported by the balance of reasons. This account allows internalists to explain the motivational power of judgements of supererogation with an internal connection.

An interesting question that could be dealt with in further research is how this internalist account of judgements of supererogation compares with an externalist account and the implications of this for the wider debate between internalists and externalists about moral judgements. Another issue that arises as a result of this paper concerns the consequences

that this account has for other debates about the nature of moral judgements. It is tempting to think that the claim that expressivism is in a better position than cognitivism to accommodate internalism is reinforced by the idea that different moral judgements bring with them different kinds of necessary connections but this is a question for another project.

Bibliography

- Attfield, Robin (1995) *Value, Obligation and Meta-Ethics* (Amsterdam: Rodopi Bv Editions)
- Anscombe, G. E. M (1958). 'Modern Moral Philosophy' in *Philosophy* (Vol.33 No. 124 pp.1-19)
- Audi, Robert (2005) 'Wrongs Within Rights' in *Philosophical Issues* (Vol.15 pp.121-139)
- Brink, David (1989) *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press)
- Broome, John (2007) 'Does Rationality Consist in Responding Correctly to Reasons' in *Journal of Moral Philosophy* (Vol.4 No.3 pp.349-371)
- Carwell, Diane. (2007). 'Subway Rescuer Receives the City's Highest Award' Available: <http://www.nytimes.com/2007/01/05/nyregion/05life.html>. Last accessed 31st July 2011.
- Chrisman, Mathew and Ridge, Mike (Work in Progress) "Ought and Must"
- Colby, Anne and Damon, William (1992) *Some Do Care: Contemporary Lives of Moral Commitment* (New York: The Free Press)
- Curtis, Barry (1981) 'The Supererogatory and the Foolish' in *Journal of Value Inquiry* (Vol 15 No.4 pp.311-318)
- Daly, Michael. (2010). "Real heroes, such as subway savior Wesley Autrey, don't do it for the glory" Available: http://articles.nydailynews.com/2010-05-27/local/27065434_1_wesley-autrey-subway-hero-real-heroes . Last accessed 31st July 2011.

- Dancy, Jonathan (2004) *Ethics Without Principles* (Oxford: Clarendon Press)
- Darwall, Stephen (1983) *Impartial Reason* (London: Cornell University Press)
- Darwall, Stephen (2006) *The Second Person Standpoint* (Cambridge: Harvard University Press)
- Gert, Joshua (2004) *Brute Rationality* (Cambridge: Cambridge University Press)
- Harman, Gilbert (1999) *Reasoning, Meaning and Mind* (Oxford: Oxford University Press)
- Heyd, David (1982) *Supererogation: Its Status in Ethical Theory* (Cambridge: Cambridge University Press)
- Heyd, David (2008) 'Supererogation', *The Stanford Encyclopedia of Philosophy* (Autumn 2008 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/supererogation/>.
- Flanagan, Owen (1986) 'Admirable Immorality and Admirable Imperfection' in *The Journal of Philosophy*, (Vol. 83 No. 1 pp. 41-60)
- Jollimore, Troy (2006) 'Morally Admirable Immorality' in *American Philosophical Quarterly* Vol. 43 No.2 pp.159-170)
- Joyce, Richard (2001) *The Myth of Morality* (Cambridge: Cambridge University Press)
- Kawall, Jason (2009) 'Virtue theory, ideal observers, and the supererogatory' in *Philosophical Studies* (Vol. 109 No. 2)
- Kamm, F. M. (2001) *Morality, Morality Volume Two: Rights, Duties and Status* (Oxford: Oxford University Press)
- Kitcher, P. S. (1984). "In Defence of Intentional Psychology", in *Journal of Philosophy* (Vol.81 pp.89-106)

Mason, Ellinor (2008) 'An Argument Against Motivational Internalism' *Proceedings of the Aristotelian Society* (Vol.108 pp.135-156)

Mellema, Gregory (1991) *Beyond the Call of Duty: Supererogation, Obligation and Offence* (New York: State University of New York Press)

Mill, John Stuart (2004) *Utilitarianism*, Kindle version, accessed 20 June 2011 from Amazon.co.uk

Provis, Chris (1981) "Reason and Emotion" in *Canadian Journal of Philosophy* (Vol.11 No.3 p.439-456)

Pybus, Elizabeth M. (1982) 'Saints and Heroes' in *Philosophy* Vol. 57 No.220 pp.193-199 (Apr 1982)

Schroeder, Mark (2007) "Teleology, Agent-Relative Value, and the 'Good'" in *Ethics*, (Vol. 117 pp. 265-295)

Slote, Michael (1989) *Beyond Optimizing* (Cambridge: Harvard University Press)

Smith, Adam (2007) *The Theory of Moral Sentiments* (New York: Cosimo)

Smith, Michael (1994) *The Moral Problem* (Oxford: Blackwell)

Stich, S. (1996) *Deconstructing the Mind* (New York: Oxford University Press)

Stocker, Michael (1979) 'Desiring the Bad: An Essay in Moral Psychology' in *Journal of Philosophy* (1979 pp738-52)

Suikkanen, Jussi (2009) 'Consequentialism, Constraints and the Good-Relative-to: A Reply to Mark Schroeder.' in *Journal of Ethics & Social Philosophy* (pp.1-8)

Svavarsdóttir, Sigrun (1999) 'Moral Cognitivism and Motivation', *Philosophical Review* (Vol. 108 No. 29 pp.161-219)

Rosati, Connie S. (2008) 'Moral Motivation', The Stanford Encyclopedia of Philosophy (Fall 2008 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/moral-motivation/>

Urmson, J. O. (1958) 'Saints and Heroes' Reprinted in *Moral Concepts* Joel Feinberg (ed.) (Oxford: Oxford University Press 1969)

van der Burg, Wibren and Taekema, Sanne (2004) 'Motivation by Ideal' in *Philosophical Explorations* (Vol. 7 No.1 pp.91-98)

van de Ven, N. Zeelenberg, M. and Pieters, R. (2011) 'Why envy outperforms admiration' in *Personality and Social Psychology Bulletin* (Vol.37 No.6 pp.784-95)

Wallace, R. Jay (2009) 'Practical Reason', The Stanford Encyclopedia of Philosophy (Summer 2009 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2009/entries/practical-reason/>

Watson, Gary (1987) 'Responsibility and the Limits of Evil.' in Schoeman, Ferdinand, ed.. *Responsibility, Character, and the Emotions*. (New York: Cambridge University Press)

Watson, Gary (2007) 'Morality as Equal Accountabiliy: Comments on Darwall's the Second Person Standpoint' in *Ethics* (118 No.1 pp.37-51)

Wedgwood, Ralph (2007) *The Nature of Normativity* (Oxford: Oxford University Press)

Williams, Bernard (1981) *Moral Luck* (Cambridge: Cambridge University Press)

Wolf, Susan (1982) 'Saints and Heroes' in *Journal of Philosophy* (Vol.79 No.8 pp.419-439)

Zimmerman, Michael J. (1996) *The Concept of Moral Obligation* (Cambridge: Cambridge University Press)