Forthcoming in *The Journal of Value Inquiry* (2018)
Penultimate draft - please cite the final version.

INTENTIONS, MOTIVES AND SUPEREROGATION

Claire Benn, cmabenn@gmail.com

Abstract: Amy saves a man from drowning despite the risk to herself, because she is moved by his plight. This is a quintessentially supererogatory act: an act that goes above and beyond the call of duty. Beth, on the other hand, saves a man from drowning because she wants to get her name in the paper. On this second example, opinions differ. One view of supererogation holds that, despite being optional and good, Beth's act is not supererogatory because she is not praiseworthy; the other agrees that Beth is not praiseworthy but holds that her act is nevertheless supererogatory because, while supererogatory actions are generally performed by praiseworthy agents, an individual need *not* be praiseworthy for their act to be supererogatory. In this paper, I raise a problem for this latter position, which I shall call the Anti-Motivation View of supererogation. While the Anti-Motivation View rejects the claim that the agent's motivations are important, it accepts that the agent's *intentions* are. Thus, this view assumes that a clear, coherent distinction can be drawn between intentions and motivations in the context of supererogation. This distinction is often attributed to Mill; however, his original discussion reveals an inconsistency. Consider Clara, who saves a man from drowning because she wants to torture him later. According to Mill, Clara's act is not morally good; those who follow Mill will have to accept that it is therefore not supererogatory. Mill asserts that Clara's act differs from Amy's and Beth's not just in motivation but also in intention. I question this explanation and demonstrate that the distinction between motivation and intention is not as clear as the Anti-Motivation View has supposed and that the gap between the Anti-Motivation View and the

alternative Praise-Based View is either much greater or much smaller than previously thought.

1 Introduction

Supererogatory actions go beyond the call of duty. They are actions that, while being morally good,[1] are neither morally required nor morally forbidden. On these two features—goodness and optionality—all accounts of supererogation agree. I focus in this paper on two further features proposed by some accounts of supererogation: praiseworthiness and intentionality.

With respect to supererogation and praiseworthiness—and the motivation of the agent more generally—two camps have emerged. According to one, the praiseworthiness of the agent is a necessary condition for an action to be supererogatory, while the other rejects that necessity. However, I challenge the idea that the latter camp can sufficiently distinguish themselves from the former without giving a radically permissive account of the supererogatory, given that both camps acknowledge the importance of the *intentions* of the agent.

In §2, I outline the two different positions on motivations taken by accounts of supererogation by introducing two cases of rescuing a drowning man. In §3, I explore the importance of intention for supererogation. I then turn, in §4, to the origins of the distinction between intentions and motivations. This gives us the third drowning man case. In light of this case, I challenge the distinction between intentions and motivations as construed by those who reject the claim that for an action to be supererogatory the

1. More specifically, they are morally better than a permissible act that we could have performed instead. For brevity, I use the term 'morally good' throughout this paper.

agent must be praiseworthy. In §5, I examine possible responses to the three cases and argue that the two positions on praiseworthiness and supererogation are either much more radically different or much more similar than previously thought. I conclude in §6 with a final argument that the role that intentions play that make them so important is one that motives can have too.

2 Motivations

Let's begin with a straightforward case:

> **Amy and the First Drowning Man**: Amy saves a man from drowning because she is moved by his plight.

Under the supposition that saving the drowning man is not morally required nor morally forbidden[2] and saving a life is morally good, Amy's action is supererogatory. On this all accounts of supererogation agree.

> **Beth and the Second Drowning Man**: Beth saves a man from drowning because she wants to get her name in the paper.

Is Beth's action supererogatory? It is over this case that accounts disagree. The answer depends on the relationship that is assumed between praiseworthiness (or the agent's

2. Due to the fact, say, that the level of risk to the agent involved is too high for this action to be obligatory but not so high as to be morally impermissible. I assume that all the rescue cases discussed have this form. Alternative cases of the same structure could be given, if the reader objects to the particulars of this one.

motivations more generally) and supererogation. According to one position, which I shall call the Praise-Based View, praiseworthiness is a necessary part of what it takes for an action to be supererogatory; the other, which I call the Anti-Motivation View, denies this. I start by exploring the Praise-Based View.

Praise and praiseworthiness form an important aspect of many accounts of supererogation.[3] Marcia Baron, for example, claims that supererogatory actions are "beyond duty and […] are morally good and praiseworthy."[4] Russell Jacobs claims that "[s]upererogatory acts, are by definition, acts that are morally good or morally praiseworthy, but not the agent's duty to perform."[5] Both Joseph Raz and Gregory Mellema describe supererogatory actions as ones that are praiseworthy to do (and not blameworthy not to do).[6]

As can be seen from the above statements, these authors phrase their position as if praiseworthiness is an attribute of an *act*. However, saying an *act* is praiseworthy is just shorthand for saying that the action is morally good (or perhaps morally optional and morally good) in which case it adds nothing to the features of supererogatory actions already established. Outside discussions of supererogation, praiseworthiness is more often seen as an attribute of the *agent* and their motivation. Thus, I believe the

3. See for example Robin Attfield, "Supererogation and Double Standards," *Mind* 88, no. 352 (1979): 481–99; Michael Clark, "The Meritorious and the Mandatory," *Proceedings of the Aristotelian Society* 79 (1978): 23–33; Barry Curtis, "The Supererogatory, the Foolish and the Morally Required," *Journal of Value Inquiry* 15 (1981): 311–18. It is also a premise of the debate between Elizabeth M. Pybus, "Saints and Heroes," *Philosophy* 57, no. 220 (1982): 193–99; Patricia M. McGoldrick, "Saints and Heroes: A Plea for the Supererogatory," *Philosophy* 59, no. 230 (1984): 523–28. See also the discussion of their disagreement in Russell A. Jacobs, "Obligation, Supererogation and Self-Sacrifice," *Philosophy* 62, no. 239 (1987): 97.
4. Marcia Baron, "Kantian Ethics and Supererogation," *The Journal of Philosophy* 84, no. 5 (1987): 239.
5. Jacobs, "Obligation, Supererogation and Self-Sacrifice," 97.
6. Joseph Raz, "Permissions and Supererogation," *American Philosophical Quarterly* 12, no. 2 (1975): 164. Gregory Mellema, "Quasi-Supererogation," *Philosophical Studies* 52 (1987): 142.

Praise-Based View of supererogation ought to be interpreted as taking praiseworthiness to be an attribute of the agent, not of the act, because we praise *someone* for performing an act.[7] So let us assume that what these authors are really saying is as follows:

> **The Praise-Based View**: For an action to be supererogatory, in addition to having certain features itself (such as it being optional and good), the act in question must also be *performed by an agent who is praiseworthy in virtue of their motivation for performing it.*

Thus, according to the Praise-Based View, Beth's action is not supererogatory because Beth is not praiseworthy, given that her motivation is to get her name in the paper. This does not mean that her action is not morally good; but her motivation excludes it from being supererogatory.

An alternative view of supererogation rejects the necessity of a praiseworthy motivation on the part of the agent:

> **The Anti-Motivation View**: For an act to be supererogatory, it must have certain features (such as optionality and goodness); however, while agents who perform supererogatory acts are *often* praiseworthy, this is not necessary: their motivations are not important to assessing whether the action they perform is supererogatory.[8]

7. For further discussion, see Phillip Montague, "Acts, Agents, and Supererogation," *American Philosophical Quarterly* 26, no. 2 (1989): 101–11.
8. Those taking such a position include: J.O. Urmson, "Saints and Heroes," in *Moral Concepts*, ed. Joel Feinberg (Oxford: Oxford University Press, 1969), 60–73; John Rawls, *A Theory of Justice*, Rev. Ed. (Cambridge, Mass.: The Belknap Press of Harvard University Press, 1999); David Heyd,

The Anti-Motivation View therefore agrees with the Praise-Based View that Beth's motivation renders her not *praiseworthy* but, contrary to the Praise-Based View, considers Beth's action supererogatory nonetheless because it does not matter that she is not praiseworthy: she saves someone from drowning, just as Amy does. They both do something supererogatory. The question is then: why reject the necessity of praiseworthiness in an account of the supererogatory?

One reason comes from a concern about the praiseworthiness of self-sacrifice. Nancy Stanlick, for example, argues that "supererogatory acts are, in fact, not necessarily praiseworthy, morally commendable, permitted, or even good except under very stringent conditions."[9] Jean Hampton echoes her concern, stating that "not all self-sacrifice is worthy of our respect or moral commendation."[10] Indeed, far from praising agents who perform supererogatory acts, we should perhaps worry that their self-sacrifice demonstrates that they do not value themselves.[11] As Jason Kawall asks: "What are we to make of the selfless actions of a woman whose self-esteem has been crippled by a verbally abusive husband and a traditional society which teaches that women are first and foremost caregivers? Or consider a cult member who does not lack

Supererogation: Its Status in Ethical Theory, Cambridge Studies in Philosophy (Cambridge: Cambridge University Press, 1982); Alfred Archer, "Are Acts of Supererogation Always Praiseworthy?," *Theoria* 82, no. 3 (September 2016): 238–55.

9. Nancy A. Stanlick, "The Nature and Value of Supererogatory Actions," *Journal of Social Philosophy* 30, no. 1 (February 1999): 210. Stanlick draws on Susan Wolf's work, who questions "the assumption that it is always better to be morally better" where 'morally better' is understood as always putting the interests of others ahead of our own (Susan Wolf, "Moral Saints," *The Journal of Philosophy* 79, no. 8 (1982): 419–39).

10. Jean Hampton, "Selflessness and the Loss of Self," *Social Philosophy and Policy* 10, no. 1 (1993): 1.

11. Ibid., 8.

self-esteem but who has placed the cult leader on a high pedestal such that he would do anything (thus going beyond duty) to benefit the leader."[12]

This line of thought rejects any necessary link between acts of self-sacrifice and the praiseworthiness of the agents who perform them. If supererogatory actions were defined in terms of self-sacrifice and were, *because of this*, considered praiseworthy, then these objections would have some bite. However, the best characterisation of the Praise-Based View is not that the agent necessarily has praiseworthy motivation simply in virtue of performing an otherwise supererogatory (or self-sacrificing) action. As Badhwar claims, sometimes "even altruism fails to be a virtue."[13]

So let us set aside any assumption that the optionality and goodness of a supererogatory act somehow entail the praiseworthiness of the agent in performing it. Instead, let's consider the Praise-Based View to be claiming that the praiseworthy motivation of the agent is a necessary condition of supererogation *in addition to* the act being optional and good. This will exclude as supererogatory those cases of abuse and subservience described above. However, even this position has been rejected by some theorists. Alfred Archer, for example, argues that the best reasons for including praiseworthiness are inadequate, given that 'supererogatory' does have a natural language meaning—'superfluous'—that does not entail praiseworthiness, as well as the fact that there are plenty of historical and contemporary accounts of supererogation that do not include praiseworthiness, instead defining supererogation simply in terms of

12. Jason Kawall, "Self-Regarding Supererogatory Actions," *Journal of Social Philosophy* 34, no. 3 (September 2003): 495.
13. Neera Kapur Badhwar, "Altruism Versus Self-Interest: Sometimes a False Dichotomy," *Social Philosophy and Policy* 10, no. 1 (1993): 90–117. See also Curtis, "The Supererogatory, the Foolish and the Morally Required"; Clark, "The Meritorious and the Mandatory," 32.

optionality and goodness.[14] He also gives some positive reasons for rejecting the inclusion of praiseworthiness. These include the fact that the most interesting issues about supererogation concern the first two conditions—goodness and optionality—not praiseworthiness[15]; that building praiseworthiness into our account of supererogation leaves us with a much less unified account of the deontic concepts[16]; and that supererogation is best understood as an act description and so information about the agent's motivations should be excluded.[17]

I do not evaluate the arguments for and against the Praise-Based View of supererogation here. Instead, my focus is on the fact that the proponents of the Anti-Motivation View, while rejecting the importance of the motivation of the agent as important, argue that the agent's *intentions* are important. I question whether this distinction can be maintained in an account of the supererogatory and whether it can give the verdicts that proponents of the Anti-Motivation View suppose it can.

---

14. Archer, "Are Acts of Supererogation Always Praiseworthy?" He cites as evidence: supererogation as understood in Christian theology (David Heyd, "Supererogation," ed. Edward N. Zalta, *Stanford Encyclopedia of Philosophy*, 2016.), Dale Dorsey, "The Supererogatory and How to Accommodate It," *Utilitas* 25, no. 3 (2013): 355–82; Michael Ferry, "Does Morality Demand Our Very Best? On Moral Prescriptions and the Line of Duty," *Philosophical Studies* 165, no. 2 (2013): 573–89, https://doi.org/10.1007/s11098-012-9968-6; Heyd, *Supererogation*, 1982; Douglas W. Portmore, *Commonsense Consequentialism: Wherein Morality Meets Rationality* (Oxford: Oxford University Press, 2011).

15. Archer, "Are Acts of Supererogation Always Praiseworthy?"

16. Ibid. Archer argues that our deontic concepts would be less unified in the sense that supererogation would not involve a connection to an assessment of the agent's motives, while obligation and prohibition do not (Ibid., 9.). And, as Archer later discusses, it is less parsimonious to reserve the term 'supererogatory' for those actions that are optional, good and praiseworthy, as another term would then be needed for those actions that are optional, good and not praiseworthy (Ibid., 16.). On the other hand, there may well be reasons, as Kawall supposes, to reserve the term supererogation for when "everything goes right" (Kawall, "Self-Regarding Supererogatory Actions," 495.). Nevertheless, it is not the purpose of this paper to determine the merits of this argument against the exclusion of praiseworthiness; but rather to reveal a tension that arises from including the agent's intentions and excluding their motives, as proposed by the advocates of the Anti-Motivation View.

17. Alfred Archer, "Supererogation and Intentions of the Agent," *Philosophia (United States)* 41, no. 2 (2013): 451. Philip Montague notes the same: "standard definitions of the concept of supererogation are formulated entirely in terms of references to features of actions" (Montague, "Acts, Agents, and Supererogation," 102.).

3 Intentions

The proponents of the Anti-Motivation View reject the claim that an agent needs to be praiseworthy in order for their act to be supererogatory because they reject the claim that the agent's *motivations* matter to whether an act is supererogatory. However, those same proponents do maintain the importance of the *intentions* of the agent. This is because the intentions of the agent play two significant roles in accounts of the supererogatory.

Firstly, intentions are *necessary* in the sense that for an act to be supererogatory it must be intentional:

> **The Intention Requirement**: For an act to be supererogatory, it must be intended by the agent under a description in virtue of which it is morally good.

This requirement rules out as supererogatory those actions that are "accidental, unconscious or involuntary."[18] Suppose that, while climbing an apple tree for fun, you accidentally knocked an apple off its branch into the hands of a passer-by. Despite the fact that your action has the consequence of giving an apple as a gift to a stranger—something that could easily be understood as supererogatory—we would not call this action supererogatory because it was performed accidentally: you had no *intention* of giving that apple to that stranger.[19] Similarly ruled out are actions that are not within

18. Heyd, *Supererogation*, 1982, 116. I disagree with Heyd's further claim that the agent "must have the intention of promoting the good by his action." (Ibid., 133.)
19. It seems clear from these kind of examples that accidental actions cannot be supererogatory. Archer gives the example of Louise who makes a large donation to a famine charity by bank transfer but does

our control, such as automatic responses like sneezing and myoclonic jerks, as well as those performed by while hypnotised or sleepwalking, and those performed against our will.[20]

Note that to fulfil the Intention Requirement the agent does not need to intend their action as self-consciously altruistic. As Archer has argued, the act in question need only be intentional under a description that picks out features *in virtue of which* the act is morally good (or, in his terms, morally meritorious).[21] The *agent* does not have to be conscious of the moral nature of their act or intend it as morally good. If an agent intends only to, for example, 'give money to someone in need', their action would still be considered supererogatory, whether or not they also have an intention the content of

---

so accidentally: she "intended to transfer the money between two of her own accounts to enable her to buy an expensive car" (Archer, "Are Acts of Supererogation Always Praiseworthy?"). This is not supererogatory. However, it should be noted that it is possible to accidentally, and therefore unintentionally, do what we intended to do: for example, "If I intend to be shaking in order to signal my confederate, and this intention makes me nervous, so that I shake, I am shaking because I intend to do so—though not intentionally" (Kieran Setiya, "Intention," ed. Edward N. Zalta, *Stanford Encyclopedia of Philosophy Online*, 2015.). Consider my modified version of Archer's case: Eleanor makes a donation to a famine charity by bank transfer but does so accidentally: she intended to transfer the money to her savings account and she intended only *after doing so* to send some money to the famine charity but she accidentally did it the wrong way round. It just so happens it was the same amount as she intended all along. This action was accidental but was nevertheless *in accordance* with her intentions. A test of whether an action is in accordance with our intentions is whether we regret the action or whether we would have done otherwise. Archer states that Louise "would like to retrieve the money but is unable to do so" (Archer, "Are Acts of Supererogation Always Praiseworthy?") whereas, on my case, Eleanor probably would not. Note, however, if *regret* is really what matters then there are actions that could be supererogatory despite being performed unintentionally and *not* in accordance the agent's intentions. We could imagine a case where Louise did not regret the action and would not retrieve the money even if she could, despite the fact that it was a complete accident and not in accordance with her intentions because, *after* having sent the money, the plight of those suffering from famine became vivid to her and so she does not regret sending it. These counterfactuals may be more important than the original intention. Unfortunately, it is beyond the scope of this paper to delve more deeply into this possibility.

20. For more on the issue of forced supererogation, see Bashshar Haydar, "Forced Supererogation and Deontological Restrictions," *The Journal of Value Inquiry* 36 (2002): 445–54.

21. Archer, "Supererogation and Intentions of the Agent," 458. See his Moral Intention Requirement Two (MIR 2).

which is explicitly 'to do good'. Intending to give money to someone in need is supererogatory because it is intended *under a description in virtue of which* it is good.

Donald Davidson argues that "some verbs describe actions that cannot be anything but intentional: asserting, cheating, taking a square root, and lying are examples."[22] The Intention Requirement implies that we should add 'supererogating' to this list: it is not possible to accidentally perform a supererogatory act.[23]

Secondly, intentions are sufficient for an act to be supererogatory insofar as we want to include *attempts* to bring about some good even if they do not succeed:

> **The Intention Sufficiency**: For an act to be supererogatory, in addition to its being optional, it is enough that it was *intended* under a description in virtue of which it is morally good even if it fails to be.[24]

The classic examples of supererogatory action—such as the soldier who throws himself on the grenade or the hero who runs into the burning building—all have a feature in common: they all *succeed* in doing good. The person who jumps on the grenade succeeds in saving the life of his companion; the person who runs into the burning building succeeds in saving the life of the child in danger. However, while success is a feature of some paradigmatic examples of supererogatory actions, it is not a necessary condition. An action can be good in terms of the consequences that the agent *intends* to

22. Donald Davidson, *Essays on Actions and Events* (Oxford, England: Clarendon Press, 1980), 45.
23. Except in the sense that we can accidentally do something that *when done intentionally* is supererogatory or that we can accidentally do something that is *usually considered* supererogatory.
24. The sufficiency of intentions for supererogation is explicitly endorsed by, for example, Heyd, *Supererogation*, 1982, 133.

bring about by performing it, even though she might fail to actually bring those consequences about.[25]

Take Captain Oates, who famously said "I am going outside and may be some time." He chose to die in the hope that it would allow his fellow explorers to survive. However, they still died 16 days later. While Oates' self-sacrifice failed to help them, what is of importance is that he *tried* to help them.

Now it is true that in some sense Oates did succeed in giving his fellow explorers a better chance of survival. This might lead some to think that all cases of attempts can be re-described in this way. However, it cannot account for cases where, for example, someone tries to save a person who is in fact already dead. Their efforts do nothing to increase the deceased person's chances of surviving. The only way to re-describe their actions is to say that they *tried* to rescue someone. This is merely to re-iterate the claim that *intending* to produce some good is sufficient.[26] Thus, as Jackson puts it, "the intention to effect such important goods […] is sufficient for an act to be supererogatory even if it fails."[27]

---

25. The relationship between attempts, completions and moral goodness may well be quite complex. There may be a difference, for example, between thinking that the moral goodness comes straightforwardly from an attempt (and thus completions, as a subset of attempts—namely, successful attempts—also have this goodness) and thinking that completions are what are morally good, but attempts share or sufficiently track what it is that make completions themselves good. It makes no difference for my discussion here, however, as either account has the same effect in terms of which actions are classed as supererogatory. For more on attempts, see Gideon Yaffe, *Attempts: In the Philosophy of Action and the Criminal Law* (Oxford University Press, 2010).

26. The sufficiency of intending can in part be further explained because, from the point of view of the deliberating agent, intending to do something and intending to *try* to do something look the same (T.M. Scanlon, *Moral Dimensions: Permissibility, Meaning and Blame* (Harvard University Press, 2008), 46.).

27. M.W. Jackson, "The Nature of Supererogation," *The Journal of Value Inquiry* 20 (1986): 295. There is a caveat however about what the agent *ought* to have known or believed about the act and its consequences. If it is only through wilful ignorance or negligence that I believe that my action would be of a certain sort, then its status as supererogatory would be undermined. The standard could be made more or less strict depending on your views about the epistemological requirements that apply to deliberating agents.

The Intention Requirement and the Intention Sufficiency look like two separate conditions. In fact they are closely related: intentions are the means by which we pick out the appropriate act *description* of an act to assess as good or not.

Elizabeth Anscombe gives the classic example of how the very same act can have multiple descriptions: we can imagine an act that can be equally well described as "'sawing a plank', 'sawing oak', 'sawing one of Smith's planks', 'making a squeaky noise with the saw', 'making a great deal of sawdust'' and so on and so on."[28] She points out that "a man may know that he is doing a thing under one description, and not under another."[29] For example "[h]e may know that he is sawing a plank, but not that he is sawing an oak plank or Smith's plank; but sawing an oak plank or Smith's plank is not something else that he is doing besides just sawing the plank that he is sawing."[30] The same applies to intending to do something as we can intend to do something under one description but not another: he could *intend* to saw a plank but not intend to be sawing *Smith's* plank.

It is important to identify the appropriate description under which to evaluate an action because different descriptions may well have different ethical value. For example, sawing a plank may be ethically neutral, but sawing Smith's plank may be morally bad, if he did not have Smith's permission to do so. The agent's intentions help us to pick out the relevant act description which we can assess as morally good or not, and thus as supererogatory or not.

Thus, the necessity and sufficiency of intentions are actually two sides of the same coin: the intention separates the act we assess from its accidental outcomes *and*

28. G.E.M. Anscombe, *Intention*, 2nd Editio (Harvard University Press, 1957), 11.
29. Ibid.
30. Ibid., 12.

from its successful completion. Suppose I leave behind the white goods I bought when I move out of my flat, though I am not required to. I do this because I believe that they have an electrical fault and I hope that the landlord will be sued by the next tenants. In fact the white goods happen to be in perfect condition and add to the rental value of the property. The appropriate act description is not 'giving a gift of some white goods' or 'increasing the rental value of the property' because both of these were not intended. A better description is of an (albeit failed) attempt to harm my landlord's interests: this was my intention. It is clear from *this* description that my act is not supererogatory. Or suppose that I run into a burning building only to discover that the child inside has already succumbed to the smoke. This should be assessed under the description of an (albeit failed) attempt to help save a child, rather than under the description of 'the fruitless risking of one's life.' Thus, in both cases, it is the intentions of the agent that allows us to specify the act description under which we can assess the act as supererogatory or not.[31]

The main proponents of the Anti-Motivation View hold intentions to be necessary and sufficient in the ways outlined above. They nevertheless reject the importance of the agent's motivations in giving an account of the supererogatory. In order to maintain such a position—and thus to distinguish itself from the Praise-Based View—the Anti-Motivation View must successfully distinguish intentions from motivations. I turn now to casting doubt on whether this can be done as cleanly and clearly as might first be thought.

31. In this, therefore, the role of intentions in accounts of the supererogatory is perhaps very different to the role it has in accounts of duty. As Heyd says, "At least some theories of duty maintain that a duty or an obligation can be discharged or met unintentionally, and *a fortiori* with no intention of benefiting anyone. And as for the second meaning of 'intended', some philosophers (notably Moore and Ross) would deny that mere intention (not complemented by actual success) can suffice to make an act duty-fulfilling" (Heyd, *Supererogation*, 1982, 133.).

4 The Third Drowning Man

The distinction between intentions and motives in discussions of supererogation is often attributed to J.S. Mill.[32] In defending Utilitarianism from detractors who claimed that it was far-fetched to expect—or to want—people to act from "the inducement of promoting the general interests of society", Mill tells us not to confuse the "rule of action" with "the motive to do it."[33] He claims that, while it is the business of ethics to tell us our duty, no system of ethics "requires that the sole motive of all we do shall be a feeling of duty" which is important because in fact "ninety-nine hundredths of all our actions are done from other motives."[34]

Mill then discusses two cases like that of Amy and Beth. Mill agrees that Amy and Beth both do "what is morally right."[35] The rightness of Amy's and Beth's acts is the same because they are the same act, and they are the same act because Amy and Beth have the same intention (to save the man from drowning), despite the fact that they have different reasons, and thus motives, for doing so (being moved by his plight and fame, respectively).

However, this discussion in the main body of text in Mill stems from an exchange he had with the Rev. J. Llewellyn Davies, "an opponent, whose intellectual and moral fairness is a pleasure to acknowledge", which he discusses in a footnote

32. For example, Archer, "Are Acts of Supererogation Always Praiseworthy?"; Heyd, *Supererogation*, 1982.
33. John Stuart Mill, *The Collected Works of John Stuart Mill, Volume X - Essays on Ethics, Religion and Society*, ed. John M. Robson (Toronto: University of Toronto Press, 1969), 219.
34. Ibid.
35. Ibid.

accompanying the examples above.[36] And in this footnote a curious third case of a drowning man is raised by Davies (I have given the rescuer a name):

> **Clara and the Third Drowning Man**: Clara saves a man from drowning in order to "inflict upon him more exquisite tortures."[37]

Davies asks whether this too is clearly an example of a 'morally right action.' Recall that Mill says that Amy and Beth both do something morally right despite the fact that they do so for different reasons. So you might think that Mill would respond the same way here: that given that she saves the man from drowning Clara also does something morally right, despite her reasons for doing so. Surprisingly, this is not Mill's response.

Mill claims that Clara's action differs from Amy's, who saves from duty, and of someone who saves from beneficence, not only in terms of *motive*, but also in terms of the *act itself.* This is because, he argues, Clara's saving the drowning man was "the necessary first step of an act far more atrocious than leaving him to drown would have been."[38] The act itself is different, he argues, because the *intention* is different and "the rightness or wrongness of saving a man from drowning does depend very much […] upon the *intention*" but denies that it depends on the motive, as the phrasing of Davies's original question had implied.[39]

The case of Clara demonstrates the difficulty of distinguishing motives from intentions because we have to ask: why is rescuing someone in order to torture them later a difference in intention rather than just a difference in motivation, while rescuing

36. Ibid., 219, footnote *.
37. Ibid.
38. Ibid.
39. Ibid.

someone in order to get one's name in the paper is a difference in motivation but not a difference in intention?

It is indeed true that Clara has a different intention to Beth: Clara intends to torture the man later, Beth does not. However, Beth has a different intention from Amy: Beth intends to get her name in the paper, Amy does not.[40] The reason that the Anti-Motivation View claims that both Amy and Beth do something supererogatory is that they *share* an intention: to save a man from drowning. So the question is: does Clara intend to save the man from drowning?

The answer is yes: Clara intends to save the man from drowning. As argued earlier, the Intention Requirement is needed to rule out those actions that are not intentional, that is, those actions that are accidental, unconscious or involuntary. Clara's act of pulling the man ought of the water in order that he does not die is none of these: it is deliberate. It is not even merely foreseen: it is an essential part of realising her ultimate goal as she cannot torture him later if he dies now. Thus, just as the Anti-Motivation View accepts that Beth intends to save him, despite having a further goal of getting her name in the paper, it ought to accept that Clara also intends to save him, despite the further goal of torturing him later.[41] Thus, they, like Amy, intend their action under a description—to save someone from drowning—in virtue of which it is good.

---

40. I assume here that we should see Beth as intending to get her name in the paper. For those who doubt this characterisation, see my discussion in §6 of how, even if she is taken to be *motivated* by getting her name in the paper, the two cases of Beth and Clara still ought to be treated in a similar fashion.

41. For those who are still sceptical we can strengthen the cases in two ways. First, we can note that Clara only intends to torture the man in question, not to torture him *to death*. Second, we can even suppose that the success of saving the man, or his ultimate survival, matters very little to Beth (maybe even less to her than to Clara): perhaps Beth is equally likely to reach her ultimate goal of getting her name in the paper whether or not he survives the night at a local hospital; or perhaps she is likely to get *more* press coverage if he suffers and eventually dies. I do not think this emphasise on her attitude towards his future would make a difference to the Anti-Motivation View's assessment of her act as supererogatory. However, it puts pressure on the idea that we should both characterise Clara's action as one of torturing but not of saving but also characterise Beth's action as one of saving and not of self-publicity.

Now, some may respond that, given Clara's ultimate aim, perhaps her action is more properly described as one of torturing than of saving. In response to this it should first be noted that Clara has so far only saved the man: she has the *intention* of torturing him later but has as yet not tortured him at all. In this sense, it would be strange to limit our description of her act to one of torturing and not at least *also* as one of saving, given that the saving has occurred and the torturing has not. Second, proponents of the Anti-Motivation View explicitly claimed that, despite the fact that Beth has the ultimate goal of getting her name in the paper, her action is still one of saving, not one of, say, self-publicity. Thus, on the Anti-Motivation View itself, the agent's ultimate goal cannot be the *sole* determinant of the correct description of the act.

The motivation for calling Clara's act one of torture is reasonable only if we suppose that we have to pick *one* description of her act. Her intention to torture the man later is perhaps more morally salient than her intention to save him from drowning. However, what the cases of Beth and Clara demonstrate is that agents can have *multiple intentions* with respect to the same action. In the cases described, the action is an *intended means* to an *intended end*.

The possibility of multiple intentions is problematic given that intentions are supposed to help us pick out the correct act description under which to morally assess the act in question. Multiple intentions pick out multiple descriptions, each of which could have different moral evaluations. And so the problem of multiple intentions is: which intentions are relevant for assessing whether an act is supererogatory?

The problem of multiple intentions is one that the Anti-Motivation View must confront. And how the Anti-Motivation View responds will shape the plausibility of the account: it turns out to be either much closer to the Praise-Based View than

previously thought or, alternatively, a much more radical account of supererogation than often supposed.

5 Two Responses

The Anti-Motivation View can respond to the issue of multiple intentions in two different ways. The first is to accept that only some of the agent's intentions are relevant to determining whether the action in question is supererogatory. The second is to accept that *everything* an agent intends is relevant to picking out the description under which to evaluate the act in question. In either case, I will show that intentions and motives are treated in a much more similar way than the Anti-Motivation View might originally have supposed. Let's take a closer look at each response.

Let's start with the consequences of selecting only some of the agent's intentions as relevant. The advocates of the Anti-Motivation View could reject the claim that intended ends make a difference to whether or not an action is supererogatory. This response allows them might stick to their guns regarding Beth: we ought to ignore any other intentions she might have beyond saving the man, such as getting her name in the paper, when assessing whether or not her action is supererogatory. However, this response also involves giving a different verdict to Mill in the case of Clara. We can ignore the impermissible intended end of her action and instead acknowledge only that she intends to save the man from drowning, just as Beth does.[42]

42. Note that just because someone intends a morally impermissible end does not entail that the means to that end is itself impermissible. That would entail that Clara is forbidden from saving the man at all, which cannot be correct (for more on this, see Scanlon, *Moral Dimensions: Permissibility, Meaning and Blame*. Note that nothing in this paper relies on Scanlon's claims concerning the role of intention in

Thus, on this position, the problem of multiple intentions is dealt with by supposing that an act is supererogatory so long as *one* of the agent's intentions is under a description in virtue of which the action is good. Thus, all but one of the intentions of an agent are treated like the Anti-Motivation View proposes to treat motives: they are excluded from making a difference to the description under which an act is assessed as supererogatory or not. In this way, intentions and motives are treated in a much more similar way than previously supposed.

The main problem with this response to the problem of multiple intentions is that it endorses an incredibly inclusive account of the supererogatory. As such, it has some potentially counterintuitive consequences. Not only would Clara's act of saving a man from drowning in order to torture him later be considered supererogatory, so too would acts like the following:

> **Discriminatory Giving**: I have two colleagues who have birthdays on the same day. I intend to give a present to one and not to the other in order to make the other feel excluded, disliked and discriminated against.

> **Passive Aggression**: On my way to meet you after we have had a small tiff, I help out someone who would otherwise be stranded. I intend to help them out in order to make you really anxious about where I am in payback for the argument we had early that day.

determining permissibility; nevertheless, his assessment (and explanation) of this particular case seems both correct and illuminating).

Confidence Trickster: I help an elderly neighbour who can't see very well fill in some necessary paperwork, with which he would otherwise have struggled. I intend to help him now in order that I will later have access to his bank accounts for the purpose of robbing him.

Betrayal of Trust: I intend to take my ex-girlfriend out for an expensive meal and listen to her work problems in order to weasel my way back into her life and continue my emotional and physical abuse of her.

Actions like these would be ruled out as supererogatory by the Praise-Based View. However, on an inclusive version of the Anti-Motivation View they would all count as supererogatory, because in each case the action is intended under a description that is good: giving a gift, helping a stranger, helping an elderly neighbour, and taking someone out for dinner, respectively.

Just as some might have resisted calling Clara's action one of saving, it might seem unnatural at first to consider these actions to be ones of gift-giving or helping, rather than acts of discrimination, passive aggression, deception and abuse. However, recall that this temptation only arises from the assumption that intentions are such that they pick out only one description of the act in question. The existence of multiple intentions reveals that often there multiple relevant descriptions of an act. And just as Clara's further intention of torturing the man later does not mean that she does not also intend to save him from drowning, my further (albeit nefarious) intentions do not mean I do not truly intend to give a present, help the stranger and my neighbour and treat my ex-girlfriend to dinner. Each of these actions is a deliberate part of my plan.

Also recall that I do not need to intend my action in a way that is self-consciously good: it is enough that I intend it under a description *in virtue of which* the action is good. The formulation of this requirement explicitly rejects the need for the intentions to be altruistic in content. Thus, my intention of helping the elderly neighbour to fill in paperwork means that one description of my act is 'helping', while my other intention—of gaining access to his bank accounts—means that another description of my act is 'deception'. But, on the response under consideration, so long as just *one* of the descriptions picked out is good, then the act is supererogatory: in each case, my act meets this minimal condition.

On such a response, the Anti-Motivation View and the Praise-Based View would no longer disagree only over cases like Beth. On this version, the two views are in fact much further apart than previously assumed. Such an inclusive version of the Anti-Motivation View would be a highly radical account of the supererogatory, which impacts its comparative plausibility and appeal.

An alternative response to the problem of multiple intentions is to accept that all intentions—including intended ends—can determine whether or not an act is supererogatory. This would allow the Anti-Motivation View to maintain that Clara's action is *not* supererogatory, because we would take into account her other intention: torturing the man later. It would also allow them to reject as supererogatory the examples of discrimination and deception detailed earlier.

However, this brings the Anti-Motivation View and the Praise-Based View much closer together than the Anti-Motivation View first supposed as they would then both agree that, if *any* of the things an agent intends is under a description that is morally bad, then their action is not supererogatory. Thus, supposing that Beth's intention of getting her name in the paper is selfish, then the two views would give the same verdict:

her act of saving the man from drowning is not supererogatory. The Praise-Based View deemed Beth's action not to be supererogatory because she was not praiseworthy; however, this was precisely because she intended something that is selfish. Both positions would therefore track the same reasons—and return the same verdict—in this case.

It is true, however, that a moral difference exists between Beth's action and Clara's: Beth's intended end is perhaps merely self-interested, while Clara's is morally abhorrent. So perhaps the Anti-Motivation View could claim the following: no act that is intended as a means to something morally prohibited can be supererogatory. In this way, perhaps Clara's action would be ruled out as supererogatory but Beth's not.

However, what makes a difference to whether the act is supererogatory on this response is not the agent's intentions nor their motivations: it is, rather, a fact about the nature of the *end* to which the action is a means. Both views are therefore sensitive to the same considerations but simply disagree about how bad one of the agent's intentions needs to be in order to disqualify it from being supererogatory. This is no longer about intentions being relevant and motives not, but rather about the bar an action must meet for it to be good enough (or perhaps 'pure' enough) to be considered supererogatory.

Thus, this response still brings the Anti-Motivation View closer to the Praise-Based View: while it looked originally like the two views gave different verdicts on some important cases, in fact they will not only agree on most cases, but their judgement as to whether or not an act is supererogatory will track exactly the same considerations.

The consequences of each of these responses demonstrates that Beth and Clara ought to be treated in a much more similar way than proponents of the Anti-Motivation View supposes.

6 A Further Critique of the Distinction

So far my discussion has been under the assumption that we should consider Beth as *intending* to get her name in the paper. This is, I think, natural given that she performs the action *so that* she will get her name in the paper. However, I now want to address those who feel an inclination to insist that this is truly a *motivation* and not an *intention*. If such a position can be maintained, some might feel that there is a sliver of hope of distinguishing Beth and Clara in the way original proposed by Mill. The consideration—and rebuttal—of this view reveals a more general critique of the different roles that the Anti-Motivation View assigns to intentions and motives in determining whether or not an action is supererogatory. As I show, a closer examination of Anscombe's position on the distinction between intentions and motivations reveals that, just as an agent's intentions can help determine the correct act description under which to evaluate an act, so can the agent's motives.

Mill distinguishes intentions and motivations in the following way. The intention of the agent is "what the agent *wills to do*" while the motive of the agent on the other hand is "the feeling which makes him will so to do."[43] In her analysis of intentions, Anscombe acknowledges that the different senses of intention and motive suggests a straightforward distinction in theory: "A man's intention is *what* he aims at or chooses; his motive is what determines the aim or choice."[44]

However, even on Anscombe's account, intentions and motives are much more similar than that Anti-Motivation View has traditionally acknowledged them to be. She argues that intentions—what someone aims at—are just one kind of motive: forward-

43. Mill, *The Collected Works of John Stuart Mill, Volume X - Essays on Ethics, Religion and Society*, 219, footnote *.
44. Anscombe, *Intention*, 18.

looking motives. She outlines two kinds of motives that are not intentions: "backwards-looking motives" such as revenge, gratitude, remorse, and pity and "motives-in-general" which ask someone to "see the action in this light."[45] Nevertheless, all three are different ways of answering the question 'why' about someone's action.

The idea that intentions are a species of motive is already a problem for the Anti-Motivation View needing as it does to draw a sharp distinction between intentions and motives. However, the Anti-Motivation View could accommodate the idea that intentions are a kind of motive so long as it is possible to distinguish motives-that-are-intentions and motives-that-are-not-also-intentions. However, the problem runs deeper as it is not always easy to tell them apart and this casts doubt on the claim that there is a clear distinction between them such that all intentions matter but no motives do.

Firstly, Anscombe notes that when someone is asked for a *motive*, they "might say 'I wanted to…', which would please [those who make the distinction between intentions and motives]; or 'I did it in order to…', which would not; and yet the meaning of the two phrases is here identical."[46] This is a problem for the proponents of the Anti-Motivation View, who want to argue that intentions are important for an act being supererogatory but motivations are not, especially as (as Anscombe notes) the interchangeability of intention and motive impacts the evaluation of an agent's actions: "When a man's motives are called good, this may be in no way distinct from calling his intentions good—e.g. he only wanted to make peace amongst his relations."[47]

45. Ibid., 21.
46. Ibid., 18.
47. Ibid. In light of these problems, Anscombe goes so far as saying that in terms of the relative importance of considering motives as opposed to intentions, she is "very glad not to be writing either ethics or literary criticism, to which this question belongs" (Ibid., 19.)!

Secondly, even if there is still hope of a clear distinction, it is implausible that all intentions matter but no motives do. To see why, we must begin by acknowledging that even if someone performs an action because of a motive rather than an intention, it does not mean that they do not perform that action *intentionally*. Suppose John kills Joshua. In answer to the question 'why', he might say that he did it so that he could inherit Joshua's fortune (a forward-looking motive and also a further intention), or because of revenge for past wrongs (a backwards-looking motive) or because of love of Joshua's wife (an interpretive 'see the action in this light' motive as her had no particular aim in mind). No matter which of three answers is the reason why he killed Joshua, it is still true that John *intended to kill* Joshua. Thus, whether or not someone performed an action because of a motive does not mean that they did not perform the action intentionally.

Of greater significance for our discussion, it seems implausible that an action could be transformed from being an un-supererogatory act to being a supererogatory one simply by exchanging a bad intention for a bad motive. Recall that, on the second response to the problem of multiple intentions, Clara's act of rescuing the drowning man is ruled out as supererogatory because she has a further intention of torturing him later. However, suppose that we find out that the real reason Clara saves him is because she knows that the man in question is dying from a long and extremely painful disease that cannot be alleviated even by suicide as he is Catholic. She saves him from revenge for perceived past slights, knowing that she has condemned him to a life of pain and suffering. Revenge is one of Anscombe's paradigmatic examples of a motive that is not also an intention. However, it strikes me as implausible that finding out that Clara has a backwards-looking motive rather than a forward-looking intention should change whether or not she has performed a supererogatory act. In both cases, she saves

someone who would otherwise have drowned at some risk to herself, and in both cases she is interested in the pain and suffering of the person that she rescues. These facts should be relevant in our assessment of either both cases or neither.

Thus, if intended ends are taken to play a role in determining the description under which an agent's actions should be evaluated, so too should their motives. Suppose that I give you a present out of the motive of friendship without any further intention (such as to bribe you).[48] While I might have a motive, rather than an intended end, I nevertheless intend to give you a present. Furthermore, it is because it comes from a motive of friendship that explains why one of the correct description under which to evaluate my action is 'an act of friendship' rather than, say, 'an act of bribery.'

If we rule out motives as part of what helps determine what act is performed, then we cannot pick out the relevant act descriptions under which to evaluate the act. It doesn't seem to be the case that whether or not an act is performed in order to achieve some other aim should make a difference to the kind of act it is. Thus, if we insist on ruling out Clara's action because of the further end that it serves and thus argue that her intended end makes a difference to the *kind* of act it is, then we must accept that motives can also determine the kind of act the agent performed.[49] Thus, the Anti-Motivation View moves closer and closer to the Praise-Based View.

---

[48] My thanks to Lucy Campbell for this example and for an illuminating discussion of Anscombe's position on intentions and motives.

49. We might even need to go so far as to say that we must take into account what an agent merely foresees as motivating side-effects might also make a difference to the moral quality of an act even if they aren't intended. Consider these two cases. Scuppered Competition: In order to help my friend be first in his class, I intend to distract his nearest competition by deliberately exploiting her weakness for late night movie marathons. I know that this will cause her to do very badly in the important test the next day, giving my friend an edge. Thin Walls: In order to help my friend be first in his class, I intend to stay up all night quizzing him and helping him study. Given the thin walls in his student halls, I know that this will cause his neighbour to be kept up leading her to do badly in the important test the day that I am

The issue is thus that we must take both motives and further intentions to be relevant to the moral evaluation of an act or we must ignore them both. Either way, we end up treating motives the same as intended ends. Thus, we cannot distinguish the role that intended ends and motives play in determining the content of the intentions we execute in action and thus we cannot distinguish the role of intentions and motives in determining the description of an act under which we evaluate that act.[50]

7 Conclusion

We sometimes talk as if there is a distinction between intentions and motives, as revealed in such pithy sentiments as: "If a rich man has an ugly daughter, he is concerned about her suitor's motives. But a poor man with a beautiful daughter is concerned about her suitor's intentions."[51] And it is upon this seeming difference that the Anti-Motivation View hopes to distinguish itself from the Praise-Based View of supererogation.

However, I have argued that, contrary to Mill's interpretation of the three drowning men, the Anti-Motivation View cannot claim that Amy and Beth's actions are supererogatory and Clara's are not. If proponents of the Anti-Motivation View accept that Clara's action *is* supererogatory, then they will have to propose an extremely inclusive and potentially counterintuitive account of supererogation. If, on the other hand, they reject that Clara's action is supererogatory then they must also reject that Beth's action is supererogatory or at least accept that they are sensitive to the very same

helping my friend prepare for. If we think that the latter case is also not supererogatory it must be because of the foreseen though unintended side-effects of my actions.

[50] Again, my thanks to Lucy Campbell for help in formulating the issue in these terms.

51. H. Gross, *A Theory of Criminal Justice* (New York: Oxford University Press, 1979), 111. My thanks to Leora Dahan Katz for bringing this quotation to my attention.

considerations that led the Praise-Based View to reject Beth's action as supererogatory. In which case, the difference between the Anti-Motivation View and the Praise-Based View is substantially reduced.

Thus, either the Anti-Motivation View, contrary to first impressions, is not nearly so different from the Praise-Based View, or it must propose a radical account of the supererogatory according to which even saving someone to torture them later is supererogatory.

Acknowledgements My thanks to David Heyd, Georgie Statham, Silvia Jonas, Olla Solomyak, Leora Dahan Katz, Sharon Berry, Elvira Di Bona, Moyra Tourlamain, and Lucy Campbell. I am also grateful to the participants at the Israeli Philosophical Association as well as the anonymous reviewers for this journal for their helpful feedback.