The Supererogatory, and How to Accommodate It

DALE DORSEY
University of Kansas

Many find it plausible to posit a category of *supererogatory* actions. But the supererogatory resists easy analysis. Traditionally, supererogatory actions are characterized as actions that are morally good, but not morally required; actions that go 'beyond' the call of our moral obligations. As I shall argue in this article, however, the traditional analysis can be accepted only by a view with troubling consequences concerning the structure of the moral point of view. I propose a different analysis that is extensionally correct, avoids the problems of the traditional view, and, incidentally, also defuses any objection to act-consequentialism, or any other first-order moral theory, on grounds that it cannot accommodate the supererogatory.

A traditionally noted feature of act-consequentialism is that it doesn't seem to leave room for the supererogatory. Trouble is, supererogatory acts seem to exist. Urmson writes:

> We may imagine a squad of soldiers to be practicing the throwing of live hand grenades; a grenade slips from the hand of one of them and rolls on the ground near the squad; one of them sacrifices his life by throwing himself on the grenade and protecting his comrades with his own body. [Doing so] is clearly an action having moral status. But if the soldier had not thrown himself on the grenade would he have failed in his duty? Though clearly he is superior in some way to his comrades, can we possibly say that they failed in their duty by not trying to be the one who sacrificed himself? If he had not done so, could anyone have said to him, 'You ought to have thrown yourself on that grenade'? Could a superior have decently ordered him to do it? The answer to all of these questions is plainly negative.[1]

Urmson's case appears to illustrate that, at least on occasion, individuals can act in ways that go beyond the call of duty. The soldier sacrificed himself in a way that clearly is morally good, but – *contra* act-consequentialism – is certainly not his *duty*.

So far, so plausible. But the category of the supererogatory resists easy analysis. Traditionally, supererogatory actions are characterized as actions that are morally good, but not morally required; actions that go 'beyond' the call of our moral obligations. As I shall argue in this article, however, the traditional analysis can be accepted only by a view with troubling consequences concerning the structure of the moral point of view. I propose a different analysis that is

[1] J. O. Urmson, 'Saints and Heroes', *Essays in Moral Philosophy*, ed. A. Melden (Seattle, 1958), pp. 198–216, at 202–3.

© Cambridge University Press 2013
Utilitas Vol. 25, No. 3, September 2013
doi:10.1017/S095382081200060X

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

extensionally correct, avoids the problems of the traditional view, and, incidentally, also defuses any objection to act-consequentialism, or any other first-order moral theory, on grounds that it cannot accommodate the supererogatory.

I. THE TRADITIONAL VIEW

Analyses of the supererogatory have generally assumed three focal points, the conjunction of which I shall call 'the traditional view'. Though some further limit actions that can properly be called 'supererogatory', rarely do any views not accept the basic principles captured in the following brief statement by Rawls:

> [Supererogatory acts] are acts of benevolence and mercy, of heroism and self-sacrifice. It is good to do these actions but it is not one's duty or obligation. Supererogatory acts are not required, though normally they would be were it not for the loss or risk involved for the agent himself.[2]

The first principle noted by Rawls can be captured as follows:

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

Permissible not Required holds that one is never morally required – but is always morally permitted – to perform a supererogatory action.

However, *Permissible not Required* is clearly insufficient to capture what it means for an action to go *beyond* the call of duty. One would not, for instance, describe the permissible, but not required, action of double-knotting my shoes as supererogatory. The difference between saving one's friends by jumping on a grenade and double-knotting one's shoes is that the former and not the latter appears to have a comparatively *positive* moral valence. The former and not the latter is:

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

Note that *Morally Good* is a comparative claim. It holds that φ-ing is morally better than some relevant baseline. Of course, this baseline is not just any old action; actions can be especially morally good as compared to, say, the bombing of Hiroshima. But this doesn't mean they're supererogatory. Rather, *Morally Good* holds that supererogatory actions are not simply permissible, but have a particularly positive

[2] John Rawls, *A Theory of Justice* (Cambridge, Mass., 1971), p. 117.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

moral status in comparison to actions that, as it were, are *merely* in accord with one's moral obligations.

Consider now the third feature of the traditional view, also noted by Rawls. Many hold that one essential feature of the supererogatory is that supererogatory actions are supererogatory in part *because* they involve some non-trivial sacrifice to the agent. Though this claim is controversially strong, a somewhat weaker claim is surely an important aspect of the traditional view:

> *But for Sacrifice*: A subset (*S*) of supererogatory actions would have been morally required but for the fact that they require non-trivial sacrifice on the part of the agent.[3]

Take Urmson's soldier. One of the essential features of his action that renders it supererogatory is that it is *sacrificial*. But were it the case that this soldier could have saved his friends *without* sacrifice, he certainly would have been morally required to do so. If I am in a position to donate half my yearly salary to Oxfam International, but only at significant cost to my own well-being, doing so is supererogatory. If my donations fail to affect my well-being, or affect it only trivially, making these donations is morally required.

A note on *But for Sacrifice*. *S* is a proper subset of all supererogatory actions. Some – like myself – hold that most, even all, supererogatory actions will involve some sacrifice on the part of the agent. Others believe that not all supererogatory actions would be morally required were it the case that there were no agential sacrifice.[4] In the remainder of the article, I will limit my discussion to cases that fall *within* S (or, at least, that seem to do so given my considered judgements). These cases generate serious problems with the traditional view, even if the relative size of *S* is very small.[5]

The traditional view might be supplemented by a number of additional principles concerning the structure of supererogatory action. For instance, some[6] might argue that the idea of the supererogatory should be limited to acts that are in some way beneficial, or are

[3] *But for Sacrifice* may not seem like an element in an analysis of the supererogatory, but rather a simple first-order truth of extension of this concept. I am neutral on these interpretations. Suffice it to say, any attempt to accommodate the traditional view must accommodate *But for Sacrifice*.

[4] See, for instance, Terry Horgan and Mark Timmons, 'Untying a Knot from the Inside Out: Reflections on the "Paradox" of Supererogation', *Social Philosophy and Policy* 27 (2010), pp. 29–63, esp. pp. 47–8. Thanks to Doug Portmore for calling my attention to this issue.

[5] See Horgan and Timmons, 'Untying', pp. 50–9.

[6] See, for instance, David Heyd, *Supererogation: Its Status in Ethical Theory* (Cambridge, 1982).

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

particularly praiseworthy. But *Permissible not Required*, *Morally Good*, and *But for Sacrifice* seem to be the beating heart of traditional analyses of supererogatory actions, whether further principles are required or not.

II. A CHALLENGE FOR THE TRADITIONAL VIEW[7]

James Dreier notices a problem for the traditional view:

> [The supererogatory] is puzzling. Morality, we are inclined to think, is a matter of what reasons one has *from the moral point of view*. When there is a supererogatory act available, it would be better for you to perform it. So surely you have a reason, from the moral point of view, to perform the act. You may have some reason not to perform it, but at least typically you will have no reason *from the moral point of view* to refrain from it (if you do have some such reason, then it will ordinarily be outweighed by the reason you have to perform, because by hypothesis it is better to perform). But now it is hard to see how it could be permissible, from the moral point of view, to refrain from doing something that you have an undefeated reason (from that very point of view) to do. Everything from the moral point of view speaks in favor of [performing a supererogatory act], and nothing at all speaks against it. In what sense is it 'all right,' 'permissible,' 'not wrong' to fail to act? There seems to be no sense at all.[8]

One way to state this challenge more precisely is to introduce a modicum of terminology. First, consider the distinction between a *supererogatory* action, and a *merely erogatory* action. Though both supererogatory and merely erogatory actions are permissible, supererogatory action goes 'beyond' one's duty. Merely erogatory action does not. Consider the following case. Imagine that you can react in one of three ways to a person down on her luck. You can assist her by going out of your way to buy her a nutritious meal. Second, you can offer her one dollar. Finally, you can do nothing. Assume that the third option is morally disallowed and that the first is supererogatory. One might say, plausibly, that the second is 'merely erogatory'; it counts as the fulfilment of one's moral obligations, but not in a way that is particularly morally special. The supererogatory action is morally special in comparison to *it*.

Second, consider the idea, mentioned by Dreier, of a *moral reason*. A moral reason to φ is a fact that, morally speaking, counts in favour of φ-ing. Moral reasons come in different strengths. In comparison to

[7] This problem is well known. See e.g. Horgan and Timmons, 'Untying', pp. 36–8. See also Douglas Portmore, 'Are Moral Reasons Morally Overriding?', *Ethical Theory and Moral Practice* 11 (2008), pp. 369–88, esp. pp. 378–81.

[8] James Dreier, 'Why Ethical Satisficing Makes Sense and Rational Satisficing Doesn't', *Satisficing and Maximizing: Moral Theorists on Practical Reason*, ed. M. Byron (Cambridge, 2004), pp. 131–54, at 148.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

some moral reason s, a moral reason r might support φ-ing in a weaker, stronger or equivalently strong way. Reasons can also add up to support actions in stronger or weaker ways: it could be that the collection of reasons in favour of φ-ing is sufficient to render it the case that there is a stronger balance of moral reasons in favour of φ-ing rather than some other act ψ, for which there is also moral reason.

Given the traditional view, if ψ-ing is merely erogatory, and φ-ing is supererogatory, φ-ing must be supported by moral reasons (or collections of moral reasons) that are stronger than the moral reasons that support ψ-ing. The reasoning for this runs as follows. If ψ-ing is both merely erogatory *and* supported by stronger moral reasons than φ-ing, it would appear that φ-ing is certainly not morally meritorious; one does not behave in a particularly morally good way if one behaves in a way for which there is weaker moral reason than the act that is merely erogatory. For the same reason, it cannot be the case that φ is supported by reasons that are equivalent in strength to ψ. This would entail that φ and ψ are of equivalent moral importance or quality. But this is incompatible with *Morally Good*. Hence it must be the case that supererogatory actions are supported by stronger moral reasons (or a stronger collection of moral reasons) than merely erogatory actions. With this terminology in mind, the puzzle is easy to see: given *Morally Good*, supererogatory actions will be supported by stronger moral reasons than merely erogatory actions. But given *Permissible not Required*, supererogatory actions cannot be morally required: merely erogatory actions remain permissible. But, in any collection of potential actions a person might perform, it seems right to say that this person ought to perform the action that is supported by the strongest balance of moral reasons. And hence either the supererogatory action will be required (violating *Permissible not Required*) or it will not be supported by stronger moral reasons (violating *Morally Good*).

Of course, there is an easy way out. The purported problem seems to assume that intra-moral rationality – put roughly, the relationship between moral reasons and moral requirements – is of a particular kind: one is morally required to perform the action for which there is strongest moral reason. Call this 'the standard view' of intra-moral rationality. If the standard view is correct, the traditional view of the supererogatory cannot be accommodated. But if we reject the claim that we are morally required to perform the action for which there is strongest moral reason (or strongest balance of reasons), we can say it is permissible to ψ even if there is stronger moral reason to φ. However, though this is clearly the right response for partisans of the traditional view, it is not enough simply to say that one can be permitted to refrain from performing actions for which there is strongest moral reason. To solve the puzzle one has to adopt an alternative account of the

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

relationship between moral reasons and moral requirements such that the peculiar character of supererogatory action can be accommodated.

The following two sections consider three potential alternatives. None, I argue, is successful. The first two (each a version of a satisficing approach to intra-moral rationality) cannot accommodate *But for Sacrifice*. The third can accommodate the principles of the traditional view, but ends up with very implausible results in so doing. I conclude that we must reject the traditional view in favour of an alternative analysis of the supererogatory.

III. SATISFICING

One alternative account of intra-moral rationality is a *satisficing* view. According to Dreier, 'Ethical satisficing is structurally similar to rational satisficing. Ethical satisficing theory says that it can be morally right to choose an alternative that is good enough ... even though there is a better alternative available.'[9] On a scheme of moral satisficing, one violates a moral obligation only if one acts in a way that is not morally good enough. The content of satisficing views will differ markedly; in this section, I consider the simplest form of satisficing, one that sets an absolute target (call this t), and holds that one need only perform actions that meet this target threshold. Call this 'simple satisficing'.

Simple satisficing accommodates *Permissible not Required* and *Morally Good*. If we accept that one is morally required only to perform an action that is at least as morally good as the target threshold, this can leave open the possibility that actions that are morally *better* than the threshold are permissible, but not morally required. Furthermore, moral satisficing can explain why supererogatory action is plausibly morally heroic or 'beyond' the call of one's duty: if one's duty is only to conform to the relevant threshold, any action that is morally better than the threshold is, plausibly, especially morally good in comparison to action that *barely* passes the threshold.

Despite its virtues, simple satisficing cannot accommodate the traditional view. To embrace moral satisficing, one must reject *But for Sacrifice*. Assuming that ψ-ing is 'morally good enough' (i.e. just barely passes the threshold), satisficing permits ψ-ing whether or not a morally *better* action involves any sacrifice at all. And hence there can be no supererogatory action that *would* have been required but for sacrifice, as it were.

But simple satisficing is not the only form of satisficing. Dreier advocates a form of intra-moral rationality that may be more promising.

[9] Dreier, 'Satisficing', p. 142.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

In considering the purportedly supererogatory action of helping to return a stranger's hat that has blown away in the wind, Dreier argues that there are, in fact, two moral points of view:

> To borrow from virtue theory, one point of view we can adopt is the point of view of the perfectly virtuous agent . . . or, less ambitiously, just the beneficent agent. From this perspective, there is everything to be said in favor of fetching the stranger's hat and nothing to be said against it. Failing to fetch the hat is falling short of perfection (in this dimension) and not permissible at all. But we can also adopt a less ambitious perspective – that of the just person (maybe 'dutiful' would be a better word). From the point of view of justice, there isn't anything to be said in favor of going to all that trouble to get a stranger's hat. If I do go to all the trouble, that doesn't make me more just. The suggestion is that judgments of wrongness are made from the point of view of this less demanding virtue, whereas judgments of what would be better or worse are made from the more ambitious point of view.[10]

Dreier's account is interesting, and much more can and should be said about it than I can or will say here. For Dreier's view, intra-moral rationality is a mix of two different moral points of view. The point of view of 'justice' determines the moral permissibility and impermissibility of actions. For Dreier, however, there is no reason of justice to fetch the stranger's hat. If so, refraining from so doing is permissible, not required. The second point of view, the point of view of 'beneficence', determines the relative goodness or badness of actions, but does not say anything about the permissibility or impermissibility of these actions. Reasons of beneficence rank-order actions, but that φ-ing is supported by greater reasons of beneficence can never count in favour of a requirement to φ. Putting these points of view together, one performs a supererogatory action if one performs an action that is better, as determined by the point of view of beneficence, than the merely erogatory action (which one might interpret as the morally worst action that is compatible with the requirements of justice).

But Dreier's view inherits the problems of simple satisficing. It also fails to accommodate *But for Sacrifice*. Take Urmson's case. Dreier, to claim successfully that saving one's comrades at a sacrifice of one's own life is supererogatory, must say that, from the point of view of justice, 'there isn't anything to be said in favour' of saving one's fellow soldiers. Of course, there is something to be said in favour of rescuing one's comrades from the point of view of beneficence, and this is what renders one's action supererogatory. But Dreier's view cannot accept the further claim that were there no sacrifice involved, one is morally required to save one's comrades. If 'there isn't anything to be said in favour' of so doing from the point of view of justice, and justice is

[10] Dreier, 'Satisficing', p. 149.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

the point of view that determines 'judgements of wrongness', it seems difficult to see how Dreier might deliver the verdict that it is wrong not to save one's comrades under any circumstances at all. It would seem that any reason to save one's friends is a reason of *beneficence* not justice, and hence this reason cannot influence the extent to which it would be wrong not to save one's friends when it is of no sacrifice.

Of course, Dreier claims that the domain of justice is the domain of 'right and wrong'. Hence we might be tempted to suggest that my requirement to save my comrades if it is of no sacrifice is a requirement of the point of view of justice. Doing so in this case *does* 'make me more just'. But one cannot make this claim and maintain the integrity of Dreier's proposal. To allow Dreier to claim that, e.g., saving one's comrades by sacrificing one's life is supererogatory, he must say that the reason to save one's comrades is a reason of beneficence: it is something that makes the act 'better', but not morally *required*. But to say that one is required to save one's comrades when there is no sacrifice involved would seem to collapse the distinction between reasons of beneficence and reasons of justice. Such a position seems to require reasons of betterness to play a role in determining one's moral obligations, after all. But if this is the case, it is difficult to see how such reasons wouldn't play a role in the case in which saving one's comrades *would* entail sacrifice. If so, we must now search for an alternative form of 'justice rationality' that would allow one to perform a suboptimal act from the point of view of justice when sacrifice is involved. But this just *is* the challenge for the traditional view.

The following point is critical. It could very well be that some supererogatory action would be supererogatory even if no sacrifice were involved. However, even if there is *one action* that falls within S, one and only one action of which it can rightly be said that without sacrifice it would be required, satisficing views fail.

IV. THE PORTMORE/RAZ VIEW

The failures of satisficing approaches trace to the same source: neither could allow that a lack of agential sacrifice would, of itself, render a previously supererogatory action morally required. To accommodate *But for Sacrifice*, then, one must allow that agential sacrifice (or lack thereof) can influence the content of moral obligations.

But the three tenets of the traditional view entail that agential sacrifice must play this role in a special, and perhaps non-straightforward, way. By *But for Sacrifice*, any supererogatory action φ, falling within S, must be supported by reasons that would be sufficient to require φ, were it not for the sacrifice to the agent involved. But, by *Permissible not Required*, φ cannot be required as it stands. Hence

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

the sacrifice involved in φ-ing must *itself* must have some sort of force in determining our moral requirements. But this force is either the product of *per se* moral reasons or not. If this force is the product of *per se* moral reasons, the sacrifice, which justifies a merely erogatory act ψ in comparison to φ, would itself have *moral significance* or *per se* moral weight, strong enough morally to justify in comparison to the reasons in favour of φ-ing. And if this justificatory strength is the result of moral factors themselves, it cannot be that φ-ing is morally special in comparison to ψ-ing; both are justified by significant moral concerns.[11] And if this is right, by *Morally Good*, φ-ing is not supererogatory. Hence the force of agential sacrifice to justify merely erogatory action (and render supererogatory action unrequired) must, at least in part, be the product of reasons that themselves lack *per se* moral weight. Call any view that accepts the possibility that the non-moral significance of, e.g., prudential sacrifice can influence the moral permissibility of actions a version of the 'Portmore/Raz view'.

Two versions of the Portmore/Raz view are offered, not coincidentally, by Douglas Portmore and Joseph Raz. According to Raz, the supererogatory is the outcome of a set of reasons known as 'exclusionary permissions'. Raz explains this idea as follows: 'The permission to refrain from performing an act [of] supererogation is an exclusionary permission, a permission not to act on certain reasons. An act is a supererogatory act only if it is an act which one ought to do on the balance of reasons and yet one is permitted not to act on the balance of reasons.'[12]

For Raz, an exclusionary permission is a not a *per se* moral reason (it is not part of the 'balance of [moral] reasons'), but is rather a 'second-order' reason, viz., a reason that allows a person to ignore the balance of moral reasons, or to refuse to grant them weight in one's moral deliberation. Surely there are strong moral reasons to save one's comrades by jumping on a grenade. But the soldier is permitted, given his exclusionary permission, to exclude these reasons in his moral deliberation. And hence, on this view, refraining from so doing is morally permissible given that the soldier maintains an exclusionary permission to ignore the moral reasons in question.

[11] See B. C. Postow, 'Supererogation Again', *Journal of Value Inquiry* 39 (2005), pp. 245–53. Incidentally, this very problem fells many of the views discussed by J. P. Vessel in 'Supererogation for Utilitarianism', *American Philosophical Quarterly* 47 (2010), pp. 199–219. Each view there discussed does allow for the possibility of a range of morally permissible acts, but cannot deliver the claim that the supererogatory act in question is particularly morally special in comparison to others, because *per se* moral factors justify both the supererogatory and the merely erogatory action.

[12] Joseph Raz, *Practical Reason and Norms*, 2nd edn. (Oxford, 1990), p. 94.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

This proposal appears to accommodate the traditional view: *Permissible not Required* is accommodated, given that (in the case of the soldier) jumping on the grenade is perfectly permissible: the balance of first-order moral reasons favours doing so. But, given this exclusionary permission, one can fail to take these reasons into account, rendering the supererogatory action unrequired. *Morally Good* is also accommodated. The balance of first-order moral reasons, after all, favours the supererogatory action. Though the soldier has 'exclusionary permission' to ignore the moral reasons in question, the fact that there are such reasons allows Raz to claim that following through on the sacrifice is morally superior: it is an act that conforms to the balance of first-order moral reasons one nevertheless has permission to exclude.

Raz's view can also be interpreted to accommodate *But for Sacrifice*. Its ability to do so depends on what *triggers* an exclusionary permission. One thought might be that the exclusionary permission is present in any case in which the 'first-order' reasons are reasons of beneficence. But this interpretation fails to accommodate *But for Sacrifice*, because this exclusionary permission would be present whether or not the beneficent action requires agential sacrifice, and hence Raz's view would be no less problematic than Dreier's. But one *could* say that the exclusionary permission is triggered only in cases in which the sacrifice to the agent in question is strong enough. If we do this, Raz's view successfully accommodates *But for Sacrifice*. For supererogatory actions falling within S, the exclusionary permission is *triggered* by the sacrifice involved; no sacrifice, no permission. On this interpretation, Raz's view does exactly what I argued must be done to accommodate the traditional view: it treats agential sacrifice as lacking 'first-order' moral importance, but nevertheless justifies merely erogatory action (by triggering an exclusionary permission to ignore the moral reasons that favour supererogatory action).

Portmore's version of the Portmore/Raz view uses a slightly different mechanism, but with a similar upshot. Following Joshua Gert,[13] Portmore exploits an important ambiguity in the way reasons operate. Note that the common understanding of reasons, i.e. as facts that 'count in favour of' particular actions, is vague.[14] Reasons may have *requiring* strength, i.e. may count in favour of a requirement to φ, or *justifying* strength, i.e. may count in favour of the permission to φ. According to Portmore, non-moral reasons (such as prudential reasons to avoid first-personal sacrifice) can possess moral *justifying* strength. For Portmore, non-moral reasons (including prudential reasons) can count in favour

[13] Joshua Gert, *Brute Rationality* (Cambridge, 2004).

[14] Thanks to an anonymous reviewer for asking me to address this ambiguity.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

of moral justification to ψ rather than φ, despite the fact that φ-ing is morally better (i.e. supported by stronger moral reasons) than ψ.[15] Non-moral reasons (including prudential reasons) are sufficient to morally justify merely erogatory action when, say, the prudential reason not to φ is of *sufficient* weight compared to the moral reason to perform the supererogatory action.

Portmore's view can accommodate *Permissible not Required*, because the reasons that tell in favour of merely erogatory actions have the power to justify merely erogatory action (and hence render supererogatory action unrequired), but do not have the power to *require* merely erogatory action, maintaining permission to perform supererogatory action. In addition, it can accommodate *Morally Good*. The moral justification allotted to a merely erogatory act is a result of *non-moral* reasons; this allows us to say that a supererogatory act is *morally* superior (since it is supported by a stronger balance of moral reasons), though not required (given that the merely erogatory action, though morally worse, is morally justified as a result of non-moral reasons). Further, it can accommodate *But for Sacrifice*. In cases like that of Urmson's soldier, were there no sacrifice, there would have been no relevant non-moral reasons that would justify a morally worse act, and hence only the previously supererogatory act (i.e. saving one's comrades) would be permissible.

The Portmore/Raz view, in either iteration, can accommodate the traditional view. Indeed, as I argued above, *only* the Portmore/Raz view can adequately accommodate the traditional view. What remains is to determine whether the Portmore/Raz view is plausible. I claim that it is not.

V. THE BOUNDS OF MORAL JUSTIFICATION

The best way to see this is to compare two cases (for the purposes of brevity, I focus on Raz's statement of the view; the problems are the same for any version of the Portmore/Raz view). The first is:

> *Gus*: Gus finds himself the recipient of an inheritance from a wealthy relative. This inheritance will allow Gus to buy a new car, which Gus desires to do, and which will allow Gus to see much more of his significant other, who lives in a distant town. Alternatively, Gus could donate his inheritance to Oxfam International, which would save ten people from death.

On reflection, it is plausible to say that Gus's donation would be supererogatory. Not only this, it seems plausible to say that were there

[15] Portmore, 'Override', pp. 372, 375 n. 12.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

no cost involved in saving ten people from death, Gus would surely be morally required to do so. Should Gus fail to save these ten from death at no cost to himself, this is surely grounds for moral complaint. Hence we should accept:

1. Gus's potential donation would be supererogatory, falling within *S*.

So far so good. But before I introduce the second case, let's shift gears slightly. There is surely some reasonably serious harm to which we may morally permissibly subject an individual for the purposes of saving some number of people (perhaps large) from death. Moral theories will differ on the *seriousness of harm-to-number of people saved* ratio; act-consequentialism, for instance, will claim that one person harmed will justify one person saved, for any harm short of death. Others will hold that there are agent-centred restrictions against harm, and that any reason to harm can only be outweighed by a larger amount of good, a larger number of individuals saved. For any plausible moral theory, however, there is some number of people (n) we could save from death that would morally justify, say, torturing one person, even leaving other things equal (e.g. that the agent is not among those who would be saved). If so, the moral reason to save n from death is of (at least) equivalent strength to the moral reason not to torture. Further, with each harm of decreasing seriousness, the moral reason not to harm someone in that less serious way is of equivalent strength to the reason to save fewer individuals. If we are licensed to torture an arbitrary person (call him 'Jerry') to save, say, 1,000 people from death, surely we are morally justified in harming Jerry in a less significant way to save some lesser number of people. With all this in mind, the following seems plausible:

2. Other things (including non-moral reasons) being equal, one is morally justified in beating Jerry up to save ten from death.

However, if we accept (2), a problematic result arises. Consider now the second case:

Stan: Stan knows that Jerry has just inherited a substantial amount of money from a wealthy relative. Were Stan to intimidate Jerry into giving him the money as a result of beating Jerry up, this would be a prudential benefit to Stan, given that this would allow him to buy a new car, from which he will derive pleasure, and which will allow Stan to see much more of his significant other, who lives in a distant town. (Assume that Stan would avoid punishment.)

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

If we accept (1) and (2), and the account of exclusionary permissions that will successfully accommodate the traditional view, it would seem that Stan is perfectly justified – *morally justified*, mind you – in beating up Jerry to buy a new car. Here's why. Gus's sacrifice triggers an exclusionary permission to refrain from saving ten. But Stan's potential sacrifice (i.e. not getting his new car) is *identical* to Gus's. And so one would assume that this sacrifice would *also* generate an exclusionary permission to beat up Jerry.

The obvious response is to hold that the relevant exclusionary permission is triggered only in Gus's case, not Stan's case. But, given (2), this is not tenable. As we have already seen, it is at least as morally important (that is, as important from the perspective of first-order moral reasons) to save ten as it is not to beat up Jerry. And so one would expect that if an exclusionary permission allows one to exclude the moral reasons in favour of saving ten in Gus's case, it must allow Stan to exclude the moral reasons not to beat up Jerry. Of course, there *are* differences between the two cases. Importantly, Gus faces moral reasons of beneficence; Stan faces moral reasons of non-maleficence. But this, in itself, cannot make a difference to the possibility of exclusionary permission unless that difference is reflected in the first-order moral significance of the reasons in question. And given (2), it is not: saving ten is at least as morally important as refraining from beating up Jerry.[16] The problem is identical for Portmore's view. Given the relative moral significance (as laid out in (2)) of saving and refraining to harm, any statement of the Portmore/Raz view will have the result that Stan is morally justified in beating up his neighbour given the moral justificatory significance of agential sacrifice. If agential sacrifice (by whatever mechanism) morally justifies pursuit of one's own interests against an action supported by moral reasons of strength *s*, it should also morally justify pursuit of one's interest against any other action supported by reasons of strength *s*-or-weaker.

[16] One could, perhaps, adopt a form of particularism and claim that further features of the case (such as the fact that the sacrifice-creating act is an instance of beneficence rather than an instance of non-maleficence) might defeat or disable the exclusionary permission (or, in Portmore's language, justifying reason). (See, for instance, Jonathan Dancy, *Ethics without Principles* (Oxford, 2004), ch. 3.) I don't have any *per se* beef with particularism. But such a move seems unavailable in this case. It would appear that all of the relevant facts of the case are accounted for in the favouring/disfavouring considerations, rather than in the enabling/disabling conditions of the case. In particular, the only relevant potential disabler of the exclusionary permission, in Stan's case, is the fact that to gain this prudential advantage, he would have to act in a maleficent way towards Jerry. But this is clearly a disfavourer – which is itself not strong enough to override justification to save ten, and which in turn is not strong enough to override justification to follow through on one's prudential interest in a new car – not a disabler.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

Care is required here. The mere fact that Stan could be morally justified in beating up Jerry to buy a new car shouldn't by itself be regarded as a fatal result of the traditional view. After all, many views – including act-consequentialism, for instance – will hold that such an act can be morally justified, especially if Stan's well-being is furthered by a new car more than Jerry is burdened by the beating. But the problem for the Portmore/Raz view is not this verdict *per se*, but rather the fact that Stan can be morally justified in beating up Jerry even if his interests are, comparatively, of insignificant moral weight (given that the justificatory power agential sacrifice, according to the Portmore/Raz view, outstrips its *per se non-moral* significance). *This* verdict is simply incredible, and must be rejected. If Stan is justified, this justification cannot be *moral* justification.

One important implication of the above argument is that it doesn't matter in the slightest what the relevant moral reasons *are*, or what moral factors might favour refusing to beat up Jerry versus what factors favour saving ten from death. One could, in principle, declare (in Kantian fashion) that the reason not to beat Jerry up is derived from an interest in not treating Jerry as a mere means, or that this reason is derived from Jerry's moral *rights*, or that Stan is under some particularly morally relevant *relationship* to Jerry, or is *accountable* to Jerry, in a way that tells against beating him up. Furthermore, one could assume that the reason to save ten is provided by their overall aggregate welfare, respect for persons, or any other reason. Because the reason – *whatever it is* – not to beat up Jerry is not stronger than the reason to save ten (in so far as one is allowed to beat up Jerry to save ten), and because the relevant agential sacrifice is significant enough to justify refusing to save ten (as seen in Gus's case), it must *also* be significant enough to justify beating up Jerry. To say otherwise is to accept a logic of exclusionary permission that seems utterly arbitrary.

The partisan of the traditional view could deny either (1) or (2). But (1) seems obvious; indeed, Gus's potential donation seems a paradigmatic instance of the supererogatory; furthermore, if we can save such individuals from death at no cost to ourselves and fail to do so, those whom we failed to save surely could complain on moral grounds. This leaves (2). One might complain that I have no genuine reason to believe (2); after all, that we are morally justified in beating up one to save ten, rather than, say, eleven, seems arbitrary. But the specific number identified by (2) is neither here nor there; the number can change without changing the seriousness of the problem. Imagine that Gus could save 100 people from death as a result of his Oxfam donation. Many would say that even under these conditions, Gus's donation remains supererogatory; his prudential reason is strong enough morally to justify his refusal to save 100 from death. But to deny

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

that one could be morally justified in beating someone up to save 100 people from death is positively absurd. Hence Stan's prudential reason, which is identical to Gus's, must also have the power morally to justify beating up Jerry, in so far as refraining from beating up Jerry is not morally more important than saving 100. One might put this in general terms: there is some number ('*n*') of individuals Gus could save from death at the cost of his new car and new life with his significant other, for which an exclusionary permission is no longer applicable. For the Portmore/Raz view to survive, we must say that we are unjustified in harming someone like Jerry to save *any* sub-*n* number of people from death. But this claim is worse than dubious.

The Portmore/Raz view accommodates the traditional view. I have argued, however, that given straightforward assumptions, and straightforward reasoning, this view implausibly expands the bounds of moral justification. This is not to say, of course, that there is no way to avoid this conclusion. But the problem here seems robust enough to warrant the search for an alternative to the traditional view. In what follows, I offer just such an alternative that is (a) satisfying in itself and (b) keeps the boundaries of moral justification right where we want them.

VI. THE SUPEREROGATORY

To begin, consider the notion of a moral requirement. Moral requirements are, well, just that: if I fail to conform to a moral requirement, this entails that I will have behaved immorally, or in a morally unjustified way. But there are many different sorts of requirements – not just moral – that I face. I face legal requirements, prudential requirements, requirements of etiquette, requirements of my neighbourhood association. Sometimes these requirements will conflict. But in cases of conflict, it seems natural to ask ourselves what we ought to do *really*, or *all-things-considered*. More generally, in the case of conflicting requirements, how should I live? For the sake of brevity, I will refer to this 'all-things-considered' requirement, which is distinct from, e.g., moral, legal or prudential requirements, as the 'rational' requirement, or rational 'ought'.

I take this conceptual territory to be familiar. Furthermore, we have an intuitive grasp on some of the answers to the questions asked above: most would agree that it is more important to conform to a moral norm than a norm of one's neighbourhood association. This is reflective of a standard assumption, viz. that morality is the most important domain, at least in determining how I all-things-considered ought to act. In particular, a standard analysis of the relationship between rational and moral requirements holds that the requirements of morality are

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

rationally *supreme*: if one is morally required to φ, one is thereby *rationally* required to φ.

But just for the sake of argument, let's say that morality is *not* supreme. Let's say that, in some cases, immorality can be be rationally permitted.[17] If we do this, it seems to me a natural – and alternative – way to understand the supererogatory emerges. Take the following case.

> *Rose*: Rose is a retiree with a substantial pension, and lives comfortably. Rose could get by with less, but this would require her to give up some things she enjoys doing. Assume now that Rose is morally required to assist others with her resources rather than spending her resources on herself. Assume also that Rose is not rationally required to assist those whom she could assist. Imagine now that Rose dedicates substantial time and money to a local family, themselves down on their luck. Rose provides for their food, lodging, and child care, which is burdensome, and leaves her unable to live the life she would otherwise want to.

How would we describe Rose's action? *Ex hypothesi*, Rose behaves in accordance with her moral obligations. But also, *ex hypothesi*, Rose is not rationally required to do so. One might, of course, simply describe Rose's action as an instance of action that conforms to a moral requirement. But that doesn't seem to say it all; given our assumptions, Rose's action is not *just* morally required, Rose's action is morally required in a way that isn't *required of her*. Though it is morally required, it is – one could correctly say – 'beyond the call of duty'.

I want to be very clear about what I take Rose's case to establish.[18] I won't assert that her case, by itself, makes for an argument that one can, as a matter of practical rationality, behave immorally. Perhaps the assumptions I've described in the case are plausible; perhaps they are not. What matters at this point in the argument, though, is not the plausibility of these assumptions *per se*, but rather the way in which we might be tempted to describe Rose's actions *under* these assumptions. If we believe the case as described, i.e. that Rose is, in fact, morally required to help but rationally permitted not to help, we would *in fact* reasonably describe Rose's helping as an instance of the supererogatory: an action that is morally better than *Rose's duty*. It is surely, as Rawls says, 'an act of benevolence and mercy, of heroism and self-sacrifice'.

[17] See Roger Crisp, *Reasons and the Good* (Oxford, 2007), ch. 1; Philippa Foot, 'Morality as a System of Hypothetical Imperatives', *Virtues and Vices* (Oxford, 1978), pp. 1–18.

[18] Thanks to an anonymous reviewer.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

Reflection on Rose's case seems to offer an alternative to the traditional view. Rather than holding that supererogatory actions are morally good but not morally required, one might say that, very roughly, supererogatory actions are morally good, but not *rationally* required.

This proposal requires rethinking of each of the tenets of the traditional view. First, *Permissible not Required*, on this view, becomes:

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

Morally Good becomes:

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

But for Sacrifice becomes:

> *But for Sacrifice II*: A subset (S) of supererogatory actions would have been rationally required but for the fact that they require non-trivial sacrifice on the part of the agent.

Though this proposal, like the traditional view, permits of a number of potential additions and supplementary principles, these theses are the heart of what I call 'the anti-rationalist view'.

The anti-rationalist view is attractive when compared to the traditional view. The key advantage is that the anti-rationalist view has no need of an alternative account of intra-moral rationality. The anti-rationalist view is perfectly free to say that one is morally required to conform to the strongest balance of moral reasons. This is an advantage for two reasons. First, this view is prima facie plausible (as noted by Dreier). If we can *both* accommodate the supererogatory *and* accept the standard account of intra-moral rationality, this is a comparative boon. But even aside from its prima facie plausibility, the ability to avoid an alternative account of intra-moral rationality saves the anti-rationalist view from the implausible results of the Portmore/Raz view. Because those wishing to accommodate the traditional view must accept some version or other of the Portmore/Raz view, any problems with the Portmore/Raz view bleed into the traditional view. And problems there are: the Portmore/Raz view implausibly expands the boundaries of moral justification (as explored in section V). But because the anti-rationalist view can accept a standard account of intra-moral rationality, this problem is avoided.

One immediate objection should be discussed. The anti-rationalist view is compatible with a standard account of intra-moral rationality, i.e. with the view that one is morally required to perform the action

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

for which there is strongest moral reason. But, on the anti-rationalist view, supererogatory actions are those which are morally better than is rationally required. But given that one is morally required to perform the morally best actions, this entails that sometimes morally required action will be supererogatory. This might be thought a non-starter.[19]

Of course, I do allow that morally required action can, on occasion, be supererogatory. But this is a feature of my analysis, not a defect. It *would* be a defect if, conceptually speaking, supererogatory action cannot be morally required. But this is not the case. The *concept* of the supererogatory for which we seek an illuminating analysis is expressed by, e.g., the thought that some actions go 'beyond' one's duty. The traditional view offers a conception of this concept: that supererogatory actions are morally good, but not morally required. But the traditional view is not conceptually true; it is a theory of the supererogatory that can (and, I argue, should) be denied. Thus it is no objection to the anti-rationalist view that the anti-rationalist view (sometimes) treats morally required action as supererogatory.

Furthermore, the anti-rationalist view is faithful to the concept of the supererogatory. Take Urmson's soldier. Urmson appears to indicate that jumping on a live grenade to save others is morally superior to refraining from so doing, but that it is not a feature of any individual's 'duty'. But one has a *duty* to perform only those actions one is *rationally required* to perform. If conforming to a moral requirement is not rationally required, one has no duty to conform to such a requirement. Hence it is perfectly acceptable to say of this case that one has no duty to conform, and also to say that, e.g., we could not 'possibly say that they failed in their duty by not trying to be the one who sacrificed himself', or that 'If he had not done so' we would not have 'said to him "You ought to have thrown yourself on that grenade" '. Furthermore, because he had sufficient rational permission not to jump on the grenade, no 'superior' could 'have decently ordered him to do it' (given the strength of the prudential reasons against doing it).

I do not wish to gainsay the popular view that supererogatory action cannot be morally required. Nevertheless, it seems to me that this intuition is misleading. It sounds right to say that supererogatory actions cannot be morally required because we have a tendency to identify our *duty* with our *moral duty*: we have a tendency to believe that moral requirements are rationally supreme. As noted here, this is a common presumption in our general thought about the relationship between practical rationality and moral requirements. Under these

[19] See, for instance, Dreier, 'Satisficing', p. 149.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

conditions, supererogatory acts cannot be morally required. But if we reject this connection, we are perfectly licensed to say that *sometimes moral requirements go beyond our duty*. Sometimes we have no duty to conform to moral requirements any more than we have a duty to conform to, e.g., legal requirements or requirements of etiquette. If so, there is no pressure to declare that supererogatory actions cannot be morally required.

VII. HOW TO ACCOMMODATE THE SUPEROGATORY

My proposal is to treat supererogatory actions not as morally better than is morally required, but as morally better than is rationally required. But there are two potential ways this analysis might go wrong. First, it could be that the anti-rationalist view simply doesn't work; it might be that there is no account of all-things-considered reasons that could accommodate the anti-rationalist view. Second, it could be that, even if there is such an account of the rational 'ought', this proposal, like the Portmore/Raz view, is just implausible. Because the second objection awaits a response to the first, I discuss the first objection here, and leave the second until section VIII.

Is the anti-rationalist view possible? Obviously, for this view to work we must be anti-rationalists about morality; we must believe that one can be all-things-considered justified in refusing to perform morally required action. However, anti-rationalism comes in many shapes and sizes, many of which clearly cannot accommodate the existence of supererogatory actions. To see this, consider the various ways one might construe the relationship between moral and rational requirements.[20] One common thing to say about this relationship runs as follows:

> *Authority*: If x has moral reason to φ at t, x has a practical reason to φ at t.

Authority holds that morality generates practical reasons: that φ-ing is supported by moral reasons entails that there is some practical reason to conform to it. But *Authority* is very weak. *Authority* is compatible with the claim that moral considerations are trumped by other considerations, such as prudential considerations. However, one must obviously accept *Authority* to accommodate the existence of supererogatory actions on the anti-rationalist view. Without doing so, any practical reason whatsoever not to φ renders φ-ing practically

[20] A helpful introduction to the relationship between morality and practical reasons is offered by David Brink, 'Kantian Rationalism: Inescapability, Authority, and Supremacy', *Ethics and Practical Reason*, ed. G. Cullity and B. Gaut (Oxford, 1997), pp. 255–92.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

irrational even if φ-ing is morally required. This would be tantamount to the rejection of *Permissible not Required II*.

However, to accommodate the supererogatory, one must deny a stronger association between morality and rationality. One must deny:

> *Supremacy*: If x has stronger moral reason to φ at t rather than to ψ at t, it is irrational for x to ψ rather than φ at t.

Supremacy holds that moral reasons are rationally overriding: whenever morality holds that one ought to φ rather than ψ, it is all-things-considered irrational to ψ rather than φ. Given the standard account of intra-moral rationality, *Supremacy* is equivalent to:

> *Supremacy**: If x is morally required to φ at t rather than to ψ at t, it is irrational for x to ψ rather than φ at t.

The anti-rationalist view cannot accept *Supremacy* (and hence must also reject *Supremacy**, given its embrace of a standard account of intra-moral rationality). To accept the existence of supererogatory actions (on the anti-rationalist's analysis), it must be the case that one occasionally has rational permission to behave in a morally suboptimal way, which *Supremacy* denies. On this view immorality – failure to conform to one's moral obligations – can, on occasion, be rationally justified given the possibility of other sorts of reasons, including prudential, aesthetic, legal, etc., reasons. This is not to say that conformity to moral requirements will never be rationally required. But, one might say, in cases in which non-moral (including prudential) considerations are *strong enough*, one is rationally permitted not to conform to the strongest balance of moral reasons.

But the denial of *Supremacy* and the acceptance of *Authority* does not guarantee that supererogatory actions, like Rose's, will be rationally *permitted*. It must be that people are rationally *allowed* to perform actions that are morally better than those that are rationally required. Hence, the anti-rationalist view must accept:

> *Permission*: If, for x at t, a particular act ψ is rationally permitted, and, for x at t, φ-ing is supported by stronger moral reasons than ψ-ing, φ-ing is rationally permitted for x at t.

Permission holds that all action that is morally better than is rationally permissible is rationally permitted. To accommodate the anti-rationalist view, one must accept *Permission* but deny *Supremacy*(*).[21]

[21] Though there may be many additional possibilities, my preferred way to guarantee *Permission* in light of a denial of *Supremacy* would be to adapt Gert's distinction between rational requiring strength and rational justifying strength. One can accept *Permission*

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

(One further point. I noted in section I that the comparative size of S relative to the set of all supererogatory actions is controversial. Horgan and Timmons, for instance, hold that at least some supererogatory actions are supported only by 'favouring' rather than 'requiring' reasons, and hence that one needn't be required to perform them even if they involve no sacrifice (such as taking a stranger on a fun outing). But my view seems to treat moral reasons as always of rational requiring strength, and if so S is coextensive with the set of all supererogatory actions. However, my view is flexible on this point. To capture the possibility of a narrower S, one might add a clause to *Authority*, holding that any moral favouring reason (if, in fact, such things exist) is also a *pro tanto* rational favouring reason. If this is right, one might then amend *Permission* to claim that if a particular act φ is morally favoured with respect to a rationally permitted act ψ, φ-ing is rationally permitted (though merely morally favoured acts would not be required). This would hold that actions supported by favouring reasons are permissible, not required; they are morally better than is rationally required, and hence they are supererogatory without being within the scope of S.)

This view can accommodate the anti-rationalist view successfully. Take Rose. If we accept the standard account of intra-moral rationality, we should say that Rose (given the strongest balance of moral reasons) is morally required to assist the family. But, if we deny *Supremacy* and accept *Permission*, we have the power to accommodate the suggestion that Rose's morally required action is, in fact, supererogatory. Rose has prudential reason to avoid assisting this family. Thus though assisting the family is morally good, indeed *morally required*, it need not be rationally required in light of Rose's sacrifice in so doing, and hence Rose is rationally *permitted* not to assist the family. But given that

if one accepts that the prudential reason to ψ can, at best, rationally justify, but cannot rationally require ψ-ing. In other words, non-moral reasons lack rational requiring strength, but maintain rational justifying strength. On this view, one will have rational permission to perform a morally suboptimal act if the prudential reason to do so is of sufficient comparative weight to the moral reason to avoid the morally suboptimal act. *Supremacy*, on this view, fails: I can be rationally justified in acting in a morally suboptimal way (depending, of course, on the weight of the non-moral reasons involved). Nevertheless, because non-moral reasons cannot require a person to behave in a morally suboptimal way, acting in a morally better way is always permitted, satisfying *Permission*.

One complication: sometimes non-moral reasons can rationally require. If, for instance, morality is indifferent between φ-ing and ψ-ing and one has stronger prudential reason to ψ, some might hold that it is irrational (given this prudential reason) to φ. Such a verdict, however, does not require a radical revision. One might say that though non-moral reasons have rational requiring strength, the rational requiring strength of moral reasons lexically dominates, or trumps, the rational requiring strength of non-moral reasons. This entails that when morality is not indifferent between φ and ψ, non-moral reasons cannot require one or the other. But in a case of moral indifference, given that there is equal moral rational requiring reason to φ rather than ψ, the comparatively insignificant non-moral rational requiring reason can tip the balance.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

assisting the family is morally better than is rationally required, Rose has rational permission to assist the family as well (given *Permission*). Hence assisting the family, for Rose, is supererogatory. The same applies in the case of Urmson's soldier. One might say that, given the lives at stake, there is surely greater moral reason for the soldier to throw himself on the grenade than to refrain. But given that this action, quite literally, is an instance of *self-sacrifice*, we should expect that the denial of *Supremacy* entails that, at least in this case, the soldier is rationally permitted not to hurl himself on the grenade.

If we accept *Permission* and reject *Supremacy*, we can accept that a supererogatory action is morally better than the merely *rationally erogatory* action (satisfying *Morally Good II*). This view can also accept that supererogatory acts are rationally permitted, but not required (satisfying *Permissible not Required II*). And this view can also accept that, were it not for the required sacrifice – that is, the prudential reasons against – supererogatory actions would be rationally required; without prudential sacrifice to rationally justify a morally suboptimal act, one is rationally required to perform the action for which there is overriding moral reason (satisfying *But for Sacrifice II*). To accept *Permission* but deny *Supremacy* in the way just illustrated is necessary and sufficient to accommodate the existence of supererogatory actions on the anti-rationalist view.

VIII. OBJECTIONS

So the anti-rationalist view is possible to accommodate. But that doesn't mean that it's plausible. I respond to three potential reasons to reject it here. First, I argue that the anti-rationalist view does not require an implausible first-order account of supererogatory actions. Second, I argue that though some might object to the denial of *Supremacy**, one should not consider the ability to accommodate *Supremacy** an advantage of the traditional view. Third, some might argue that this form of anti-rationalism view faces problems – just as the Portmore/Raz view did – with *Stan*. I argue that this is not so.

VIII.1. Practical reason and the supererogatory

The anti-rationalist view holds that any action that is morally better than is rationally required is eligible for the epithet 'supererogatory'.[22] But this view might generate a range of very implausible verdicts depending on the first-order theory of all-things-considered practical reason one accepts. Consider, e.g., Sidgwick's classic *dualism* of

[22] Subject to any further conditions one might wish to place on the category of the supererogatory, explored in section VIII.3.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

practical reason.[23] This view (or, at any rate, my interpretation) holds that one is always rationally permitted to behave in accordance with the demands of either prudence or morality. On this view, because no acts of any moral quality at all are rationally required, *any* act that even has the barest moral worth in comparison to advancing one's self-interest is supererogatory. But if this is right, the anti-rationalist view will implausibly extend the category of the supererogatory.

Of course, some substantive theories of rationality, when combined with the anti-rationalist view, may result in an implausible set of supererogatory actions. But this is no objection. If we accept the anti-rationalist view, our considered judgements about which actions are supererogatory are fair game when it comes to evaluating theories of rationality, as they have been all along in evaluating theories of morality (and intra-moral rationality) under the traditional view. If so, there is no reason to believe that the true theory of our rational requirements will offer an implausibly inflated or deflated set of supererogatory actions. If the dualist view is committed to a range of supererogatory actions that stretches considered judgement, we should reject that view. This move is no more problematic than the traditional rejection of, e.g., act-consequentialism on the basis of the traditional view.

Furthermore, note that the anti-rationalist view can accept that *exactly* the same actions that the traditional view declares are supererogatory are, in fact, supererogatory. When we seek to determine whether any action is supererogatory (whatever the analysis) we ask whether that action is among those that make up 'how we should live'. If φ is not part of 'how we should live', but is perhaps morally *better* than we can permissibly live, any view that accepts *Supremacy* will interpret this as an instance of action that is morally better than is morally required, and hence that φ-ing is supererogatory. But a view that rejects *Supremacy* will treat 'how I should live' not as capturing my moral requirements, but as capturing my *rational* requirements. Hence this view will suggest that φ is not necessarily morally better than is morally required, but is morally better than is *rationally* required, and is therefore supererogatory.[24] The anti-rationalist view is sensitive to precisely the same considered judgements as the traditional view, and hence needn't deviate from any verdicts that the traditional view can coherently accept.

[23] Henry Sidgwick, *The Methods of Ethics*, 7th edn. (Indianapolis, 1981 [1907]), pp. 507–9.

[24] For a more in-depth argument to this effect, see Dale Dorsey, 'Weak Anti-Rationalism and the Demands of Morality', *Noûs* 46 (2012), pp. 1–23.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

VIII.2. Denying Supremacy[*][25]

The anti-rationalist view is coherent and extensionally correct. But it is, after all, an *anti-rationalist* view. On my proposal, to accommodate the existence of the supererogatory, we must be *anti-rationalists*. This might be in principle implausible, as well as implausible in application. After all, is it plausible to say that Rose is morally required to assist the family in question, but rationally permitted not to do so?

I don't have the space to defend moral anti-rationalism in every nook and cranny; I have tried to do so elsewhere.[26] But as a general point, I think there are two reasons why we should temper our scepticism about the form of anti-rationalism at play here. First, to adopt *Supremacy*[*] is to adopt a rather mild form of anti-rationalism about the moral point of view. Recall that to reject *Supremacy*[*] one need not reject *Permission*, or claim that non-moral reasons will be decisive against reasons to conform to moral requirements. Rather, all one need believe is that moral agents, at least on occasion, will have rational permission to perform morally suboptimal acts. My view can accept the claim that conforming to morally decisive actions will often be rationally decisive, depending on the relative strength of the moral and non-moral reasons involved. If it is morally required of me not to harm someone else for a mild chuckle, it is perfectly compatible with my view to say that refraining from such harm is rationally required: the prudential reason in favour of harming someone, in this case, is too weak in comparison to the moral reason against to rationally justify such harm.

Of course, *Supremacy*[*] is, to some, very plausible. But it seems to me that the ability to accommodate it comes to very little *qua* advantage. The traditional view grants no greater rational authority to moral considerations than the anti-rationalist view. In fact, as defined here, both views must deny *Supremacy*. The traditional view suggests that though moral requirements are rationally overriding, moral *reasons* themselves do not fully determine our rational obligations, and hence it is not the case that moral reasons have overriding practical authority as defined by *Supremacy*. Moral reasons are (on Raz's view) *justly ignored* in determining *how one ought to live*. Hence the extent to which one's rational requirements are determined by moral reasons on the anti-rationalist view is precisely equivalent to the extent to which moral reasons determine one's rational requirements on the traditional view, when combined with *Supremacy*[*]. This still leaves a number of paradoxes surrounding the traditional view, which I argue defeat it.

[25] I argue for the claims in this section in much greater detail in Dorsey, 'Weak Anti-Rationalism'.

[26] See Dorsey, 'Weak Anti-Rationalism'; Dorsey, 'Against the Supremacy of Morality', MS.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

But those concerned about the rational authority of *morality* should not be worried to any *additional* extent about the anti-rationalist view.

Here's another way to see this point. Both the traditional and anti-rationalist views will say, of Rose (for instance), that the strongest balance of moral reasons favours her assisting the family. Both views will say that the strongest balance of moral reasons needn't capture 'how Rose should live'. The difference between the views relies only on their acceptance or rejection of the standard account of intra-moral rationality, and hence whether the denial of *Supremacy* entails the denial of *Supremacy**. For the purposes of argument, I'm willing to admit that to accept *Supremacy**, and hence to describe Rose's potential failure to assist the family as morally permitted, might very well be plausible.[27] But this advantage seems to me dwarfed by the advantages of the anti-rationalist view. Ultimately any choice between the traditional view and the anti-rationalist view must be made on the basis of their relative strengths and weaknesses. Different people will have different intuitions on this score. But as far as I'm concerned, given the mildness of the form of anti-rationalism in play and the identical treatment of *Supremacy*, any implausibility that remains in the denial of *Supremacy** is more than compensated for by the ability to accept the standard picture of intra-moral rationality, and avoid a bloated conception of moral justification. At the very least, enough has been said to grant the anti-rationalist view a very serious hearing.

VIII.3. Reconsidering Stan

At the end of section VI, I argued that the anti-rationalist view is to be preferred to the traditional view on grounds that it can avoid the traditional view's troubling implication that Stan could be morally justified in subjecting Jerry to beatings even if Stan's welfare is comparatively morally insignificant. The final objection, however, holds that this advantage is chimerical. My view might be committed to a conclusion that is, for all intents and purposes, identical. After all, I hold that prudential reasons might be balanced against moral reasons from the perspective of all-things-considered obligations, and it might be the case that in certain circumstances, prudential reasons are sufficient to grant rational permission to perform actions that do not conform to moral requirements. Why, then, haven't I *rationally* justified Stan's beating of Jerry? And if I have, haven't I implausibly made his refraining from so doing supererogatory?

There are two questions that should be separated here, however. The first concerns whether my view is committed to the claim that Stan's

[27] Thanks to an anonymous reviewer.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

beating of Jerry for the purpose of purchasing a new car can possibly be *justified*. The second concerns whether my view is committed to the claim that Stan's beating of Jerry for the purpose of purchasing a new car can possibly be *supererogatory*. Take the first question. The anti-rationalist view holds that one is morally required to donate in Gus's case, and is morally required not to beat up Jerry in Stan's case. But if Gus's prudential interest in a new car can rationally justify failing his moral obligation to save ten, and refraining from beating up Jerry is not morally more important than saving ten, it would appear that Stan's prudential interest in a new car can rationally justify ignoring his moral obligation not to beat up Jerry.

I have two responses, the first more tentative than the second. First, assuming the truth of a prima facie plausible conjecture, the anti-rationalist view can avoid this conclusion in a way that the traditional view cannot. Notice that, for the anti-rationalist view, non-moral considerations cannot count against the permission to conform to moral reasons (given *Permission*). But this says nothing about the ability of non-moral reasons to count against the permission to conform to other *non*-moral reasons. But – and here's the conjecture – there seems a *bevy* of non-moral reasons that count against beating up Jerry that need not count in favour of saving ten. Reasons of *neighbourliness*, *association*, to say nothing of *etiquette*, may in fact require Stan not to beat up Jerry, the force of which may be enough to outweigh Stan's prudential interests (especially when combined with the moral requirement not to do so). Notice that the traditional view cannot make the same claim. If reasons of etiquette, neighbourliness, etc., are genuinely non-moral, they have no power morally to require actions and hence cannot tell against Stan's moral justification to beat up Jerry. (Furthermore, if they're *per se* moral considerations, they are no help. As we have already seen, the moral factors involved have already been accounted for in (2), i.e. the moral justification to beat up Jerry to save ten.) Of course, this response requires a controversial account of the moral point of view, viz., that the moral point of view does not include reasons of neighbourliness, etiquette, association, etc. But the schematic point stands: if there are genuinely non-moral reasons that require one not to beat up Stan, which I hereby leave as a plausible conjecture, the anti-rationalist view can (in principle, anyway) avoid the problem faced by the traditional view in Stan's case.

But let's say, for the moment, that this response is not available. Let's say that the anti-rationalist view is committed to Stan's rational justification to beat up Jerry (perhaps because my conjecture fails or for some other reason). Even if this is correct, the anti-rationalist view is in a better position than the traditional view. As noted throughout, the anti-rationalist view and traditional view take very different

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

justificatory stances towards Stan's dastardly deed. The problem with the traditional view was never that Stan is *justified* in beating up Jerry, but rather that the traditional view is saddled with an implausible claim about morality, viz. that Stan is *morally* justified in beating up Jerry. I admit that some will think that both verdicts are implausible. But if Stan is to be justified, my approach is comparatively palatable: I needn't, but the traditional view must, claim that Stan is morally justified.

Take now the second question. Must my account hold that, were Stan not to harm Jerry, his action would be *supererogatory*? Of course, all the same responses apply: if my conjecture is right, the anti-rationalist view is not committed to this result, and if it is, the traditional view is no less committed. But leave this aside. Both views are licensed to reject the claim that Stan's action is supererogatory. One could, in principle, restrict the epithet 'supererogatory' to only a subset of morally good, but not rationally required, actions by further supplementing one's analysis of the supererogatory. Some hold that an action is supererogatory only if it is somehow *beneficent*.[28] If we accept this view, Gus's donation (which is beneficent) would be supererogatory, but Stan's failure to beat up Jerry (which isn't) wouldn't. On whether this further limiting constraint is all-things-considered plausible, I am officially neutral.

One might suggest that admitting that the traditional view can avoid marking Stan's failure to beat up Jerry as supererogatory saps any motivation for adopting the anti-rationalist view. But this is not so. The problem with the traditional view is that it relies on an extensionally incorrect theory of moral justification. The traditional view must accept the claim that Stan's harming of Jerry is *morally justified*, whether or not it is officially 'supererogatory'. This renders the traditional view, in my book, unacceptable in comparison to the superior anti-rationalist alternative.

IX. CONCLUSION: ACT-CONSEQUENTIALISM AND THE SUPEREROGATORY

The traditional view is puzzling. To accept the combination of *Permissible not Required*, *Morally Good* and *But for Sacrifice*, we must accept a form of intra-moral rationality that stretches the limits of moral justification. However, we can accommodate the existence of the supererogatory without altering the structure of intra-moral rationality if we assume that supererogatory actions just are those that are morally good, but for which one does not have decisive practical

[28] Heyd, *Supererogation*, p. 137. For a contrary view, see Gregory Mellema, *Beyond the Call of Duty: Supererogation, Obligation, and Offence* (Albany, 1991), ch. 2.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press

reason. This approach to the supererogatory perfectly captures the category of actions for which the supererogatory was designed: those actions that go beyond the call of one's duty.

But my view is significant for a further reason. As noted in the introduction, a traditional objection holds that act-consequentialism cannot accommodate the supererogatory. After all, act-consequentialism requires agents to maximize the value of consequences in every case, leaving no room for actions to be 'better' than is morally required. Morality, according to act-consequentialism, demands the best. This objection requires the traditional view for its cogency: it must be that someone can go beyond duty from *within* the moral point of view. But if my analysis is correct, this objection fails. The supererogatory is not an *intra*-moral category. The supererogatory exists between the demands of morality and the demands of practical rationality. Hence it is illegitimate to reject act-consequentialism on the grounds that this theory cannot accommodate the supererogatory.[29] The supererogatory is not a method to evaluate first-order moral theories, but rather a method by which to evaluate theories of all-things-considered practical rationality.

But while we're on the subject, in addition to not being an objection to act-consequentialism, the existence of the supererogatory is grist for the consequentialist's mill. If the supererogatory is best understood as a way in which the demands of practical reason separate from the demands of morality, to accommodate the supererogatory we must adopt a first-order moral theory that can plausibly *explain* this cleavage. But, famously, the demands of act-consequentialism do not plausibly match up with that which a given person is rationally required to do.[30] Hence, because act-consequentialism can plausibly explain the distinction between moral requirements and rational requirements, act-consequentialism is well placed to explain the existence of the supererogatory: sometimes morally required action is simply supererogatory, or beyond the call of that which we ought, rationally, to do.[31]

ddorsey@ku.edu

[29] Consider, for instance, Jason Kawall, 'Virtue Theory, Ideal Observers, and the Supererogatory', *Philosophical Studies* 146 (2009), pp. 179–96.

[30] Paul Hurley, *Beyond Consequentialism* (Oxford, 2009); Dorsey, 'Weak Anti-Rationalism'.

[31] Many thanks are due to Doug Portmore, who offered extensive comments on this article on two separate occasions. I would also like to thank Derek Baker, Jack Bricke, Ann Cudd, Ben Eggleston, Howard Nye, Matt Talbert, various participants at the 2010 Rocky Mountain Ethics Congress, as well as a very helpful audience at St Andrews University.

https://doi.org/10.1017/S095382081200060X Published online by Cambridge University Press