

The Rationally Supererogatory

CLAIRE BENN¹

The Polonsky Academy for Advanced Study
The Van Leer Jerusalem Institute
cmabenn@gmail.com

ADAM BALES

The Faculty of Philosophy
University of Cambridge
atb39@cam.ac.uk

The notion of supererogation—going above and beyond the call of duty—is typically discussed in a moral context. However, in this paper, we argue for the existence of *rationally* supererogatory actions: that is, of actions that go above and beyond the call of rational duty. In order to establish the existence of such actions, we first need to overcome the so-called paradox of supererogation. That is, we need to provide some explanation for why, if some act is rationally optimal, it is not the case that we are rationally required to carry out the act. We argue that a response to this ‘paradox’ can be found by reflecting on normative conflicts: cases where what is best according to some normative domain is different to what is best according to some other normative domain.

1 Introduction

Morally supererogatory actions are, roughly, those actions that go above and beyond the call of *moral* duty.² Over the last seven decades, such actions have been much discussed. In this paper, we will be interested in a more neglected notion: the notion of the *rationally* supererogatory. By analogy, and again roughly, rationally supererogatory actions are those that go above and beyond the call of *rational* duty. In the following, we will argue that just as some actions are morally supererogatory, some are rationally supererogatory. Specifically, we argue that the rationally supererogatory arises at points of normative conflict.

¹ Both authors contributed equally to this paper, and the order of author names was chosen at random.

² We use the terms ‘duty’, ‘obligation’ and ‘normative requirement’ interchangeably for the purposes of this paper.

2 Rational Supererogation

To understand what is meant by the rationally supererogatory, we need clarity on what is meant by rationality and what is meant by supererogation. Let's begin with the former. Rationality means many things in many contexts. For example, sometimes we talk about epistemic rationality. At other times, we talk about rationality as a broad notion of being responsive to all reasons.³ However, we are interested in the notion of rationality as *prudence*.⁴ As a first gloss, prudence is a domain of self-regarding normativity of choice, where this is to be contrasted with morality, which is primarily focused on the other. So, on this view, prudence relates to how we ought to act so as to promote our own wellbeing.

Yet this is hardly the last word on the matter. After all, many people think that we have self-regarding *moral* duties (and perhaps parts of prudential rationality are other-regarding). If so, the prudence–morality distinction won't neatly follow the self–other distinction. Nevertheless, this initial characterisation captures prudence to a rough approximate, so we will adopt it in the majority of this paper.⁵

For now, the important point is that references to rationality in this paper should be read as references to prudential rationality. In particular, when we talk about what is

³ The broad notion of all-things-considered rationality has been employed in discussions of supererogation. For example, Dale Dorsey conceives of *moral* supererogation as comprised of those acts that go beyond the call of our *rational* duties (2013, 2016). However, he states that 'when I use the terms "rational requirement", "rational authority", I simply mean to refer to *normative* requirements, *normative* authority' (2016, 18). Thus, 'rational requirement' is used to refer to an 'all-things-considered' requirement. Our account of the 'rationally supererogatory' is therefore importantly different from Dorsey's as it deals with a narrower notion of rationality as prudence, rather than a broad all-things-considered normative notion.

⁴ Why not simply use the word 'prudence', rather than 'rationality', throughout this paper? Because this term is used somewhat differently in philosophy, as compared with its more everyday use. This raises the risk that our intuitions about prudence will be influenced by this everyday notion, even when we are supposed to be focusing on the more technical notion.

⁵ Throughout, we will assume that there is some distinction between prudence and morality, even if this distinction doesn't perfectly follow the self–other distinction. Still, this assumption is commonly accepted, especially within discussions of moral supererogation.

rational in certain hypothetical scenarios, we are talking about what is prudentially rational in these scenarios.

So much for rationality. What does it take for an action to be rationally *supererogatory*? Well, to act in a supererogatory manner is, broadly, to go above and beyond the call of duty. Now, while there are many ways to make this precise, at the heart of all accounts sit two conditions:

DEONTOLOGICAL CONDITION: Supererogatory actions are neither required nor forbidden; they are optional.

AXIOLOGICAL CONDITION: Supererogatory actions are better than some permissible alternative.⁶

Supererogation is typically discussed in moral contexts, where these conditions refer to *moral* optionality and *moral* betterness. However, it is easy to see how these conditions could refer to other normative domains, including rationality. In this case, the conditions stipulate that an act is rationally supererogatory only if it is: (a) rationally optional; and (b) rationally better than a rationally permissible alternative. So our aim is to demonstrate that some actions meet these conditions.

Two clarifications are called for. First, we are *not* defending the claim that it can be rational to perform *morally* supererogatory actions. This is an important claim, as it sheds light on why it is not irrational, in some sense, to do what is morally best when doing so imposes a serious personal cost and it would be permissible to do

⁶ This condition stops supererogation arising in cases where the options available are merely equally good.

otherwise.⁷ Still, important or not, in this paper we are interested in the rationally supererogatory, not the rationality of the morally supererogatory.

Second, it is worth noting that there is debate over whether the deontological and axiological conditions jointly provide sufficient conditions for an action to be supererogatory. For example, on some views, such actions must also meet an evaluative condition: roughly, they must be praiseworthy. Still, we set this matter aside until §6.1 and focus, for now, on just the deontological and axiological conditions. After all, the main challenge to the possibility of supererogatory action in *any* domain, the paradox of supererogation, arises from just these two conditions.

3 The Paradox and Demandingness

It follows from the above conditions that a supererogatory act must be accompanied by a permissible, and worse, alternative. Consequently, supererogation can arise in a domain only if it is permissible to act suboptimally by the standards of that domain.⁸ The 'paradox' of supererogation demands a justification for this permissible suboptimality: if an act is suboptimal, according to some normative domain, why doesn't this domain forbid carrying out this act? Or, to put it another way: if some action is the best, according to some normative domain, why doesn't that domain require us to carry out this action? Framed either way, the paradox of supererogation notes a tension between the deontological and the axiological conditions.

⁷ Philippa Foot, for example, discusses why a farm boy's decision to be hanged rather than join the SS is not irrational (2004). She argues that a theory of what is rational in an overarching, all-thing-considered sense must be sensitive to moral considerations in a way that does not bring 'the rationality of moral action under the rationality of self-interested action, or that of the maximum fulfilment of present desire' (8).

⁸ Note the lack of the biconditional here. The permissibility of suboptimality doesn't suffice to make room for supererogation, as optimality might itself be impermissible (see Slote 1989, 127–28).

Resolving the paradox in a way that allows for the supererogatory is harder in the rational case than the moral. After all, the literature is dominated by optimising views of rationality, according to which we are rationally required to do the rationally best.⁹ As supererogation entails permissible suboptimality, optimising views entail that no acts are rationally supererogatory. The prominence of such views might leave us pessimistic about the prospects for rational supererogation.

Further, consider the common response to the paradox of supererogation in the moral domain: the appeal to over-demandingness. This response starts from the fact that doing what is morally best often involves making a substantial personal sacrifice. It is then argued that morality is not so demanding as to always require us to make such sacrifices in order to do good for others.¹⁰ This plausibly frees us from the requirement to always do what is morally optimal.

Now, if this response makes central reference to cost and sacrifice then it appears that no similar response can be applied to the paradox of *rational* supererogation. The rationally optimal decision is, by its very nature, the decision that is best for one's self. So rational optimality precludes self-sacrifice, if this sacrifice is thought of in prudential terms. Consequently, it seems that we cannot justify acting rational suboptimally by appealing to the idea that acting optimally involves self-sacrifice.

⁹ For example, decision theory enjoins one to *maximise* expected utility (see Savage 1954; Jeffrey 1965). Insofar as expected utility measures rational goodness, it follows that rational decisions are optimally good. Optimality also appears in various non-decision-theoretic accounts (see, for example, Rawls 1999, 23).

¹⁰ The centrality of sacrifice in discussions of moral supererogation can be seen in the accounts of, for example, Stanlick 1999; Straumanis 1984; McGoldrick 1984; Pybus 1982; Jackson 1986; Rawls 1999; Jacobs 1987; Portmore 2003. Benn has argued that an appeal to *comparative* cost (though not sacrifice) is in fact needed to understand concepts central to the notion of supererogation, such as optionality (Benn 2017).

However, we propose another interpretation of these over-demandingness objections. Perhaps over-demandingness is not specifically about *prudential* sacrifice, but about sacrifice of something that is valuable by the lights of another domain of normativity. In this case, the notion of demandingness can be generalised as follows:

DEMANDINGNESS: A domain of normativity, D, is demanding insofar as it requires us to sacrifice something of value by lights of another domain of normativity, E.

In discussions of moral supererogation, D is morality and E is prudential rationality. And the solution to over-demandingness here is allowing a realm of permissible, morally suboptimal actions that leaves space for the pursuit of prudential rationality.¹¹ However, DEMANDINGNESS applies *whenever* we are asked to act contra any normative domain, not just when we are asked to act contra prudential rationality. Thus, a domain of normativity can be demanding even if it doesn't involve a specifically prudential sacrifice. Prudential cost is, after all, only one kind of cost.

Returning to the paradox of rational supererogation, we can now see how rationality could be demanding: the prudentially optimal action may sometimes require an agent to sacrifice something of value according to some other normative domain.

¹¹ This suggests that considerations from one normative domain can affect other domains. Note, however, we are specifically arguing that moral considerations can influence the *deontological* status of an act (that is, roughly, whether it is permissible, impermissible or required) according to another domain. We say nothing about whether moral considerations affect the *axiology* of the prudential domain: that is, that what is morally good can affect what is prudentially good. Now, some, such as Foot, have argued for just that (1959). However, so long as the axiological rankings of morality and prudence do not simply collapse into one another even that claim presents no threat to the idea of distinct domains, or the possibility of conflict.

Of course, supererogation doesn't arise from all conflicts between normative domains: sometimes we are required to make great sacrifices. However, cases of demandingness are promising places to look for supererogation. Further, having found a case of the supererogatory, issues of demandingness explain *why* the action is supererogatory, rather than required.

Then, just as DEMANDINGNESS justifies the permissibility of morally suboptimal actions and therefore moral supererogation, it can also justify the permissibility of rationally suboptimal acts and therefore the rationally supererogatory. Just as with morality, rationality is surely not so demanding as to always require us to make sacrifices by the light of some other normative domain. For this reason, perhaps we are, at times, rationally permitted to act suboptimally. Thus, rational supererogation may arise when acting rationally optimally requires such a sacrifice.

4 Normative Conflict

DEMANDINGNESS give us a roadmap of where to look for the rationally supererogatory: in places of normative conflict. We begin our search considering two cases where conflicts arise between rationality and some other normative domain.

4.1 *The Moral Domain*

The over-demandingness defence of moral supererogation is based on the idea that morality is not so demanding that it requires us to make substantial *prudential* sacrifices. Likewise, we argue, rationality is not so demanding that it requires us to make substantial *moral* sacrifices. Consider:

THE VOLUNTEER: Kavya, a doctor, could volunteer to provide free cataract surgery to those in need. Yet, in terms of her own well-being, Kavya would be marginally better off if she instead went on a writing course she has been longing to attend.

This is a classic case in which the morally supererogatory arises: it is morally better for Kavya to volunteer than to attend the course and yet morality is not so demanding that she is morally *required* to volunteer. Thus, the prudential cost of acting morally optimally makes room for the morally supererogatory.

The same move can, we argue, then be made in the other direction. Given that Kavya will be prudentially best off if she refuses to volunteer, it is prudentially permissible for her to do so. However, it is also, we submit, prudentially permissible for Kavya to volunteer. Prudential rationality is surely not so demanding as to require that she sacrifice something of substantial moral value for the sake of the minor prudential gain of attending the writing course. So attending the course is prudentially optional. It therefore meets the deontological condition for supererogation.

Further, attending the course is prudentially better than the permissible alternative of volunteering, and so we have the axiological condition. Therefore, attending the course is rationally supererogatory: it is rationally permissible and rationally better than the permissible alternative of volunteering. It follows that we have an over-demandingness argument for the rationally supererogatory, one that mirrors the standard over-demandingness argument for the morally supererogatory.

Indeed, this discussion reveals how to find a range of cases of the rationally supererogatory: we should look for such cases amidst examples of the morally supererogatory. After all, over-demandingness arguments for moral supererogation involve demonstrating that some act is both morally better and prudentially worse than another permissible act. This other act, while morally worse, must therefore be *prudentially* better than the morally optimal act. Furthermore, this act will often be rationally optional, as rationality is often not so demanding that it will require us to sacrifice something of moral value. So, many familiar cases involving moral

supererogation will also involve rational supererogation (in relation to a different action).

4.2 *The Epistemic Domain*

Morality and prudential rationality do not exhaust the domains of normativity and conflicts can occur elsewhere. Consider:

THE DIARY: Sebastian has inherited his deceased grandfather's diary, which he knows his grandfather would have been happy for him to read. However, he also suspects that if he reads it, he will discover that his grandfather was deeply sexist. Sebastian would find this upsetting, such that his well-being would be best promoted by not reading the diary.

Despite the upsetting nature of discovering this sexism, Sebastian is, we suggest, rationally permitted to read the diary. Rationality would be demanding indeed if it *forbade* him from reading it and sating his curiosity. Yet he is also rationally permitted to not read the diary given that he would be better off if he doesn't do so. So not reading the diary is rationally optional. This gives us the deontological condition for supererogation. Further, refraining from reading the diary is prudentially better than the permissible alternative of reading it. So, we have the axiological condition. Refusing to read the diary is therefore rationally supererogatory.¹²

¹² This argument for rational supererogation arising out of conflict with the epistemic lends plausibility to the idea of *epistemic* supererogation. For an independent defence of epistemic supererogation, see Hedberg 2014.

Further, we can explain why this is the case. It is plausibly *epistemically* better to read the diary than to not do so. Thus, refusing to read the diary comes at an epistemic cost. Prudential rationality, we suggest, is not so demanding as to require that Sebastian undertake an action that leads to this cost. Cases of supererogation therefore arise when prudential rationality and the epistemic conflict.

So THE VOLUNTEER and THE DIARY provide two examples of the rationally supererogatory: where prudence conflicts with what is morally best and where it conflicts with what is epistemologically best. And if you accept other domains of normativity in addition to the ones mentioned here, the possibility of rationally supererogatory action may well arise there too when what is best according to that domain conflicts with what is prudentially best.¹³

5 Beyond Conflict?

Still, it's not clear that all cases of rational supererogation arise from conflicts between prudential rationality and some other normative domain.

5.1 Two Cases

Consider two further examples of the rationally supererogatory. First:¹⁴

THE MATHEMATICIAN: Omari, who enjoys a full and balanced life, wants to be the best mathematician in his school. One day, his school is

¹³ The domain of 'aptness' has received a lot of attention recently (e.g. Srinivasan 2018) and it is possible to imagine cases where what is apt conflicts with what is rationally (in the sense of prudentially) best. If so then aptness might be another normative domain that gives rise to prudential supererogation.

¹⁴ Similar cases have been discussed in Slote (1989, 118–30) and McElwee (2017, 510). We discuss their arguments in more detail in §5. Related cases are also discussed in Kawall (2003, 490). However, Kawall is most naturally read as discussing self-regarding *moral* supererogation. We will return to this matter later.

closed because of snow and his parents tell him that he may make a snowman in the garden. Although he is tempted, Omari instead spends two hours revising maths, bringing him closer to realising his mathematical goal.

Assuming that the benefits of extra revision outweigh the costs, revising is rationally optimal. Yet, common-sense suggests that it would also be rationally permissible for Omari to give revision a miss. Snow is, we can stipulate, exceedingly rare where Omari lives, so making a snowman will not set a poor precedent. Further, those of us who remain young at heart can well understand Omari's temptation to make a snowman. Omari is permitted to build the snowman. Hence, revising is rationally supererogatory: it is optional and better than the permissible alternative of building the snowman. Yet this case does not, at least in any obvious way, involve a conflict between rationality and some other normative domain. So, insofar as this is a case of supererogatory action, it seems as though the supererogatory can arise even in the absence of such a conflict.

Nor is this an isolated case. The case for rational supererogation without apparent normative conflict does not depend on anything specific about Omari's circumstances. To see this, consider a second case, which we adapt from Michael Slote (1989, 133–36):¹⁵

DISTRACTIONS: While cooking dinner, Zhang gets distracted from time-to-time: he looks out the window, checks social media, and rearranges

¹⁵ Slote also discusses an alleged case of rational supererogation involving infinitely-many options, none of which are optimal. We set such cases aside both because we prefer to focus on realistic cases and because we're dubious that this case truly involves supererogation.

his fridge. These actions aren't justified by Zhang's desires: he doesn't much care about how the fridge is arranged, for example, nor does he enjoy cooking more if he does it at a relaxed pace. Rather, Zhang is just engaged in trivial time wasting.

Slote suggests, and we agree, that Zhang acts rationally permissibly by engaging in such time wasting. However, Zhang would be slightly better off if he didn't fritter away his time, as he could then, say, spend slightly more time with a friend later in the evening. So focused cooking is both optional and better than the permissible alternative of unfocused cooking. Focused cooking is rationally supererogatory. We have another case where rationality supererogation does not appear to arise from a conflict between rationality and some other normative domain.

However, the paradox of supererogation arises for these cases. If revising is the best thing for Omari to do, why isn't he required to revise? If focused cooking is the best thing for Zhang to do, why isn't he required to cook in a focused manner? These questions were answered for earlier cases by appeal to normative conflicts. Such an explanation cannot straightforwardly be given for the cases of Omari and Zhang, as there does not appear to be any such conflict in these cases. We have a problem.

5.2 *Ordinary Capacities*

Fortunately, there's a natural way to resolve this problem. In particular, in discussing moral supererogation, J.O. Urmson claimed that morality should not demand that which is beyond the capacity of ordinary people (1969, 70). Similarly, we might hope that the

above problem can be resolved by accepting that rationality should not make such demands.

Indeed, Slote has endorsed just such a view of rationality. In particular, Slote points to situations where: (a) it is difficult for someone to do what is rationally optimally; (b) it is unreasonable to expect someone to do what is rationally optimally; or (c) doing what is rationally suboptimally is 'so normal and tolerable' that we are 'unwilling to consider it rationally unacceptable' (Slote 1989, 135).¹⁶ Slotes suggests that in these kinds of situations— for brevity, we'll call them 'Slotian situations'—it can be rationally permissible for the agent to do what is rationally suboptimally. Thus, Slote's claim about the rational domain is very much in-keeping with Urmson's claim about the moral: the ordinary capacities of humans mean that it's unreasonable to demand that we act optimally when it is, for example, difficult to do so.

If we accept this view, then the paradox of supererogation can be answered with regards to the above cases. Consider *THE MATHEMATICIAN*. Here, it may be difficult for Omari to resist the temptation to build the snowman. As a consequence, it may be unreasonable to expect Omari to revise rather than building the snowman. Further, acting suboptimally (by building the snowman) is likely to strike us as 'so normal and tolerable' that we may well be unwilling to judge it irrational. The Slotian nature of the situation provides a justification for why it's rational to act suboptimally in this case.

Likewise, consider *DISTRACTIONS*. Here, it may well be difficult for Zhang to ensure that his focus never wanders. Further, it is unreasonable to expect that Zhang will cook with perfect focus and so failing to do so is 'so normal and tolerable' that we would hardly condemn someone as irrational for such a failure. Again, the situation is

¹⁶ The general viewpoint mentioned here is discussed throughout Slote 1989, chap. 5 and 6.

Slotian and so we can explain the permissibility of Zhang acting suboptimally. The paradox of supererogation has been addressed.

5.3 *Conflicts Again*

Nor need we stop there. Instead, we could seek a deeper explanation of why Slotian situations often go hand in hand with permissible suboptimality. And, reflecting on the first half of this paper, an interesting possibility arises. Perhaps such situations arise in cases of normative conflict. As normative conflict can lead to permissible suboptimality, this would explain why Slotian situations accompany some cases involving permissible suboptimality.

In order to see how the cases discussed involve *normative* conflict, we must begin by seeing how conflict of any sort arises in these cases (we will then turn to what makes these conflicts normative). Well, we can see where the conflict lies by considering *what* makes it difficult to do what is rationally best. Humans often experience fleeting, but strong, desires (as, for example, when we feel the tempting lure of chocolate cake). Further, these fleeting desires will often be to do things that are not prudentially optimal for us. Consequently, such desires will give rise to Slotian situations. That is, it will often be difficult to act optimally when we face fleeting (but strong) desires to act otherwise. Also, in such cases it will often be unreasonable to expect someone to act optimally and will be normal and tolerable for someone to act suboptimally. So Slotian situations naturally arise when fleeting desires favour suboptimal behaviour and so conflict with what is prudentially best.

We've already shown that normative conflicts can lead to permissible suboptimality. So if the above conflicts are *normative* (and so Slotian situations

naturally arise from normative conflicts) then we have an explanation of why such situations often accompany permissible suboptimality. But why think of the clash between fleeting desires and prudential optimality as a *normative* conflict? Let's explore two potential answers to this question.

The first appeals to the fact that it is common to distinguish two domains of self-facing rationality: instrumental rationality and prudential rationality.¹⁷ Instrumental rationality relates to the selection of means that are well placed for achieving our ends (where these ends are construed in terms of our current desires).¹⁸ Prudential rationality, on the other hand, is broader: our current desires might influence what we prudentially ought to do, but so do other considerations (perhaps, for example, considerations of objective well-being). Now instrumental rationality will be dramatically influenced by strong, but fleeting, desires given that it is responsive only to current desires. So when strong, fleeting desires conflict with prudential optimality, instrumental rationality and prudential rationality will often conflict. Furthermore, if we accept that instrumental rationality has normative force, then this conflict will be a normative one.

The second potential explanation of how the above conflict is normative appeals to the idea that some values, principles, and projects are more fundamental than others to how we define ourselves.¹⁹ In our day-to-day lives, this distinction plays an important role in how we think of ourselves: we often speak of who we really are, of our authentic selves, or of our true selves. For example, perhaps some people see being a feminist as a core part of their identity (that is, as part of their *true* self), but merely enjoy drinking

¹⁷ It is far from unanimously accepted that both of these normative domains exist, so the discussion of this section will convince only those, like us, who are willing to accept the existence of both domains.

¹⁸ In fact, there is some debate over which desires are relevant to instrumental rationality. However, the present-desire account is a popular one.

¹⁹ We are not endorsing a Millian distinction between objectively higher and lower projects. Rather, we are pointing to a distinction between people's self-conception of their different projects.

wine (without seeing this as central to who they are). Following Sripada, we will say that a person's central values (and principles, and projects) form her *deep self*, while the less central values form her *shallow self* (Sripada 2010).²⁰ However, while this terminology nicely captures the idea that some of our values sit at the core of our identity, a note of caution is called for: despite our choice of language, we do not intend to commit ourselves to the idea that there are two genuine metaphysical selves in each of us. Rather, we remain neutral on this matter, and simply take the reference to selves to be an appeal to a metaphor familiar from our day-to-day lives.

From our perspective, the important point about the distinction between the deep and shallow self is the following: given its association with our most fundamental projects, acting in accordance with the interests of our deep self will often leave us prudentially best off overall. Yet, for many people, the desire for creature comforts will come from their shallow self and such desires should not be entirely dismissed, even if it would be optimal to do so.²¹ Now we can provide a second potential explanation of why a conflict between fleeting desires and prudential optimality often involves a *normative* conflict. In particular, in such cases, there is often a conflict between the

²⁰ Sripada's notion of the deep self is a little different to our own: while we characterise the deep self purely in terms of what is central to a person's identity, to Sripada this is an important but not exhaustive aspect of the deep self. Related (though not identical) notions have been discussed in key philosophical works (such as Hume 1738, Bk 11, Pt, 111, Sec. 2; Watson 1975; Frankfurt 1971; see also Arpaly and Schroeder, "Praise, Blame and the Whole Self" and Callard 2018, 3). The important psychological roles that the concept of the deep self plays has been explored in, amongst others, Strohminger, Knobe, and Newman (2017). For example, it has been shown that when people are primed to think about their deep self, they see their life as more meaningful (Arndt et al. 2002; Schlegel et al. 2009).

²¹ Of course, one could claim that the desires of a shallow self count for nothing, prudentially speaking. We deny this for two reasons. First and foremost, we simply take it to be obvious that a life devoid of such creature comforts would seem to lack something of value, even if it otherwise goes well. Second, while we think it implausible to claim that prudence is simply about desire satisfaction, we think the claim that desires count for nothing prudentially speaking is equally implausible. Those who adamantly reject these arguments can help themselves to the alternative form of normative conflict, discussed above.

deep and the shallow self. Further, insofar as both of these aspects of the self have normative standing, such a conflict is a normative conflict.²²

These two solutions have different structures. The first identifies a conflict between prudence and a distinct domain of normativity (instrumental rationality). On the other hand, the second solution identifies a conflict *within* the domain of prudence (a conflict between those aspects of prudence that are concerned with the deep self and those aspects that are concerned with the shallow self).

Still, despite their differences, both of these solutions share a common feature: they show how normative conflict can arise from a clash between stable central interests and more fleeting peripheral interests. As situations of the kind Slote describes arise from these sorts of clashes, these solutions explain why Slotian situations will often accompany normative conflict and so why they will often accompany permissible suboptimality.

We find both of these explanations promising and illuminating. Thus, in the remainder of the paper, we take it that one or the other of these explanations is right, and so speak as if all of our examples of rational supererogation ultimately arise from normative conflicts. Yet one can, in the style of Urmson and Slote, appeal to ordinary capacities and considerations such as difficulty to answer the paradox of supererogation in the cases described without accepting the accounts just outlined. In either case, we can accept THE MATHEMATICIAN and DISTRACTIONS as examples of the rationally supererogatory.

²² This raises the possibility of a plethora of previously unimagined normative conflicts. For example, might there be *moral* (rather than prudential) conflicts between the deep and the shallow self? We think this question is worth further reflection, though we lack the space to explore this matter in any depth here.

6 Issues Arising

The above discussions form the core of our argument: rational supererogation ultimately arises from normative conflict.²³ In the remainder of the paper, we comment on three issues arising from this discussion.

6.1 *The Problem of Praise*

The first of these issues is an objection. This objection starts from an endorsement of:

EVALUATIVE CONDITION: Supererogatory actions are praiseworthy.²⁴

If we accept this condition, a problem arises, as some of the actions that we've declared to be rationally supererogatory do not appear to be praiseworthy.²⁵ This concern arises most clearly in *THE VOLUNTEER*. Here, we declared that Kavya's refusal to volunteer was rationally supererogatory. Yet such an act seems far from praiseworthy. So our claim that this act is supererogatory is under threat if we accept the evaluative condition.

Assuming that we want a unified account of supererogation, there are two ways to respond here. First, we could reject the evaluative condition. Second, we could show that accepting the evaluative condition does not rule out the possibility of rationally supererogatory acts. Now, adopting the first strategy is not as ad hoc as it might at first

²³ Noting, as above, that we will be assuming that the ordinary capacities approach ultimately involves an appeal to just such conflicts. For those who deny this optional move in our argument, rational supererogation will instead arise from two sources: normative conflict and cases where optimality is beyond the ordinary capacities of agents.

²⁴ This is framed as if praiseworthiness is a feature of the action rather than of the agent. However, the response to follow applies to either way of thinking about this notion. For discussions of this condition, see Attfield 1979; Clark 1978; Curtis 1981; Jacobs 1987; Baron 1987; Raz 1975; Mellema 1987. This condition also features in the debate between Pybus (1982) and McGoldrick (1984).

²⁵ Though this problem afflicts only some of our examples of the supererogatory and, thus, even were it a genuine problem, it would not undermine our basic case.

appear, as it has become increasingly common in the literature on moral supererogation to reject this condition (Urmson 1969; Rawls 1999; Heyd 1982; Archer 2016). Still, we will show that the second strategy can also be adopted: if we retain the evaluative condition, our account remains unproblematic.

To see why, first note that when we talk of praiseworthiness, we tend to mean moral (or perhaps all-things-considered) praiseworthiness. Yet even if there is an evaluative condition for *rational* supererogation, this presumably only requires that rationally supererogatory actions be *rationally* praiseworthy (just as the other two conditions appeal to *rational* betterness and *rational* permissibility). Now, even though Kavya attending the writing course is plausibly not *morally* praiseworthy, it's unclear why it wouldn't be rationally praiseworthy. So, we have been provided with no reason to think that an appropriately-construed evaluative condition undermines our account. Indeed, given that attending the course is prudentially optimal but comes at some cost (in moral terms), there are at least *prima facie* grounds to think that this action is prudentially praiseworthy: it involves acting in the prudentially best way despite the cost of doing so. From a purely prudential perspective, Kavya seems praiseworthy for acting in this manner.

More broadly, our cases of the rationally supererogatory arise when the rationally optimal action comes at a cost by the lights of another domain of normativity, or is beyond the ordinary capacities of agents. This effort or cost can explain why it could well be *prudentially* praiseworthy for the agent to do what is rationally optimal in such cases.²⁶ Overall, then, the EVALUATIVE CONDITION does not undermine the

²⁶ Even more can be said if we are happy to appeal to the above distinction between the deep and the shallow self. After all, acts that promote a person's deep interests (at a shallow cost) plausibly call for prudential praise, as such acts are powerful expressions of a person's agency. Furthermore, the distinction between the deep self and shallow self allows us to answer a question explored by Aristotle, of how to distinguish positive and negative senses of self-love (*NE* 1168b15-25). In particular, we can adopt the

claim that our cases (including THE VOLUNTEER) are cases of rationally supererogatory actions.

6.2 *Self-Regarding Morality*

Still, a further problem arises. In particular, we've assumed that, while prudence is self-regarding, morality is exclusively other-regarding. Yet it's not clear that this is true. Indeed, various people have argued that we have self-regarding moral duties (Hills 2003; Baier 1966). Yet, if this is so then perhaps our cases establish not the existence of the rationally supererogatory but, rather, the existence of the self-regarding *morally* supererogatory. For example, perhaps Omari studying is morally, rather than rationally, supererogatory, because studying in this way advances the moral project of fulfilling one's potential.

Of course, this doesn't impact the force of our argument for those who deny the existence of self-regarding morality. Yet it also lacks force for those who do accept such a view. After all, while some of our cases, like THE MATHEMATICIAN, could be explained in self-regarding moral terms, in other cases such an explanation would hardly be compelling. For example, would Zhang really be acting in a *morally* superior way if he cooked dinner in a focused manner as compared with an unfocused manner? Assuming that we think not, this can't be explained as a case of moral supererogation. More dramatically, would Kavya really be acting in a *morally* better manner by going on the writing course, as compared with volunteering as a doctor? Again, assuming that we think not, this is not a case of the morally supererogatory. These cases cannot

position that self-lovers gratify their deep self, while egoists gratify themselves to the exclusion of their deep self. This distinction can explain how some actions—namely, those actions that flow from our deep self and demonstrate a love of our true selves—can be rationally praiseworthy, while allowing that others—those that are merely selfish, shallow and self-indulgent—are not.

naturally be explained as cases of self-regarding moral supererogation. Regardless of whether or not we have self-regarding moral duties, these cases are best explained as cases of rational supererogation. So while accepting self-regarding morality might threaten some of our examples of the rationally supererogatory, it does not threaten our overall argument.²⁷

6.3 *Prudential Criticism*

The final issue relates to an alternative way of arguing for the rationally supererogatory. In particular, an argument for the supererogatory could start from the judgement that agents can remain immune to criticism (in some crucial sense) even while acting suboptimally. It could then be argued that this immunity reveals that the suboptimal action is permissible and so reveals that the optimal alternative is supererogatory. Whereas our argument starts from judgements of demandingness and permissibility, these arguments start from judgements of criticisability.

Brian McElwee puts forward an argument of just this sort (McElwee 2017). At the core of his argument is the claim that an action is rationally obligatory only if failure to perform it merits the 'distinctive prudential criticism' of 'foolishness, or simply imprudence' (McElwee 2017, 509). Further, McElwee claims, this charge is appropriate only when a person acts in a 'grossly deficient' manner with respect to their prudential interests (McElwee 2017, 210). Thus, McElwee's argument can be presented in terms of two premises:²⁸

²⁷ Further, it is a matter of debate whether there can be self-regarding supererogation, whether moral or otherwise. Consequently, even if our cases were read as examples of self-regarding *moral* supererogation then our result would still be noteworthy (and would see us joining the likes of Kawall (2003)).

²⁸ We frame this argument slightly differently to McElwee, partly for consistency with the language of this paper and partly to clarify the structure of his argument.

(P1) It is only fitting to charge someone as foolish or imprudent if she acts in a manner that is grossly deficient at promoting her well-being.

(P2) If it is not fitting to charge someone as foolish or imprudent for making a decision then it's rationally permissible for her to make that decision.

From (P1), it follows that it is not fitting to charge someone as foolish or imprudent *merely* because their decision was suboptimal (as it could be suboptimal and yet above the level of gross deficiency). Combined with (P2), it follows that it is permissible both to make the optimal decision and to make slightly suboptimal decisions (because in neither case will it be fitting to charge the person with foolishness or imprudence). So the optimal decision is permissible and rationally better than some permissible alternative. That is, the optimal decision is supererogatory.

This argument is interesting, insofar as it proceeds without appealing to our judgements about concrete cases. However, we worry that it trades on an equivocation in the phrase 'foolish or imprudent'. Take (P1). This premise is plausible if it refers to foolishness. After all, foolishness is a harsh judgement and so should only be applied when someone acts grossly deficiently. However, this premise is far less plausible if it refers to imprudence. Suppose that a friend goes for a walk on a warm spring day without taking water, knowing that she won't get lost and that the weather won't get oppressively hot. Such behaviour does not involve gross deficiency in self-care but is nevertheless imprudent. Gross deficiency is not required for a charge of imprudence.

Now take (P2). Here, it is plausible that if someone acts impermissibly then it must be fitting to charge her with imprudence. However, it's far less clear that the same is true of *foolishness*: plausibly, one can act rationally impermissibly without deserving the extremely harsh judgement that they acted foolishly. Indeed, it seems perfectly coherent to say, for example, that there are times when it would be rationally impermissible to forgo sunscreen when going for a walk but where it would be a bit much to call someone a fool for doing so.

Now we can see the equivocation: (P1) holds in relation to foolishness but not imprudence; (P2) in relation to imprudence but not foolishness. The argument only seems to work because the phrase 'foolish or imprudent' allows us to focus on different notions when considering each premise. Disambiguate this phrase in any which way and the argument loses its force.

Still, perhaps McElwee's argument can be rescued. After all, there are many forms of prudential criticism (reckless or ridiculous, hare-brained or rash, precipitate or disastrous and so on) and one of them may satisfy both (P1) and (P2). If so, McElwee could appeal to this prudential criticism and establish the existence of the rationally supererogatory. What would such success mean for our own paper? Well, if two independent arguments point towards the rationally supererogatory this provides particularly strong grounds to think these arguments promising. So if McElwee's argument succeeds, our arguments would work in tandem to provide strong support for the rationally supererogatory.

Further, even if McElwee's argument, like ours, established the existence of the rationally supererogatory, our paper carries out three additional tasks. First, we provide a deeper taxonomy of the kinds of rationally supererogatory actions that exist, by pointing to how supererogation arises from different sorts of normative conflicts.

Second, we provide a detailed explanation as to how the paradox of supererogation can be resolved: it can be resolved by pointing to normative conflicts and the notion of demandingness.²⁹ And third, by revealing the link between normative conflict and supererogation, our paper demonstrates how a commitment to the possibility of supererogation in one domain, such as ethics, lends weight to the possibility of supererogation in another, such as rationality.

7 Conclusion

Having spent some time lost in the details of our three 'issues arising', let's take a step back and consider the big picture once more. Well, the core challenge to establishing the rationally supererogatory is the paradox of supererogation: if some action is rationally optimal, why shouldn't we be rationally required to carry it out? In order to address this paradox, we appealed to the same rationale that justifies the existence of *morally* supererogatory acts: over-demandingness. Acting optimally may sometimes require sacrificing something of value by the lights of some other normative domain or by that domain's own lights. Rationality, we have argued, just like morality, is not always so demanding as to require such sacrifices, and so it is sometimes rationally permissible to act suboptimally. Thus, we demonstrated that we can locate the supererogatory in places of normative conflict.

By appeal to this framework, we presented two examples of the rationally supererogatory, THE VOLUNTEER and THE DIARY, both of which straightforwardly

²⁹ McElwee provides his own resolution of the paradox: suboptimality can be permissible because it is not fitting to criticise all suboptimality. Yet we think that this gets things the wrong way around: it is the permissibility (or impermissibility) of an action that partially explains the fittingness of criticising someone for carrying out that action (and *not* vice versa). In any case, McElwee's argument combines with our own to show that the paradox can be addressed even given wildly divergent normative views.

resulted from conflicts between rationality and some other normative domain. We also discussed another two cases, THE MATHEMATICIAN and DISTRACTIONS, where things were somewhat more complex. We explored how these cases could also be explained in terms of normative conflict: between the two separate normative domains of instrumental and prudential rationality or *within* the domain of prudence. Alternatively, it could be explained without any appeal to conflict, by taking it to instead arise purely as a result of the ordinary capacities of agents.

Thus, while perhaps it is not possible for the prudential domain to involve self-sacrifice, we have demonstrated it can nevertheless be over-demanding. Therefore, our four cases reveal that, just as we can go above and beyond the call of moral duty, we can also go above and beyond the call of rational duty.³⁰

³⁰ We would like to thank Jane Heal, Michael Slote, Bernhard Salow, Georgie Statham, Olla Solomyak, Casper Storm Hansen, Elena Cagnoli Fieconi, Elvira Di Bona, Leora Katz, Sharon Berry and Silvia Jonas as well as the anonymous reviewers and editors of this journal for their helpful comments on this and earlier drafts of this paper.

References

- Archer, Alfred 2016, 'Are Acts of Supererogation Always Praiseworthy?', in *Theoria* 82 (3): 238–55, <https://doi.org/10.1111/theo.12085>.
- Arndt, Jamie, Jeff Schimel, Jeff Greenberg, and Tom Pyszczynski 2002, 'The Intrinsic Self and Defensiveness: Evidence That Activating the Intrinsic Self Reduces Self-Handicapping and Conformity', in *Personality and Social Psychology Bulletin* 28 (5): 671–83.
- Arpaly, Nomy, and Timothy Schroeder 1999, 'Praise, Blame and the Whole Self', in *Philosophical Studies* 93 (2): 161–88.
- Attfield, Robin 1979, 'Supererogation and Double Standards', in *Mind* 88 (352): 481–99, <http://mind.oxfordjournals.org/content/LXXXVIII/1/481.full.pdf>.
- Baier, Kurt 1966, 'Moral Obligation', in *American Philosophical Quarterly* 3 (3): 210–26.
- Baron, Marcia 1987, 'Kantian Ethics and Supererogation', in *The Journal of Philosophy* 84 (5): 237–62, <http://www.jstor.org/stable/10.2307/2026752>.
- Benn, Claire 2017, 'Supererogation, Optionality and Cost', in *Philosophical Studies* 175 (10): 2399–2417, <https://doi.org/10.1007/s11098-017-0965-7>.
- Callard, Agnes 2018, *Aspiration: The Agency of Becoming* (New York: Oxford University Press).
- Clark, Michael 1978, 'The Meritorious and the Mandatory', in *Proceedings of the Aristotelian Society* 79: 23–33, <http://www.jstor.org/stable/10.2307/4544931>.
- Curtis, Barry 1981, 'The Supererogatory, the Foolish and the Morally Required', in *Journal of Value Inquiry* 15: 311–18.
- Dorsey, Dale 2013, 'The Supererogatory and How to Accommodate It', in *Utilitas* 25

- (3): 355–82.
- 2016, *The Limits of Moral Authority* (Oxford: Oxford University Press).
- Foot, Philippa 1959, 'Moral Beliefs', in *Proceedings of the Aristotelian Society* 59: 83–104.
- 2004, 'Rationality and Goodness', in *Royal Institute of Philosophy Supplement* 54: 1–13.
- Frankfurt, Harry G. 1971, 'Freedom of the Will and the Concept of a Person', in *Journal of Philosophy* 68: 5–20.
- Hedberg, Trevor 2014, 'Epistemic Supererogation and Its Implications', in *Synthese*, 191 (15): 3621–37, <https://doi.org/10.1007/s11229-014-0483-5>.
- Heyd, David 1982, *Supererogation: Its Status in Ethical Theory* in Cambridge Studies in Philosophy (Cambridge: Cambridge University Press).
- Hills, Alison 2003, 'Duties and Duties to Self', in *American Philosophical Quarterly* 40 (2): 131–42.
- Hume, David 1738, *A Treatise of Human Nature* (Oxford: Oxford University Press).
- Jackson, M.W. 1986, 'The Nature of Supererogation', in *The Journal of Value Inquiry* 20: 289–96, <http://link.springer.com/article/10.1007/BF00146118>.
- Jacobs, Russell A. 1987, 'Obligation, Supererogation and Self-Sacrifice', in *Philosophy* 62 (239): 96–101, <http://www.jstor.org/stable/3751025>.
- Jeffrey, Richard 1965, *The Logic of Decision* (University of Chicago Press).
- Kawall, Jason 2003, 'Self-Regarding Supererogatory Actions', in *Journal of Social Philosophy* 34 (3): 487–98, <https://doi.org/10.1111/1467-9833.00196>.
- McElwee, Brian 2017, 'Supererogation Across Normative Domains', in *Australasian Journal of Philosophy* 95 (3): 505–16, <https://doi.org/10.1080/00048402.2016.1253760>.

- McGoldrick, Patricia M. 1984, 'Saints and Heroes: A Plea for the Supererogatory', in *Philosophy* 59 (230): 523–28, <http://www.jstor.org/stable/3750369>.
- Mellema, Gregory 1987, 'Quasi-Supererogation', in *Philosophical Studies* 52: 141–50, <http://www.springerlink.com/index/NQ90065355311077.pdf>.
- Portmore, Douglas W. 2003, 'Position-Relative Consequentialism, Agent-Centered Options, and Supererogation', in *Ethics* 113 (2): 303–32, <http://www.jstor.org/stable/10.1086/342859>.
- Pybus, Elizabeth M. 1982, 'Saints and Heroes', in *Philosophy* 57 (220): 193–99.
- Rawls, John 1999, *A Theory of Justice* Rev. Ed. (Cambridge, Mass: The Belknap Press of Harvard University Press).
- Raz, Joseph 1975, 'Permissions and Supererogation', in *American Philosophical Quarterly* 12 (2): 161–68, <http://www.jstor.org/stable/10.2307/20009570>.
- Savage, Leonard Jimmie 1954, *The Foundations of Statistics* (Wiley Publications in Statistics).
- Schlegel, Rebecca J., Joshua A. Hicks, Jamie Arndt, and Laura A. King 2009, 'Thine Own Self: True Self-Concept Accessibility and Meaning in Life', in *Journal of Personality and Social Psychology* 96 (2): 473.
- Slote, Michael 1989, *Beyond Optimizing* (Cambridge: Harvard University Press).
- Srinivasan, Amia 2018, 'The Aptness of Anger', in *Journal of Political Philosophy* 26 (2): 123–144, <https://doi.org/10.1111/jopp.12130>.
- Sripada, Chandra Sekhar 2010, 'The Deep Self Model and Asymmetries in Folk Judgments about Intentional Action', in *Philosophical Studies* 151 (2): 159–76, <https://doi.org/10.1007/s11098-009-9423-5>.
- Stanlick, Nancy A. 1999, 'The Nature and Value of Supererogatory Actions', in *Journal of Social Philosophy* 30 (1): 209–22, <https://doi.org/10.1111/0047->

2786.t01-1-00013.

Straumanis, Joan 1984, 'Duties to Oneself: An Ethical Basis for Self-Liberation?', in *Journal of Social Philosophy* 15: 1–13.

Strohminger, Nina, Joshua Knobe, and George Newman 2017, 'The True Self: A Psychological Concept Distinct From the Self', in *Perspectives on Psychological Science* 12 (4): 551–60, <https://doi.org/10.1177/1745691616689495>.

Urmson, J.O. 1969, 'Saints and Heroes', in *Moral Concepts*, edited by Joel Feinberg, 60–73 (Oxford: Oxford University Press).

Watson, G. 1975, 'Free Agency', in *Journal of Philosophy* 72: 205–20.