

### Virtue Theory, Ideal Observers, and the Supererogatory

The final publication is available at [link.springer.com](http://link.springer.com); <http://link.springer.com/article/10.1007%2Fs11098-008-9250-0>

Discussion of the supererogatory in the last half-century has been sparked in large part by J. O. Urmson's 1958 article, "Saints and Heroes".<sup>1</sup> Urmson argues that there is a class of actions – the supererogatory – that cannot be adequately accounted for by traditional divisions of actions into the obligatory, the forbidden, and the permissible. He has us consider the actions of a doctor who chooses to travel to a plague-ridden city in order to help combat the crisis, and ease the suffering of those trapped there. It seems too much to demand that all people (or even all doctors) perform such actions – we would not blame other doctors who do not make such extreme sacrifices, and do not take such extreme risks to their own well-being. The doctor's actions thus do not seem to be obligatory. On the other hand, these actions are not simply permissible. It is not as if the doctor were simply choosing to walk to work rather than taking his bicycle. His actions seem morally praiseworthy, and worthy of emulation. Thus, we have an instance of the supererogatory, a morally good action which is neither obligatory, nor merely permissible. A supererogatory action is often described as one that (in some sense) goes beyond duty or what is morally required.

Since Urmson's paper, there has been a great deal of discussion of supererogation by both consequentialists and deontologists. Urmson suggests that both theories face difficulties in accounting for the supererogatory, though he does recommend a modified consequentialism as a possible solution. A wide range of responses have been evoked – some suggesting that supererogation is simply an illusion, while others have attempted to account for it within (broadly) traditional consequentialist or deontological frameworks.<sup>2</sup>

In the first half of this paper I argue that accounting for the supererogatory is just as problematic for standard virtue theories as it is for their consequentialist and deontological rivals; common formulations of virtue ethics do not yet provide developed accounts of supererogation. In

the second half I propose and defend an alternative virtue-theoretic characterization of supererogation. I focus in particular on how we should understand cases in which what a saint (or an idealized version of a saint) considers obligatory for herself (and perhaps certain others) is seen as supererogatory by other idealized saints. I will not here address the question of whether we need the category of the supererogatory. I will simply be arguing that if we wish to embrace the supererogatory, then common virtue ethical approaches face certain difficulties, and that there might be a better alternative available.

### I

In her important book, On Virtue Ethics, Rosalind Hursthouse defends an influential virtue-theoretic account of rightness:

An action is right iff it is what a virtuous agent would characteristically (i.e. acting in character) do in the circumstances.<sup>3</sup>

Note the clause that an action is morally right only if it is what a virtuous agent would do in the circumstances (while acting in character). It seems then, that it is morally required to act as a virtuous agent would; if not, the action could not be morally right.

The problem for such an approach in accommodating the supererogatory becomes clear when we consider particular moral exemplars – perhaps Dorothy Day, Mohandas Gandhi, or Albert Schweitzer. Surely these are virtuous agents. But now it would seem that virtue theories would require us to act like these saints or heroes simply in order to perform morally right actions; we would be required to do what a saint would do in the same situation. Gregory Trianosky suggests that

The fully virtuous person is willing to do both what morality requires and what it only recommends [the supererogatory], and has whatever supporting traits are sufficient to maintain this commitment in humans.<sup>4</sup>

It should be stressed that Trianosky himself does not embrace a strict virtue ethics – he maintains that moral rightness, and the supererogatory can be defined independently of virtuous persons.<sup>5</sup> But his statement does seem to reflect what we would typically expect of a fully virtuous agent. And our problem is that a standard, ‘pure’ virtue ethics will set moral rightness at the level of action performed by the fully-virtuous, and thus at an intuitively supererogatory level. There would be no space for considering certain actions as saintly, heroic, or supererogatory – and beyond the call of duty – as moral rightness would itself require such actions.<sup>6</sup>

It might seem that we could solve the problem by simply modifying this standard virtue theoretic account of moral rightness to be one of supererogation. That is,

An action is supererogatory iff it is what a virtuous agent would characteristically (i.e. acting in character) do in the circumstances.<sup>7</sup>

This account has certain attractive features. For example, it maintains the same structure as our original virtue-theoretic account of moral rightness, and draws upon the behaviour of virtuous agents in establishing a moral standard (a feature that appeals to many virtue theorists). And it also captures our expectation that moral exemplars will not just do what is (merely) right, but will go beyond the call of duty, performing supererogatory actions.

But this standard would be too liberal. Consider – repaying a loan, and telling the truth when there are few (if any) negative consequences would both count as acts of supererogation insofar as these are actions that virtuous agents would perform in a given set of circumstances. Worse still, it would seem that sleeping when a virtuous agent would characteristically sleep would similarly count as supererogatory.<sup>8</sup> The problem for both the original proposal (treating what a virtuous agent would do as right) and the revised proposal (treating what a virtuous agent would do as supererogatory) is that some of the actions that would be performed by virtuous agents would be

supererogatory, while others would simply be permissible or obligatory. And we need a way to distinguish the supererogatory from these merely permissible or obligatory actions.

An alternative solution might lie in holding that virtuous agents need not be saints or heroes. Perhaps we simply need to be good people, but not to the extremes of saints. We could then avoid the objection that virtue theories would require us to act as saints simply in order to act morally rightly. Consider the following:

An action is right if and only if it is what a satisfactorily (non-saintly) virtuous agent would characteristically (i.e. acting in character) do in the circumstances.<sup>9</sup>

This proposal retains the basic form of the standard virtue-theoretic account of rightness, characterizing rightness in terms of the actions of virtuous agents. It also allows space for supererogation as rightness is set at the level of the actions of non-saintly virtuous agents, rather than at that of saints or heroes.

An initial hurdle for such a proposal would come in establishing who would qualify as moral exemplars. We would, after all, no longer be looking for moral excellence, just moral satisfactoriness. How good would an agent have to be in order to qualify as ‘satisfactorily’ virtuous, and what would justify appeal to this particular level of goodness rather than any other? Still, for present purposes we can simply grant that some plausible account of satisfactory virtue can be given and justified (though this is no small concession).

We would now need an account of how the actions of saints would be morally good and praiseworthy, even if they do not match the actions that would be performed by the (merely) virtuous. If we take ‘right’ in the above account to mean required then we might initially agree that the actions of saints are not right in this way – after all, supererogatory actions are taken to be optional, not required. But this account of moral rightness is rather implausible. Suppose, *ceteris paribus*, satisfactorily virtuous agents would characteristically eat when very hungry, sleep when very

tired, and act foolishly when very drunk. According to this account of rightness, it would now be morally obligatory to act foolishly when very drunk, to sleep when very tired, and to eat when very hungry. As such the account is surely incorrect. And, as argued above, we cannot simply identify the supererogatory with what would be done by moral saints; we would still need a working account of supererogation, and of the relationship between saints and the supererogatory.

If we instead take ‘right’ to mean permissible (as is common) we find that the intuitively supererogatory actions of saints would not even be morally permissible: they are not what the satisfactorily virtuous would do, and actions would be right (permissible) only if they are what non-saintly virtuous agents would do. Surely supererogatory actions should at least be permissible. Note further that if, as seems to be true, average ‘satisfactorily’ virtuous agents would be willing to inflict severe shocks on others in certain circumstances (as in the famous Milgram experiments), then it would be morally permissible to do so. If ordinarily virtuous agents would characteristically unknowingly give poisoned wine to their guests (where they have no reason to suspect any such problems; a madman has simply snuck in and poisoned the wine), then it would be morally permissible to do so.<sup>10</sup> More broadly, there is a wide range of cases where people plausibly taken to be decent or satisfactorily virtuous (but not saintly) act in ways that are clearly problematic, but on the current proposal such actions would be treated as permissible.<sup>11</sup> The proposal is thus rather dubious.

One could argue that supererogatory actions are not strictly merely permissible in this way; as such, we should not expect supererogatory actions to be subsumed under this category. But note that we would now simply arrive back at our original problem. We would lack a positive account of supererogation, and obvious alternatives (such as treating what the fully-virtuous themselves would do as supererogatory) have been shown to face significant difficulties.

We can turn to another prominent virtue ethics, that of Christine Swanton. In her Virtue Ethics: A Pluralistic View Swanton does not present a detailed account of supererogation, instead holding that this is a matter that will need to be spelt out in particular virtue theories:

depending on how demanding is one's virtue ethics, an act of returning a minor favour will be morally desirable but not obligatory, or obligatory [...] Furthermore, supererogatory acts may be seen as highly desirable, or admirable, but not obligatory, or, if one's ethics is very demanding, obligatory. [...] Virtue ethics as such is neutral on this issue.<sup>12</sup>

Swanton thus wishes to leave the issue of supererogation open, and to allow for a range of distinct virtue-theoretical possibilities. Still, we might consider ways in which we might capture the supererogatory, given her explicit account of right action.

According to Swanton, a morally right action is that action (or one of a set of possible actions) that are morally best in a given situation. More precisely,

An act is right if and only if it is overall virtuous, and that entails that it is the, or a, best action possible in the circumstances.<sup>13</sup>

She gives a somewhat different presentation in her recent "Satisficing and Perfectionism in Virtue Ethics":

An act is right if and only if it best meets overall the targets of relevant virtues.<sup>14</sup>

While such an account appears to treat rightness as highly demanding (by requiring best actions), Swanton suggests that this might not be so, once we recognize a virtue of virtuous perfectionism, a character trait that focuses on our efforts at self-improvement and striving. Her thought is that, for example, the best overall action in a given case might be for an agent to volunteer overseas with an aid organization (this would best hit the targets of the relevant virtues). But for a given ordinary agent, this might be very risky or impose tremendous costs – it is an action that might be at the very limit of her abilities. Virtuous perfectionism might then recommend not striving for such a risky,

demanding goal for the ordinary agent, instead recommending a more modest course of action (perhaps giving ongoing donations to the same group). This would allow for a gap between what is right and best for a particular agent (given her limitations and a virtuous perfectionism) and what would be overall best (for a fully virtuous agent).

Swanton thus suggests that there may be acts with the following characteristics:

- They are the best acts as measured by standards of supreme (heroic or saintly) virtue.
- They are within the agent's power (and known by the agent to be so).
- Given the circumstances of the agent, they are not mandated by requirements of virtuous perfectionism in the agent (given those circumstances).<sup>15</sup>

Could this set of acts be seen as capturing the supererogatory – best overall actions that are not required of an agent? Swanton herself does not propose this, though it does seem a plausible initial attempt to characterize the supererogatory.

Still, problems would arise for anyone who would attempt to treat supererogatory actions as equivalent to such overall morally best actions. On one hand, there are actions that are supererogatory but not the overall best (morally): perhaps giving 50% of your pay cheque to charity might be morally best, but surely giving 45% could still be supererogatory. On the other hand, there are acts that are morally best, but which would not qualify as supererogatory for a given agent insofar as things would typically go badly if she were to attempt them; as Swanton notes, overreaching one's strength in such cases might result "in various kinds of motivational distortions and bad consequences"<sup>16</sup>.

We might modify the proposal to explicitly address the two problems just raised. Thus we might suggest:

An action is supererogatory for an agent iff it is better (by the standards of supreme virtue) than the action required of this agent (given limits imposed by virtuous perfectionism), is within the agent's power, and would have a positive moral outcome.

The revised proposal thus allows that there can be supererogatory actions that are not yet best overall actions, and rules out cases where an agent's attempt to perform such a better action would have negative results. This comes quite close to an adequate account, but certain questions remain. First, what makes various acts better than what is required for the given agent, given that they can be quite different from what is overall best? Recall our example where what is overall best is to volunteer overseas. We can expect a wide range of still supererogatory actions here – giving especially large donations, devoting much time to a local aid organization (perhaps concerned with very different issues), becoming a political activist for a cause, some combination of these, and so on. While it could simply be asserted that (overall) these acts better hit the targets of the relevant virtues, there is some mystery here. What are acceptable balancings of the demands of various virtues, and how are these determined - what is it about these disparate actions that makes them all qualify as better hitting the targets of the virtues? A second issue concerns the clause that actions “would have a positive moral outcome”; what would determine whether this is the case? For example, if the agent improves the lives of others while doing some harm to his own character, would this count as a positive moral outcome? Answers to such questions may be forthcoming, but at least initially they loom rather large.

Moving on, we can consider whether we can provide an adequate characterization of supererogation drawing on Michael Slote's sentimentalist, agent-based virtue ethics. Slote has recently suggested that the distinction between right and wrong can be understood (“to a first approximation”) as

the distinction between what reflects fully developed human empathic concern for others and what reflects less than fully empathic concern for other people.<sup>17</sup>

On this account, (mere) moral rightness requires that fully-developed human empathic concern be reflected or expressed in a given action. But if so, it is not clear that we have room for the supererogatory – there are not higher levels of human empathic concern available to us (given that rightness requires fully-developed human empathic concern). In other words, moral rightness already seems to demand the highest possible level of empathic concern available to humans, so we cannot simply treat the supererogatory in terms of expressing great empathy.

Furthermore, it seems that there will be occasions where actions can quite easily be appropriately motivated by (and express) fully-developed human empathy – thus, being right in this way does not seem to be sufficient for supererogation. We might imagine a case where one could perform a significant favour for a close friend at little or no cost to oneself. On the other hand, it seems there will be occasions where actions expressing fully-developed human empathy could be extremely demanding and difficult – such that even if an action falls short of expressing such full empathy, it could still be plausibly treated as supererogatory. For example, fully-developed empathy might require (in a particular case) forgiving someone who has committed a serious crime against an agent's friend – even coming close to the proper sort of full-empathy and forgiveness could plausibly be seen as supererogatory depending on the nature of the crime, and so forth. If this is correct, then expressing fully-developed human empathic concern for others would not be a necessary condition for supererogation. And we would thus need to develop a very different account of supererogation for Slote, one not grounded in any obvious or straightforward appeal to fully-developed human empathy.

We can approach the problem for virtue theories from another angle. Consider a doctrine embraced by many virtue theorists – Aristotle's doctrine of the mean.<sup>18</sup> Roughly, according to this

doctrine, a virtuous person will pick that action which is a mean between flawed alternatives, some ‘excessive’, some ‘deficient’. Thus, a virtuous person will choose the correct, courageous action, one which is neither cowardly (a deficiency) nor foolhardy (an excess). But how then are we to allow room for the supererogatory? There is a mean – this is what we are required to strive for, just as utilitarians require us to maximize happiness. Anything departing from the mean is to that extent flawed. So in what way can we ‘go beyond the mean’? If we stray from it, the action is morally flawed, reflecting an excess or deficiency.

Of course, as Robert Johnson has recently suggested (following Aristotle), the best thing that we can do in certain cases while we are moral novices might be to stray from the mean.<sup>19</sup> If we have a tendency to the vice of cowardice, it might be best for us to perform actions that fall on the foolhardy side of courage while we are learning. Here we can allow that straying from the mean is right, but that this still reflects an imperfect character. That is, the very fact that in the novice’s case straying from the mean would be right shows that the novice is not yet fully-virtuous. In any case, this does not yet help us in accounting for supererogation.

We might hold that achieving the mean itself is supererogatory; this would be akin to the previous suggestion that performing actions that a fully virtuous person would perform is supererogatory. But this move would be subject to parallel counterexamples, requiring us to claim that many actions that seem merely required or permissible are supererogatory. After all, in some cases determining the mean is fairly obvious, and the action easily performed; we would not want to count such cases as supererogatory. In any event, we would now be left in need of an account of what actions are morally right – and in what cases we are allowed to deviate from the mean.<sup>20</sup>

There is a shared, underlying problem that will need to be overcome for each of the approaches to virtue ethics that we have considered thus far. Consider: some of the actions performed by the fully-virtuous will be merely right, others supererogatory; some morally best

actions (hitting the targets of the virtues) will be merely right, others supererogatory; some actions expressing fully-developed human empathy will be merely right, others supererogatory; and some actions which achieve the mean will be merely right, and others supererogatory. The underlying issue to be addressed, then, is how are we to distinguish between those actions that are merely right on these virtue-based approaches, and those that are supererogatory?

Our discussion has been somewhat cursory, and certainly virtue theorists could attempt to modify the standard accounts in ways beyond those suggested here. And it should be stressed that Hursthouse, Swanton, and Slote have not explicitly set out to provide accounts of supererogation in their recent work. But it seems clear that virtue theories in familiar forms face initial difficulties akin to those faced by consequentialists and deontologists in accommodating the supererogatory, and that simple modifications of these prominent virtue-theoretic accounts of rightness will not adequately capture the supererogatory. As such, we have some reason to consider alternative virtue-based approaches.

## II

Consider the following characterizations of the supererogatory and the morally required:

(S) An action is morally supererogatory for an agent in a given set of circumstances to the extent that fully-informed, unimpaired, virtuous observers would deem the action to be supererogatory.

(R) An action is morally required for an agent in a given set of circumstances iff all fully-informed, unimpaired, virtuous observers would deem the action to be so.

Intuitively, the proposal suggests that we determine the status of an action by appealing to the appraisals of the ideally virtuous. They consider the individual involved (and her roles, her obligations, her capacities, etc.) and make an appraisal relative to her. Thus, an agent need not act precisely as some virtuous agent would in order to act supererogatorily (or even rightly). Nor, on the

other hand, would merely sleeping in circumstances in which a saint might sleep qualify as supererogatory. What matters is the extent to which idealized saints or heroes – virtuous ideal observers – would deem her actions supererogatory, given full-information about the circumstances, the agent, her motivations, and so on. Importantly, the proposal treats supererogation as a matter of degree; actions are not simply supererogatory or not. We will see shortly that this is a plausible feature of the account.

Before turning to that, we can briefly consider the traits of the observers. With full-information such observers would be familiar with the concept of the supererogatory (and rightness, etc.), able to understand the motives of the agents, understand the relationships and commitments of those involved, accurately predict the consequences of various courses of action, have an awareness of the affective states of those involved, the nature of any suffering that would result from given actions, and so on.<sup>21</sup> Thus, they would not lack information crucial to making good assessments – information that may not be available to an agent immersed in a situation.

They would also have fully virtuous characters to draw upon in interpreting and assessing the action before them. Ultimately, of course, we would need an account of the virtues in order to pick out the virtuous. This project lies rather beyond the scope of the current paper. Still, we might appeal to standard accounts of the virtues embraced by virtue theorists. Thus we might hold that the virtues are those traits that lead to and are (partially) constitutive of human flourishing.<sup>22</sup> If there are different ways to lead a flourishing human life, we could allow for a range of virtuous observers.

Finally, such observers must be unimpaired – they must not be coerced, distracted, under the influence of drugs which diminish mental acuity, and so on. The presence of any of these sorts of impairments could clearly lead to questionable judgments. Thus, the current position can be seen as blending elements from virtue and ideal observer theories.

Are these accounts circular? Focusing on (S), we might wonder about the basis upon which a given virtuous ideal observer would deem an action supererogatory, and worry that she would need to do so by determining to what extent virtuous ideal observers would deem the action to be supererogatory. This would be viciously circular: she herself would be one of the ideal observers at stake, and she would thus need to know whether she deems an action to be supererogatory precisely in order to determine whether she would deem that action to be supererogatory. In response, note that a virtuous ideal observer need not appeal to (S) or (R) as decision-procedures in assessing actions. That is, she would not think in terms of the judgments that she and other virtuous ideal observers would make when she deliberates on whether to deem an action supererogatory. Rather, she would focus her attention on the actions themselves. To deem an action supererogatory would be (to a first approximation) a matter of judging the action to be one that the observer would want to be performed by an agent in virtue of its morally-relevant aspects<sup>23</sup>, but one where the observer would not want the agent to be blamed or punished for a failure to perform the action as such (insofar as she judges the action to be morally better than what she would demand of agents in similar circumstances), *ceteris paribus*.

Notice how the current proposal allows us to solve the underlying problem raised earlier for prominent virtue theories. We can appeal to the attitudes of virtuous ideal observers themselves to distinguish between actions that are morally right (required, or permissible, etc.), and those that are supererogatory. The virtuous would distinguish those of their actions (or the actions of others) that are merely required and those that are optional/supererogatory; they would react to some morally best actions as merely required, others as supererogatory, and so on. Of course a significant question here concerns the basis upon which virtuous ideal observers would distinguish these classes of actions. This lies somewhat beyond the scope of the current paper, but do note that virtuous ideal observers could, at the very least, appeal to those characteristics that we often take as relevant

to the status of actions as supererogatory – such things as the costs or risks they impose on the agent, past actions or sacrifices on the part of the agent, the personal psychology of the agent (for example, if agents have certain phobias, actions that might be merely right for others could well be supererogatory for these agents), and so on. But crucially, it would be the weighing and assessment of such factors (and the resulting attitudes) by virtuous ideal observers, given their preferences, desires, and so forth that would ultimately determine whether a given action is supererogatory.

### III

The current proposal holds that the supererogatory is a matter of degree along two dimensions. First, there is a fairly non-controversial dimension, though one which is not immediately obvious from (S). Intuitively, if we think of supererogatory actions as ones going beyond duty, then we can of course go a little beyond duty, or we can go quite far beyond duty. Perhaps giving 50% of your salary to charitable organizations, given your circumstances, is supererogatory. It could then be supererogatory to a higher degree to give 55%. Of course, this is not to say that there are not limits. There could be a point – saying giving 80% of your salary – where this shifts from supererogation to foolishness. Still, this dimension is fairly clear and well-entrenched in common sense.

The second dimension draws on the fact that virtuous ideal observers can disagree over the claim that a given action is supererogatory. In a given case we might have a small number of virtuous ideal observers who deem the action supererogatory, while the majority consider it as simply required. Or in some cases almost all will view an action as supererogatory while only one or two would deem it required. That is to say, there may be a wide range of virtuous observers, and even given full information, they may disagree on whether an action is supererogatory or simply required. Our proposal is to treat supererogation as matter of degree along this dimension, so that an action is supererogatory to the extent that virtuous ideal observers would deem it so. If all

virtuous observers would deem the action supererogatory, then we can say that the action is universally supererogatory:

(FS) An action is universally morally supererogatory for an agent in a given set of circumstances iff all fully-informed, unimpaired, virtuous observers would deem the action to be supererogatory.

We are thus given two endpoints. First it could be that all virtuous ideal observers would deem a given action supererogatory, making it universally supererogatory. Second, it could be that all virtuous ideal observers would deem a given action required, making it entirely unsupererogatory. In between, it will be a matter of degree.

This is, of course, somewhat awkward, and we might have hoped to have a straightforward yes or no answer to the question of whether a given action is supererogatory. It might be suggested that the apparent haziness of whether a given action is required or beyond the call of duty is simply an epistemic matter. In each case the action is required or supererogatory tout court; we may simply have difficulty in certain cases of determining which is which.

But treating supererogation as a matter of degree along both dimensions does capture important commonsense intuitions about the supererogatory. For example, we can consider cases in which a person is faced with the possibility of sacrificing his life to save the lives of others. If the action is to save one life, we might feel his action is supererogatory. But if the action is to save two million lives, we might feel that the action is morally required (even if highly demanding). The details of the case are not essential – the key point is that we sometimes believe that morality can make very severe demands, and that different people can have different intuitions about whether an action should be understood as supererogatory or simply as fulfilling a very demanding requirement.

Furthermore, there are different, irreducible sets of moral ideals. Even given full information, it is not at all clear that the virtuous would agree on all cases. There are different ways

of leading good moral lives, and some will value benevolence slightly over honesty, others will value honesty over benevolence, and so on. We expect there to be occasional disagreements even among virtuous people. And again, it seems unlikely that these differences in ideals and intuitions will all vanish in light of full information. Given such variations, it would be best to treat supererogation as a matter of degree – with actions being more supererogatory to the extent that more virtuous observers would deem them so. While we might have hoped for simple answers, it seems that the only way to do so is to arbitrarily select one set of moral ideals and then treat this set as establishing a standard for all. The current proposal avoids such arbitrariness, and recognizes the possibility of a range of viable moral ideals.

#### IV

The proposal allows us to address the following problem. Many saints appear to consider what they do as simply morally required of them – including actions which seem supererogatory to others (consider, for example, the Trocmes and other citizens of LeChambon). Sometimes they hold that such actions are also the duty of others, but often they do not – they often see themselves as having a special vocation or calling that requires more of them than of others. In many such cases we wish to say that their actions are in fact supererogatory, but then this suggests that they must be mistaken in their belief that their actions are simply required of them. And Susan Hale has noted that this is a strange result, to the extent that we think of saints as exemplary, insightful moral agents.<sup>24</sup> Are such saints and heroes simply confused about the status of many of their actions?

An initial point we can make, given the current account, is that to the extent that a saint would still deem her action simply required, even when she is suitably idealized, to that extent her action is not universally supererogatory. Recall the characterization of a universally supererogatory action given earlier. The very fact that an ideal version of a moral saint would see the action as the actual saint's duty changes the overall status of the action. So the actual saint is mistaken to the

extent that she sees the action as absolutely required of her, but she is correct that the action is not universally supererogatory (precisely because of her attitude, once incorporated into a virtuous ideal observer). As such, the attitudes of saints who consider their actions as duties are taken into account, and do influence the overall status of their actions, even while their actions remain to some extent supererogatory.

What else can we say of cases in which an idealized saint would deem an action to be morally required of her ordinary saintly self (and perhaps certain others), but other virtuous ideal observers would deem it supererogatory? A crucial point is that the ordinary or saintly person involved could still rationally feel shame for failing to perform such an action. We can recognize that an agent may choose to live up to a certain standard – for example, Claire might try to live in a way that an ideal St. Francis would endorse. Relative to that standard, she falls short, and so could blame herself, but given that there are other virtuous ideal observers who see the action at stake as supererogatory, we could not say that she failed to perform a morally required action.

This possibility is important in addressing certain concerns about embracing supererogation. It might seem that if an action is not morally required, then there is no need to feel shame or guilt for failing to perform it. David Heyd, for example, suggests that

The decision to act beyond what is required is free not only from legal or physical compulsion, but also from informal pressure, the threat of moral sanctions, or inner feelings of guilt. It is purely optional. Such a freedom allows for the exercise of individual traits of character and for the expression of one's personal values and standards of moral behaviour. Being purely optional, the supererogatory act is spontaneous and based on the agent's own initiative.<sup>25</sup>

And this may seem to make morality too minimal and undemanding – we could simply turn our backs on these good but optional actions with a clear conscience. Indeed, as Heyd suggests, it might

even be thought irrational to feel guilt for failing to perform supererogatory actions, given that they are not actually morally required. Such a view could encourage moral mediocrity (or so one might worry).

In response, we need to recognize that we as agents embrace moral ideals and strive to live up to them. To the extent that we fail to live up to them, we can rationally feel shame for such failure. Still, punishment typically would not be appropriate in these cases, to the extent that our action is not required by all moral ideals.<sup>26</sup> These claims are obviously compatible. We can regret our failures to perform supererogatory actions insofar as these would be deemed required by our chosen moral ideals, even while others cannot demand such actions from us. Compare other cases where we have undertaken worthwhile personal commitments: surely we can be disappointed with ourselves, and indeed feel shame about our failures with respect to such commitments even if they are not universally required.<sup>27</sup>

On the other hand, note that all virtuous observers will agree on and surely require that at least certain basic standards be met by all agents. It is hard to imagine any fully-informed, unimpaired virtuous observer declaring a given agent's refraining from killing a person out of spite to be supererogatory... Intuitively, even while we expect some variation in the judgments of the virtuous in particular cases, we also expect universal agreement across a wide range. There are different sets of moral ideals embodied in different virtuous ideal observers, but there will be a shared core of actions that will be seen as required, regardless of these variations. Thus, it certainly seems that there will be a wide, commonsense range of actions that no virtuous ideal observer would deem supererogatory (and which would instead be required). This gives us a shared core of morality, while allowing for people to strive for viable moral ideals that need not be universally binding.

Susan Hale draws attention to cases in which a saint might hold other persons to high standards, and declare many actions morally required (for these people) that other virtuous ideal observers would see as supererogatory:

Witness Gandhi's demands upon his family and followers and his stern reproaches when they failed to meet his demands.<sup>28</sup>

What shall we say in cases where Gandhi's followers fail to perform an action which would be deemed required by an idealized Gandhi, but where other ideal observers would see the action as supererogatory?<sup>29</sup>

Prior to discussion of such cases, it will be helpful to briefly examine moral shame and guilt. With a slight modification, we could endorse Jennifer Manion's treatment of moral shame:

Moral shame is shame precipitated by some moral lapse, failure or omission that results in an agent's disappointment in aspects of her own moral character about which she has some significant control. As such, it is less shame about transgressing some rule or violating some sanction, but rather about failing to meet or approximate one's own moral ideal.<sup>30</sup>

We can depart slightly from Manion in holding that moral shame can be primarily a result of (and thus focused-upon) transgressing some rule or violating some sanction, particularly in cases where an agent commits herself to a particular moral code with a range of rules that she sees as binding upon her, though they are not binding upon others. She holds herself to a higher or distinctive set of moral rules, and her failure to live up to them can properly induce moral shame.

Guilt is typically held to be more 'other-focused' than shame. Bernard Williams writes "What I have done points in one direction towards what has happened to others, in another direction to what I am. Guilt looks primarily in the first direction [...] Shame looks to what I am."<sup>31</sup> Manion adds that "Usually, one's feeling of guilt concerns a rule or rule-like constraint that one has broken, the harm that has ensued, and the people affected by the harmful act."<sup>32</sup> Again, we can

accept this general characterization, though we should emphasize (with Williams) that guilt focuses especially on our impacts upon others, and not necessarily on the breaking of a moral rule.

Returning to Gandhi and his followers, we can propose three understandings of the case, depending on how certain details are fleshed out. First, if the followers involved have reached great heights of virtue and have firmly committed themselves to living up to a saint's standards (or an idealized version thereof), then all virtuous ideal observers could agree with the idealized saint that the followers should feel moral shame – and possibly guilt<sup>33</sup> - in those cases in which the idealized saint deems that the followers have fallen short of what is morally required of them. After all, his followers are (in this scenario) capable of such actions, and have voluntarily committed themselves to performing such actions. On the other hand, the virtuous observers could at the same time still deem these actions to be supererogatory – there is not an overall binding moral requirement to perform these actions; they are not actions required by all moral lives. Instead, the followers simply fail to live up to personally chosen commitments – a particular way of living morally.

For the second understanding, assume that in fact the saint's followers have not reached very high levels of virtue, but are capable of reaching these levels. Assume also that their current lower level of virtue is not due to culpable actions or patterns of behaviour; it is simply taking time for them to develop higher levels of virtue.<sup>34</sup> In such a case, other virtuous ideal observers could maintain that his followers can rightfully feel disappointed with themselves if they fail to perform what is considered morally required by their chosen standards, but should not feel moral shame. That is, given that they lack the required levels of virtue, shame is unwarranted (they are not currently capable of acting at the high level demanded). On the other hand, they can rationally feel disappointed with themselves for not acting in this way, and for not yet having the required character. And overall, the action would still be supererogatory (should they manage, on a given

occasion to perform it); the demands of their particular commitments go beyond what would be universally morally required.

Finally, for the third understanding, assume that the followers, while committed, simply lack the ability to reach high levels of virtue. They do not have the capacity to reach high levels of self-discipline, compassion, and so on. Again assume that this current inability is not due to culpable actions or patterns of behaviour in the past. This case would clearly be one in which the actions would be supererogatory for the agents involved (should they somehow manage on some occasion to perform such an action). We could hold that virtuous ideal observers would maintain that the agents should not feel ashamed, but might rationally feel disappointed, though in a different fashion than that considered in the second scenario. Here the followers could feel disappointed with themselves, but rather than striving harder to reach the saint's heights (as in the second case), they should perhaps modify their goals, and strive to achieve more realistic levels of virtue (given their own potentials for character development).

Consider the following analogy – an athlete wishes to play in a top-level professional league, and fails to make any team. If he is simply lazy and does not train as hard as he knows he should, while he has all the necessary ability, coaching, etc. it seems we could all see his action as problematic, given his commitment to this goal and his ability to achieve it. We can rationally hold as observers that he should feel (non-morally) ashamed of his failure, even while we hold that playing in a top professional sports league is not in any way universally required.

If he has the capacity to play at such a level, but just has not yet had sufficient training, etc., (so that it is extremely unlikely that in a game held now that he could play at an adequate level) we can hold that he should feel disappointment over his failure, but not shame. If he is not securely capable of the required actions yet, he has no reason to feel to shame. Still, he should be disappointed and strive harder, again even if his project is not required. Finally, if he simply lacks

the necessary body type, or other necessary conditions to reliably compete at this level (except, perhaps, for occasional moments through extreme good luck), he can rationally be disappointed, but he should also change his commitments to ones he can attain (given adequate effort, etc.).

The athlete case is described as an analogy for the following reason: it could be argued that a commitment to morality (in general) is not optional for any rational agent. We may choose not to have any athletic commitments, but we are all bound by morality. In the present case, what this means is that while we need not live up to the standards of any one particular ideal character type, we must still try to live up to at least some such ideal – we have our choice. We should strive to live up to the level of the virtues we have attained, and strive to improve these levels, to the best of our abilities. But note here also that the current proposal does not assume that all people can become moral saints. Just as we do not all possess the potential to be outstanding athletes, brilliant mathematicians, or (musical) virtuosos, it seems plausible to hold that we do not all have the potential necessary to develop saintly degrees of moral virtue. Judgments of our actions must be sensitive to our potentials.<sup>35</sup>

In addition, we must acknowledge that there are ranges of virtuous lives and standards. This is not to embrace an extreme moral relativism. Rather, it is to acknowledge that there is a wide variety of moral saints and corresponding standards for good lives and actions. Gandhi is quite different from St. Francis, who is in turn quite different from Schweitzer, and so on. And it seems unlikely that they will all come to identical judgments about all particular cases, even with full information. Despite this variation, a wide range of actions will presumably be deemed to be required by all ideal virtuous agents. On the other hand, there will be cases in which different virtuous agents will disagree on whether an action ought to be considered morally required – these will be supererogatory to some degree (assuming that this is where the disagreement lies).

Thus, the current proposal allows us to make sense of cases in which an ideal virtuous observer considers some action to be morally required of an agent, while others do not. The agent should still rationally feel shame or disappointment if she fails to perform actions demanded by her chosen standards (and corresponding virtuous ideal observer), given her commitment to a particular way of living morally, and her ability to do so. We can make sense of an agent who, in performing an action that others would deem supererogatory, would claim “I’m simply fulfilling my duty”. A person with a particular set of moral ideals can be understood as having self-imposed duties of this sort – where she is rational to feel shame for failure, but where these self-imposed ideals and duties are not full-fledged moral requirements that all virtuous ideal observers would judge as required of an agent in the given circumstances.

On the other hand, these actions are still supererogatory to at least some extent, given that they are not required by all ways of living morally; that is, there are other virtuous ideal observers who would deem her action supererogatory. She has not failed to meet any universally binding moral requirement. And in this way then, we have a viable framework for a virtue-based account of supererogation that allows for a plurality of reasonable moral standards and good moral lives.

---

<sup>1</sup> Urmson (1958).

<sup>2</sup> For deontologists against the possibility of supererogation, see Hale (1991), and Baron (1995). For a deontologist accommodating the supererogatory see McCarty (1989). For consequentialist accommodations of supererogation, see Zimmerman (1993), and Portmore (2003). Finally, for a denial that consequentialism can accommodate supererogation, see McConnell (1980).

<sup>3</sup> Hursthouse (1999, p. 28). The same basic account is presented in Oakley (1996), and Zagzebski (1996, p. 235). For a somewhat modified version of the account see Tiberius (2006).

<sup>4</sup> Trianosky (1986, p. 31).

---

<sup>5</sup> On a strict virtue ethics, judgments about character or virtue are seen as basic or primary, in that concepts of moral rightness or goodness are to be explained and understood in terms of character or virtue.

<sup>6</sup> I will treat ‘saint’, ‘hero’, and ‘fully virtuous agent’ as largely interchangeable for purposes of this paper, though obviously there are other contexts in which these ought to be distinguished.

<sup>7</sup> Brännmark defends such a proposal. See his (2006, p. 596).

<sup>8</sup> It might be objected that no virtues would be involved in choosing to sleep at a given time, and thus we need not count such behaviour as supererogatory. But minor exercises of virtues (such as prudence) could easily be involved. And the problem is that even if some slight exercise of prudence were involved, sleeping when a virtuous agent would characteristically do so hardly seems supererogatory (barring a highly-gerrymandered scenario).

<sup>9</sup> Brännmark defends a similar view, holding that an action is morally wrong iff it is one that no decent person would characteristically do in the circumstances. See Brännmark (2006, p. 596); decent people are not saints on his account. Brännmark in turn suggests that an action would be obligatory iff it would be wrong not to do it. Thus, an action would be obligatory iff all decent people would characteristically do it in the circumstances. Finally, he holds an action would be permissible iff some decent person would do it in the circumstances. The worries developed in the following paragraphs apply to these proposals, mutatis mutandis.

<sup>10</sup> We could hold that such agents are not morally blameworthy, given their non-culpable ignorance (they had no reason to suspect the wine was poisoned). But this is quite different from treating the actions themselves as morally permissible.

<sup>11</sup> For a range of such cases, drawing upon work in social psychology, see Doris (2002). See also Kawall (2002).

---

<sup>12</sup> Swanton (2003, p. 241).

<sup>13</sup> Swanton (2003, p. 239).

<sup>14</sup> Swanton (2004, p. 177).

<sup>15</sup> Swanton (2004, p. 187).

<sup>16</sup> Swanton (2004, p. 182).

<sup>17</sup> Slote (2003, p. 133).

<sup>18</sup> Aristotle (1985, Book II). For contemporary defenders of the doctrine of the mean, see Curzer (1996), Kurschwitz (1986), and Martin (1994).

<sup>19</sup> See Johnson (2003).

<sup>20</sup> Aristotle might have allowed for a solution along these lines. He writes: “Still, we are not blamed if we deviate a little in excess or deficiency from doing well, but only if we deviate a long way, since then we are easily noticed. But how far and how much we must deviate to be blamed is not easy to define in an account” (1985, II, ix, 1109b18-22). This would appear to give us space for the supererogatory – cases in which it is not blameworthy to be off of the mean. Still, Aristotle does not expand on this, and seems to have in mind epistemic worries (i.e., that it can be hard to tell exactly what is required of us), rather than allowing for action that goes beyond what is required.

<sup>21</sup> One could treat full-information as consisting in knowledge of all facts relevant to the given case. See, for example, Carson (1984, p. 58). Alternatively, one could simply attribute omniscience (or perhaps all possible true beliefs) to the observer, thereby avoiding questions of relevance.

<sup>22</sup> This is not to endorse such accounts (indeed, they face important difficulties), but they are familiar and can serve adequately for immediate purposes.

<sup>23</sup> That is, the observer would not want the action to be performed simply because of its aesthetic features, etc.

---

<sup>24</sup> See Hale (1991, pp. 279-81).

<sup>25</sup> Heyd (1982, p. 175).

<sup>26</sup> ‘Typically’ as there may be circumstances where punishment might be appropriate. For example, an individual might be a member of a religious order with a certain set of moral ideals (which would not be seen as required by most virtuous ideal observers), where there are established punishments for those members who fail to live up to these ideals.

<sup>27</sup> Gregory Mellema argues for a class of actions, the ‘quasi-supererogatory’, which are non-obligatory, praiseworthy to perform, but also blameworthy to omit. See his (1991), and (1996). For other authors who embrace the possibility that a failure to perform certain supererogatory actions is blameworthy (or suitable for feeling shame), see Badhwar (1985), and Humberstone (1974).

<sup>28</sup> Hale (1991, p. 282).

<sup>29</sup> Note that an idealized Gandhi could consider certain other actions to be supererogatory even for his followers – he is not restricted to only embracing morally required actions for his followers. Still, the more challenging cases are those where a particular virtuous ideal observer would deem certain actions to be required while other virtuous ideal observers would deem them to be supererogatory.

<sup>30</sup> Manion (2002, p. 77).

<sup>31</sup> Williams (1993, pp. 92-3). The passage is quoted in Manion (2002, p. 76).

<sup>32</sup> Manion (2002, p. 76).

<sup>33</sup> If, for example, the moral ideal at stake requires tremendous acts of charity, and an agent fails to live up to these rules, she might appropriately feel shame (for her failure to act in accordance with the rules, thus failing to live up to her ideals), and guilt (as she thinks of the continuing suffering of those she has failed to assist in failing to live up to her ideals).

<sup>34</sup> For discussion of issues in character development and culpable vice, see Jacobs (2001).

---

<sup>35</sup> These comments fall broadly in line with similar suggestions to be found in recent work by others. Mellema emphasizes the possibility of agents having greater responsibilities in areas in which they have a vocation in his (1991, chapter 5). Flescher argues more broadly for a duty to develop ones virtues in his (2003), as does Swanton in her (2004). See also Norton (1988).