SELF-PERFECTION, SELF-KNOWLEDGE, AND THE SUPEREROGATORY

KATHARINA NAUMANN

Justus-Liebig-Universität Gießen, Germany
Georg-August-Universität Göttingen, Germany
katharina.naumann@uni-goettingen.de

ABSTRACT
Supererogation seems to be an important concept of common sense morality. However, assuming the existence of such a category seems to pose a serious problem for Kantian Ethics, given the all-encompassing role of duty. In fact, Kant seems to deny the possibility of such acts when he states in the second critique that “[b]y exhortation to actions as noble, sublime, and magnanimous, minds are attuned to nothing but moral enthusiasm and exaggerated self-conceit; [...] they are led into the delusion that it is not duty [...] which constitutes the determining ground of their actions [...].” (KpV, AA V, 84f.). This paper’s aim is to show that even though it is not possible to include a category of supererogation within Kantian Ethics, the recognition of alleged supererogatory acts is not only a source of self-deception, as Kant seems to fear, but can also serve as a source of self-knowledge. Starting from the premise that within Kantian Ethics self-knowledge is indispensable to aspire to the duties of moral self-perfection, I will argue that supererogation is only a plausible concept regarding the observer-perspective: thus it describes an act that, given the observer's own moral development, seems to be beyond duty but for the agent herself is not. Given an observer has the insight that her judgment relies on her own deficiencies to act accordingly, this fulfills an epistemic function in the process of developing one’s own moral capacities. This seems to be a fairly proper phenomenological account of supererogation as it can, on the one hand, explain the common moral intuition that certain acts are supererogatory. While on the other hand, it can capture the empirical observation that agents of alleged supererogatory acts usually do not perceive their actions as supererogatory because the differing perspectives of the agent and the observer constitute an integral part of the given account.

KEYWORDS
Supererogation, kantian ethics, self-knowledge, moral development, exemplars

Thinking of the parable of the good Samaritan, the stylization of persons like Mahatma Gandhi, extensive media coverage of heroic actions or charity events,
praise for acts of forgiveness or honorary offices – it seems we all know paradigmatic narrations of people going the second mile and actions that are commonly assumed to go beyond the call of duty.

These phenomena form an important part of our moral experience and thereby seem to fulfill two apparently contradictory functions: on the one hand, they work as role models of moral perfection. The recognition they receive seems to imply that it is worthy to follow them and, therefore, point to an expansion of moral demands. On the other hand, such narrations often seem to serve the purpose of justifying one’s own moral failure by stating that such an action would have been beyond legitimate expectation, and, therefore, they point to a limitation of moral demands.

This putative tension is not only a practical problem, but, in fact, it poses a challenge for moral theory. Following a by now well-known assumption, one can accommodate these phenomena by claiming that the acts under consideration constitute a distinct ethical category, namely the supererogatory, referring to acts that are, in short, morally good but not required. Assuming the existence of such a distinct category of actions does not only seem to reflect on and account for an important concept of common sense morality, but is philosophically interesting as it promises to contribute to questions about the scope of morality as well as to matters of moral reasoning, moral psychology and moral epistemology.

But doing so seems to pose a serious problem for Kantian Ethics, given the all-encompassing role of duty. Therefore, it is highly controversial, even among Kantians, if the concept of supererogation is compatible with his writings or whether he can at least give an alternative explanation of the phenomena.

In the following I try to solve this tension by outlining an account of supererogation that integrates the supererogatory not as a basic ethical category, but rather as a useful concept of moral anthropology. Starting from the premise that within Kantian Ethics the duty to perfect oneself morally is the most fundamental and that self-knowledge is indispensable to aspire to these duties, my aim is to show that the proper recognition of supererogatory acts can serve as a source of self-knowledge.

Therefore, I will proceed as follows: first, I will point out what I take to be a quite common definition of the supererogatory and the challenges it is confronted with. Second, I will outline the problem that evolves for Kantian Ethics and explain why I believe the existing solutions to it are not sufficient. Against this background I will in a third step sketch my alternative proposal, connecting self-Perfection, self-Knowledge, and the supererogatory. Finally, I will briefly point out why this constitutes a fairly proper phenomenological account of supererogation.

1. SUPEREROGATION
Despite the prevalence in common sense morality, the supererogatory has been largely neglected within modern moral theory at least until the middle of the last century.

In his seminal paper “Saints and Heroes”, which is commonly taken to be the starting point of the debate, James O. Urmson claims that most moral philosophers only discriminate three types of action, namely the obligatory, the permissible and the forbidden, but that

“this threefold classification [...] is totally inadequate to the facts of morality; any moral theory that leaves room only for such a classification will in consequence also be inadequate.”¹

Instead he claims that

“We have to add at least the complication of actions that are certainly of moral worth but that fall outside the notion of a duty and seem to go beyond it, actions worthy of being called heroic or saintly. It should indeed be noted that heroic and saintly actions are not the sole, but merely conspicuous, cases of actions that exceed the basic demands of duty [...].”²

Following this line of thought, theories of supererogation are built upon the idea that in morality we deal with two sorts of standards, a deontic and an evaluative, which are not necessarily congruent.

Thus the definition of supererogation consists of at least two necessary conditions:

(1) Moral Worth: Performing such an act is not morally indifferent, but rather morally good, whereas its omission is not morally blameworthy.

One can here notice an asymmetry as opposed to obligatory acts, where one would usually think if performance of a certain act is good, omission is bad and vice versa. But this asymmetry can be found in disjunctive or imperfect duties as well, even though entailing a qualification (something like: if alternative means are taken), which supererogatory acts supposedly do not include. Hence the second criterion cannot only be stated as the acts not being required, but rather:

(2) Optionality: Performing such an act is completely permissible, i.e. it does not fulfill an obligation nor is it forbidden.

Whether (1) and (2) are sufficient is discussed controversially. There are at least three further conditions that are often named. And even though there is considerable discussion whether these are necessary conditions of a definition of supererogation, it is at least widely accepted that these are characteristics that often come along with supererogatory acts:

¹ Urmson 1958, 198.
² Ibid., 205.
(1) Merit: Performing such an act is meritorious and therefore praiseworthy.³

(2) Sacrifice: Performing such an act involves (great) sacrifices for the agent.⁴

(3) Altruism: Performing such an act is intended to contribute to the well-being of others.⁵

Assuming the existence of a category of supererogation does, however, lead to two serious problems which every account of supererogation must resolve: the “paradox of supererogation”⁶ and the “heroism paradox”.⁷

The first paradox deals with the question whether the supererogatory can be a plausible category after all. This argument is prominently stated by Joseph Raz:

“One important characteristic of supererogatory acts (though not only of them) is that their performance is praiseworthy while their omission is not blameworthy. But this creates a problem. If doing a supererogatory act is praiseworthy there must be reasons for doing it, and the reasons must outweigh any conflicting reasons for not doing it. But if there are conclusive reasons for performing the act then not to perform it is to act against the balance of reasons. If reason requires that the act be done then surely one ought to do it, and the “ought” is based on all the reasons which apply to the case; it is a conclusive ought. But this entails that failing to perform the act is failing to do what one ought (conclusively) to do, so why isn’t it blameworthy not to perform a supererogatory act?”⁸

³ Cf. Feinberg 1961, 280f.; Heyd 1982, 115; Raz 1975, 164. However this condition leads to the question whether supererogation isn’t better understood as a feature of agents instead of acts (cf. Montague 1989, 101ff.).

⁴ It seems to be more often implicitly than explicitly assumed that supererogatory acts involve costs on the part of the agent. This view seems to rely on the assumption that supererogatory acts are not required precisely because they surpass what can be reasonably expected of a person, at least when it is claimed that they necessarily include great sacrifices (cf. McGoldrick 1984, 525). Thus moderate versions of this claim only hold that there are at least some costs for the agent, even though they might be minor, and are therefore able to include favours and the like into the category of supererogation (cf. Feinberg 1961, 280ff.). Others contest the necessity of such a condition altogether (cf. Archer 2016).

⁵ This is claimed explicitly in condition (3) and (4) of Heyd’s definition (1982, 115), but is oftentimes implicitly assumed. Nevertheless it has been criticized for two reasons: On the one hand Mellema (1991, 22) claims that “acts of supererogation can be performed in such a manner that intended good consequences or benefits to others play no role whatsoever.” The example he has in mind is that of someone held in prison by political terrorists, who “as a man of high principle, […] is simply unwilling to renounce allegiance to his own government.” (ibid., 19f.) On the other hand Kawall (2013, 490) claims that there can be self-regarding supererogatory acts, such as this one: “A farmer is held prisoner in a fascist state. She has committed no crimes […]. Life in the prison is hard, and there is no indication that she will be released or even receive a trial. Still […] her situation is not entirely nightmarish. […] She decides to risk her life and attempt to escape.”

⁶ The term refers to Horgan/Timmons 2010.

⁷ The term refers to Archer/Ridge 2015.

⁸ Raz 1975, 164.
The second paradox deals with the problem that following the testimony of the agents of such actions, they, in contrast to the observers, usually do not conceive their actions as supererogatory but rather as obligatory. This gap is particularly striking with regard to heroic and saintly acts. So, the question is how to deal with these testimonies or, in other words, how to make these differing perspectives intelligible.

2. SUPEREROGATION AND KANTIAN ETHICS

At first sight Kantian Ethics might not seem as a suitable candidate to analyze the concept of supererogation. Due to the prevailing assumption that it represents the prototype of deontological ethics, it seems that it is conceptually impossible to integrate the supererogatory. Thus Richard McCarty states:

“All ethical theory which, following the Kantian turn, derives moral goodness from obligation, instead of the other way around, precludes in its very foundation the possibility of morally good, non-obligatory actions.”

In particular the problem seems to be the following: for Kant an action can only have moral worth if it is done from duty, but one can only act from duty when acting in accordance with the moral law. Therefore, every morally worthy action must be acknowledged as obligatory and cannot be merely permissible.

It is exactly this idea that has lead to two often articulated objections, namely that Kantian Ethics is formalistic and that it is rigoristic. Or, in other words, it is, on the one hand, criticized for not offering practical moral ideals and, on the other hand, for being too demanding. Therefore, Kantian Ethics has to deal with the two challenges supererogation promises to resolve so that it seems to be all the more attractive to examine whether there is a chance for integrating this idea. Due to a rising interest in recent years in the Doctrine of Virtue, these objections have oftentimes been rejected by claiming that one needs to take the full picture of Kantian Ethics into account. Because there Kant does not remain formal, but rather introduces material duties and challenges the limits of duties, some have argued that this might be the point of departure for this question.

But let us have a look into the ambiguous textual evidence first: Kant hardly addressed the matter directly and never used the term supererogation. Thus with

---

10 Of course one could claim, that this is not necessary as the testimony of the agents is merely a form of modesty (cf. Urmson 1958, 203) but I believe that one should initially take these at least as serious as the judgment of the observers.
11 McCarty 1989, 43.
12 Vogt (2008, 228) thus states this for example: “No action can be called good if it is not done from duty, and one can only act from duty when one acts in compliance with a duty”
reference to the given definition of supererogation, it seems reasonable to look at what Kant says about optionality of actions, about merit, praise and sacrifice in light of promoting other’s happiness. The few passages that deal with extraordinary acts might rather imply that Kant rejects the possibility of supererogation. This becomes particularly evident in the following passage of the second critique where he states:

“By exhortation to actions as noble, sublime, and magnanimous, minds are attuned to nothing but moral enthusiasm and exaggerated self-conceit; by such exhortations they are led into the delusion that it is not duty […] which constitutes the determining ground of their actions […].” (KpV, AA V, 84f.)

On the other hand, some passages of the Doctrine of Virtue concerning the imperfect duties might rather indicate that the supererogatory can be integrated, for example, when he states:

“Imperfect duties are […] only duties of virtue. Fulfillment of them is merit (meritum) = +a; but failure to fulfill them is not in itself culpability (demeritum) = -a, but rather mere deficiency in moral worth = 0 […]” (MS, AA VI, 390)

“But I ought to sacrifice a part of my welfare to others without hope of return, because this is a duty, and it is impossible to assign determinate limits to the extent of this sacrifice.” (MS, AA VI, 393)

“But Ethics, because of the latitude it allows in its imperfect duties, unavoidably leads to questions that call upon judgment to decide how a maxim is to be applied in particular cases […] So ethics falls into a casuistry […]” (MS, AA VI, 411)

“How far should one expend one's resources in practicing beneficence? Surely not to the extent that one would finally come oneself to need the beneficence of others.” (MS, AA VI, 454)

On this latter textual evidence, Thomas Hill prominently proposed that supererogation in Kantian Ethics can be understood as forming a subclass of wide imperfect duties to others. In short, he argues that this is because wide imperfect duties come with a differing degree of latitude. Whereas one always has to judge when and how a duty is to be fulfilled, the widest imperfect duties – the duty of beneficence and the duty to develop one's talents – do under certain circumstances also include the permission not to act upon them, given an agent has generally adopted the relevant maxim and has so far continuously acted upon it. Performing such a permissible act thus is supererogatory.

---

13 All citations of Kant’s work refer to the Prussian Academy Edition. Translations and abbreviations used as follows:
15 Cf. ibid., 71.
This approach has been subject to at least four objections:

First of all, by relating supererogation to wide imperfect duties of virtue, Hill seems to modify the criterion of Optionality in a significant way – as even though the singular act might be totally permissible, its performance seems to be the fulfillment of a duty after all. At least this is a weak account of supererogation.

Second, as Hill himself concedes, the problem discussed above remains; if the act is supposed to have moral worth, it must be motivated by duty. But if this is the case, should we assume that the agent took the action as her duty by mistake?

Third, Hill’s reconstruction of the account of latitude is controversial, as it is by no means clear that Kant really allows that much leeway. Jens Timmermann, for instance, argues that one only has latitude with regard to the means of fulfilling a duty.

Finally, as Marcia Baron argues, even if one agrees that there is considerable leeway within imperfect duties, it nevertheless seems to be in opposition to the Kantian idea of morality to assume a demarcation between obligatory acts and merely good acts. Because this would not only make parts of morality a matter of taste, but particularly because this would be in contradiction to the most fundamental duty, namely the imperfect duty to perfect oneself morally, which has to be understood far more rigorous than the other imperfect duties.

Therefore three alternative strategies can be distinguished: some have tried to establish an alternative account of supererogation by modifying the criterion of moral worth. Others claim that it is impossible to give a Kantian account of Supererogation, but whereas some see this as a shortage of Kantian Ethics, others have contested the need of such a category.

Baron certainly follows the last strategy. I take her objection to Hill’s account to be the most severe, thus I will take the duty of self-perfection as the starting point of my analysis. Before I do so, let me quickly point out why I, in opposition to Baron, do not believe that this must ultimately lead to a rejection of supererogation. Baron contests the need of such a category, as she finds that Kantian Ethics delivers an alternative, if

---

16 Following Baron (1987, 244f.) the problem is “that supererogationists are trying to get away from duty, but the Kantian analysis keeps pulling what is supposed to be 'beyond' duty back under the rubric of duty. Kant treats the acts in question not as beyond duty, but as ways of fulfilling one's imperfect duties.”

17 Hill (1971, 74) discusses a possible solution to it. He claims that it might be possible to modify Kant’s position to the effect that it is sufficient to be motivated by moral consideration.


20 McCarty (1989, 47ff.) claims that if the acts are optional then they cannot have moral worth but they might be admirable in an aesthetic sense, whereas Jokic (2002, 227ff.) claims that they are only “nearly morally praiseworthy”.

not even better solution, to the theoretical and practical problems that are at stake—namely describing the actions as fulfillments of imperfect duties and praising the good character of virtuous persons. But she admits

“that there are acts that can (contextually, and as acts; not as act types) be labeled “supererogatory”. But it is the wrong thing to focus on; it is not something that ethical theory should emphasize. Supererogatory acts “exist”, but they do not form an ethically useful or theoretically interesting kind.”

Now I do share Baron’s point that there cannot be a category of supererogation, but pace Baron I claim, that these singular acts, that can be labeled supererogatory, do in fact form an ethically useful kind.

3. EXAMPLES OF SELF-PERFECTION

My argument in favor of this assumption runs in three steps: I will first establish what I take to be a proper account of self-Perfection. Then I will outline why self-Knowledge is indispensable to it, and, finally, I will point out how such alleged supererogatory acts contribute to this.

3.1 Self-perfection

When introducing duties to oneself in the Doctrine of Virtue, Kant lets us know that these are fundamental when he claims that if there were no duties to oneself, then there would not be duties to others either (cf. MS, AA VI, 417). But the imperfect duty to perfect oneself as a moral being seems to constitute the most fundamental duty among these. This is because it aims at developing and improving moral capacities, and, thus, the more someone improves in that respect, the more impact this will have on the capacity to fulfill all the other duties. Aspiring to fulfill this duty can be understood as the path to virtue, which is the ideal of human morality (cf. MS, AA VI, 383).

Thus self-perfection must be understood as a continuous process, in fact a process that can never reach its end, or as Kant puts it:

---

22 Cf. Baron 1987, 239.
23 Cf. ibid., 262.
24 Ibid., 241.
25 It is perfect in quality, but imperfect in terms of degree (cf. MS, AA VI, 446).
“For a rational but finite being only endless progress from lower to higher stages of moral perfection is possible.” (KpV, AA V, 123)  

Now to understand in how far Kant admits for degrees of moral perfection, it is helpful to take a look at the different ways he uses the term virtue: he speaks of a virtuous disposition (Tugendgesinnung), obligation of virtue (Tugendverpflichtung) and duties of virtue (Tugendpflichten). The first is concerned with the purity in moral motives, i.e. to have a good will. The second is concerned with the strength in overcoming one’s inclinations so that the proper maxims will lead to according actions. And the third is concerned with the content of these actions, i.e. with the different duties of virtue.

It seems that with regard to the first use it is not intelligible to talk of degrees. You either have a good will or not, and, as Kant often points out, there is no middle thing in that. But strength in overcoming one’s inclinations is something that can be gradually developed. Because even though we have the capacity to overcome inclinations, the strength in doing so has to be acquired (cf. MS, AA VI, 397). And dependent on that one has a gradually varying capacity to fulfill the duties of virtue to a larger or lesser extent – the more one is thus able to do with respect to an imperfect duty, the more meritorious it is. And this can be understood as an external sign of a high degree of self-perfection.

3.2 Self-Knowledge

To perfect oneself morally, self-knowledge is a necessary precondition. As Kant points out, it is “the first command of all duties to oneself” (MS, AA VI, 441). Self-knowledge is indispensable because it works against the human tendency of self-deception, of which Kant distinguishes two forms. On the one hand, the tendency to euphemistically present the own actions as good (schwärmische Verachtung). The task here is to critically examine whether the maxim one acted upon was actually moral or whether the act was simply a case of moral luck and an inner lie to believe that it was done from duty. On the other hand, there is the tendency to empty commitments to the good in form of mere declarations or wishes (eigenliebige Selbstschätzung). The task here is to critically examine one’s actions, i.e. whether

---

27 And similarly in the *Doctrine of Virtue* where he points out, that for human beings perfection can only consist in a “progression from one perfection to another.” (MS, AA VI, 446)
28 Cf. MS, AA VI, 387 and 474.
29 For the distinction between obligation of virtue and duties of virtue cf. MS, AA VI, 410.
30 Cf. Blöser 2013, 52f.
31 Cf. ibid., 53ff.
32 In the Draft to the *Doctrine of Virtue* Kant calls this the “subjective precondition of all virtues”, as Esser (2013, 287) points out.
there really are or will be observable acts of will which show good attitude (cf. MS, AA VI, 441f.).

But self-knowledge is flawed with an epistemic problem, namely the opacity of one’s own motives (cf. MS, AA VI, 392). So, if an act really deserves to be called meritorious can never be known for sure, neither by the observer of such an action nor the actor herself. The closest we ever get to evidence is when someone’s act is according to a good maxim even though the circumstances in the given situation make it seem extremely difficult to act upon it.

3.3 The Supererogatory

I take it that these sorts of acts are acts that are commonly taken to be supererogatory, given that they are also done with a cheerful heart. These acts usually seem to display a high degree of self-perfection. This is at least the cases when the agent does not seem to act out of character. So, if someone performs such an act, then given the observer’s own (presumably lower) degree of virtue, this act will appear as something optional.

My assumption therefore is that supererogation is only a plausible concept regarding the observer-perspective, who can, in contrast to the agent herself, judge the action to be a merely permissible action. An (alleged) supererogatory act then is one that, given the observer’s own moral development, seems to be beyond duty, but for the agent herself is not. The observer could realize her own deficiencies in this observation – it reminds the observer to critically examine herself. As I take it, such actions are not only instances of morally good acts but rather they can be considered as good examples. To illuminate the concept of a good example, it seems to be helpful to turn to the didactics of the Doctrine of Virtue, where Kant states:

“The experimental (technical) means for cultivating virtue is good example on the part of the teacher (his exemplary conduct) and cautionary example in others [...] As for the power of examples (good or bad) [...] what others give us can establish no maxim of virtue. [...] the law itself, not the conduct of other human beings, must serve as our incentive. [...] A good example (exemplary conduct) should not serve as a model but only as a proof that it is really possible to act in conformity with duty. So it’s not comparison with any other human being whatsoever (as he is), but with the idea (of humanity), as he ought to be [...].” (MS, AA VI, 479f.)

Here the example does not simply have the function to illustrate the good, but rather it serves as a normative ideal, namely that one can always further improve one’s strength and thus be able to fulfill the duties of virtue to others beyond the present limits.

Hence, my argument does not draw upon the assumption that it is possible to integrate the supererogatory among the basic ethical categories within Kantian
Ethics, but rather as a subjective category of action assessment that has an important epistemic function in the process of developing one’s own moral capacities.

Finally I believe it resolves the tension in the textual evidence. Let us once again have a look at the quote from the second critique:

“By exhortation to actions as noble, sublime, and magnanimous, minds are attuned to nothing but moral enthusiasm and exaggerated self-conceit; by such exhortations they are led into the delusion that it is not duty [...] which constitutes the determining ground of their actions [...]” (KpV, AA V, 84f.)

Now Kant proceeds under which circumstances, these acts may be praised:

“Actions of others which are done with great sacrifice and for the sake of duty alone may indeed be praised by calling them noble and sublime deeds, but only insofar as there are traces suggesting that they were done wholly from respect for duty and not from ebullitions of feeling.” (KpV, AA V, 85)

And this seems to be quite in line with the proposed account of alleged supererogatory acts as examples that have to be done from duty but only appear as if they were optional to the observer, due to her frailty of nature. Therefore, I assume that understanding alleged supererogatory acts as instructive examples is not only a source of self-deception, as Kant fears even though the fear certainly is justified, but can also serve as a source of self-knowledge.

4. CONCLUSION

In this paper I have claimed that it is possible to accommodate the supererogatory as a subjective category of action assessment within Kantian Ethics that can play an important role in the process of moral self-perfection.

With regard to the given definition of supererogation it seems that this account fulfills the five criteria named above. The acts in question are altruistic, involve sacrifice, are meritorious, have moral worth and are, at least from the observer’s point of view, optional with regard to her capacities. Even though, because of that latter modification, supererogationists will claim that this is not sufficient as it does not establish a basic ethical category, I do believe that it nevertheless does justice to the phenomenon. Especially because it offers a solution to the heroism paradox: on the one hand, it can explain the common moral intuition that certain acts are supererogatory. While on the other hand, it can capture the empirical observation that agents of alleged supererogatory acts usually do not perceive their actions as supererogatory. And this is precisely because the differing perspectives of the agent and the observer constitute an integral part of the given account.
Therefore I claim that the outstanding characteristic of supererogation is better captured by focusing on singular (alleged) supererogatory acts and not by figuring out the status of supererogation as such – or, in other words: I believe that we rather need a functional than a conceptual analysis and that Kantian Ethics is a good candidate to offer such an analysis, even though it is conceptually anti-supererogationist.\textsuperscript{33}

\textsuperscript{33} I would like to thank the audiences at the Dublin Graduate Philosophy Conference 2016: Kant, Metaethics and Value, the Kant in Progress Workshop in July 2016 in St.Andrews and the Workshop on Supererogation with David Heyd in Basel, as well as the anonymous reviewers for this journal for their for helpful comments on earlier drafts of this paper.
Bibliography


