

## International Phenomenological Society

---

The Impossibility of Supererogation in Kant's Moral Theory

Author(s): Daniel Guevara

Reviewed work(s):

Source: *Philosophy and Phenomenological Research*, Vol. 59, No. 3 (Sep., 1999), pp. 593-624

Published by: [International Phenomenological Society](#)

Stable URL: <http://www.jstor.org/stable/2653784>

Accessed: 05/11/2012 17:23

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*International Phenomenological Society* is collaborating with JSTOR to digitize, preserve and extend access to *Philosophy and Phenomenological Research*.

<http://www.jstor.org>

# The Impossibility of Supererogation in Kant's Moral Theory<sup>1</sup>

DANIEL GUEVARA

*Cowell College*

*University of California, Santa Cruz*

It is common to think that certain acts are supererogatory, especially certain heroic or saintly self-sacrifices for the good. The idea seems to have an ordinary and clear application. Nothing shows this better than the well-known cases which J. O. Urmson adduced. Urmson argued that no major moral theory could give a proper account of the supererogatory character of such acts, and that therefore none could account for "all the facts of morality," as he put it. But his arguments were sketchy. This paper shall show, in some detail, that he was essentially right about Kant's moral theory, and that the criticism goes deep and holds up against recent sympathetic interpretation of Kant's views of duty and worth.

## §1 Introduction

It is common to think that certain acts are supererogatory, especially certain heroic or saintly self-sacrifices. That is, some acts involve too much self-sacrifice to be morally required—even though they are especially good and urgent acts. This idea has a clear and ordinary application, as illustrated by

---

<sup>1</sup> I use the following abbreviations to refer to Kant's works:

C1 = *Kritik der reinen Vernunft*, Norman Kemp Smith, trans., *Critique of Pure Reason*. New York: St. Martin's Press, 1929. [A and B editions are cited in the usual way: A/B]

C2 = *Kritik der praktischen Vernunft*, Lewis White Beck, trans., *Critique of Practical Reason*. New York: Macmillan Publishing Company, 1956.

G = *Grundlegung zur Metaphysik der Sitten*, H. J. Paton, trans., *Groundwork of the Metaphysic of Morals*. New York: Harper Torchbooks, Harper and Row, 1964. This translation was originally published as, *The Moral Law*. Hutchinson University Library, 1948.

MM = *Metaphysik der Sitten*, Mary Gregor, trans., *The Metaphysics of Morals*. Cambridge: Cambridge University Press, 1991.

Religion = *Die Religion innerhalb der Grenzen der bloßen Vernunft*, 1793, Theodore M. Greene and Hoyt H. Hudson, translators., *Religion Within the Limits of Reason Alone*. New York: Harper Torchbooks, Harper and Row, 1960.

All page references, except in the case of *Religion*, are to the Prussian Academy of Sciences edition of Kant's works. For *Religion*, the pagination is of the Greene and Hudson translation.

cases which J. O. Urmson adduced:<sup>2</sup> a soldier who throws himself on a grenade in order to save his comrades, for example, or a doctor who volunteers to help in a foreign, plague-ridden city.

Urmson argued that no major moral theory could account for the supererogatory character of such acts, and that therefore none could account for “all the facts of morality,” as he put it. But his arguments were sketchy, and many have disagreed. I want to show, in some detail, that Urmson was essentially right about Kant’s moral theory.

The basic problem for Kant’s theory, I hold, arises because he takes the moral worth of an act to wholly consist in the act’s being done from proper regard for duty—that is, from respect for the moral law.<sup>3</sup> And insofar as the acts which Urmson adduced are supererogatory, their moral worth could not wholly consist in their being done from respect for the moral law.

It is true that supererogatory acts may not be *possible* without proper regard for duty, and in some sense may even emerge from proper regard for duty. But I will argue that their moral worth cannot *consist* in proper regard for duty, or properly motivated fulfillment of duty.<sup>4</sup> It seems to follow that supererogatory *worth* is an indispensable feature of the supererogatory character of an act. This point is important: to call the acts under consideration “supererogatory” is to attribute a praiseworthiness which could not be in virtue of the praiseworthiness of mere conscientiousness in the commitment to moral ends. I will argue that this is as much a part of the idea of supererogation as is the thought that some acts are morally good, but not morally required. Indeed, it turns out that the characterization “morally good

---

<sup>2</sup> Urmson 1958, see especially pp. 201–2.

<sup>3</sup> See, e.g., C2 81–86, especially 85. Cf. C2 32, 74–75, and G 397, 400–401 & 405, MM 379, 383. It is important to bear in mind that Kant conceives of duty as a principle (the moral law) which all particular duties exemplify. Particular duties are duties in virtue of exemplifying its form, a form or principle which is necessarily a constraint (a categorical imperative) on the human will. It is also important to bear in mind that particular duties are often arrived at through casuistry, and that the picture is not at all one of a list of relatively specific inviolable commandments. Moreover, Kant relies upon the “matter” of the law, the ideal of pure rational agency (autonomy, or ends in themselves, etc.) more often than not, in the casuistry of the cases. That is, he relies upon a more or less intuitive grasp of the one indefeasible and supreme value.

<sup>4</sup> “Properly motivated fulfillment of duty,” or “proper regard for duty,” is my gloss on the so-called motive of duty: the moral motive or respect for the moral law. This keeps in sight the distinction admitted by common sense and most theories between morally good acts, and acts merely in accordance with duty, *whatever the motive* (acts which Kant calls “*pflichtmäßig*” including acts with “legality but not morality”). This rules out a view that seems to be a complete non-starter: namely that any act good for being beyond mere outward harmony with duty is supererogatory. This view seems to be a non-starter since it says, for example, that simply adding the moral motive (proper moral motivation) is supererogatory. Yet some seem to offer this as the view of supererogation in Kant’s moral theory, as Marcia Baron notes in citing Onora O’Neill’s view (Baron 1995, p. 28 n. 19).

and not morally required,” is too blunt to carve out the special character of supererogatory acts.

Moreover, it will emerge that the most interesting and fundamental reason Kant cannot account for (at least saintly or heroic) supererogation is that he cannot grant to one’s own happiness the right sort of weight or value. He cannot give it the proper value in relation to the value of morality, or in relation to the promotion of, or commitment to, the mandatory ends of morality which might require self-sacrifice. His theory of value requires a stark separation between moral value and the value of happiness (the value of autonomy and heteronomous value), and the deepest problems about his theory may lie there, in his opposition to any form of eudaimonism. But, in any case, he cannot admit the right sort of relation necessary for the idea of supererogatory worth. This deeper point, which is often overlooked in discussions of supererogation in Kant’s theory, connects with some of the more trenchant contemporary criticisms of Kant’s view of duty and worth.

My discussion relies upon the intuition that certain acts, such as those described by Urmson, *are* supererogatory, indeed, that they are paradigms. I take this to be a common intuition, often expressed by the claim that they are beyond the call of duty, or saintly or heroic. While there are extreme circumstances in which duty can demand the ultimate self-sacrifices, and in that way be heroic or saintly too, there are many circumstances in which duty does not demand so much, and yet in which people go on to be heroically good or saintly all the same. That seems evident to me, anyway. And, I am sure my view is fairly common. Indeed, I shall proceed on the assumption that a theory is discredited if in fact it makes supererogation impossible. However, in the final sections I shall consider more carefully the claim to the contrary that some Kantians wish to make.

## **§2 The inadequacy of the usual characterization of supererogatory acts**

Barring special circumstances, then, heroically good acts, like those adduced by Urmson, fall squarely within the bounds of the supererogatory, as it is usually characterized: namely, morally good but not morally required. The usual characterization is inadequate though.<sup>5</sup> It does not distinguish

---

<sup>5</sup> Yet it appears in otherwise sophisticated and informed accounts. For example Heyd says (1982, p. 115), “[a]n act is supererogatory if and only if (1) [i]t is neither obligatory nor forbidden. (2) Its omission is not wrong, and does not deserve sanction or criticism—either formal or informal. (3) It is morally good, both by virtue of its (intended) consequences and by virtue of its intrinsic value (being beyond duty). (4) It is done voluntarily for the sake of someone else’s good, and is thus meritorious.”

Except for the requirement of an altruistic motive in (4), I cannot see in these four conditions any significant difference from the characterization “morally good but not required,” assuming a reasonably broad interpretation of “morally good” and “morally required.” The condition of “intrinsic value” in (3), and of voluntariness and altruism in

supererogatory acts from acts which merely fulfill certain kinds of duties: so-called disjunctive duties.<sup>6</sup>

Say I have a duty to give a small percentage of my income to a reputable charity, and that Oxfam and CARE are the only two I have found that, after reasonable consideration, I could give to with any confidence. Thus, for example, I might have a disjunctive duty to give a hundred dollars to Oxfam or a hundred dollars to CARE. If I elect to give to Oxfam rather than CARE, then I do my simple duty, even though giving to CARE is good and not specifically required. Clearly, there is nothing supererogatory in electing to fulfill the duty that way, if the acts are equally commendable, as I am assuming they are. Whatever moral worth the act may have, it has entirely in virtue of the properly motivated fulfillment of duty.

Now, it is true that the act falls under the more general description: "giving a small percentage of my income to a reputable charity," and under that description the act is morally required (by hypothesis at least). It might be thought, therefore, that the supererogatory could be defined as that which under *every* description is not morally required. But that is too strong. For, even the (putatively) paradigmatic self-sacrificial acts adduced by Urmson fall under the description "helping others," or "beneficence;" and I think that any supererogatory act could be redescribed as the promotion of a very broad, and arguably mandatory, end like beneficence.

The general availability of such a redescription would prove straightaway that there were no supererogatory acts, if it were correct to define supererogation as suggested. But, of course, the redescription masks what is intuitively of the essence for the sort of case Urmson gave, namely the great self-sacrifice for the good in such circumstances.

What seems indispensable to the right account, then, is the idea of an act's moral worth not being simply in virtue of the moral worth which accrues to the act from (the properly motivated) fulfillment of duty. More specifically, what seems important for heroic or saintly cases is that the moral worth is in virtue of a self-sacrifice too great to be required in those circumstances, even though the act falls within the scope of mandatory ends (which can in some circumstances require the greatest self-sacrifices).<sup>7</sup> It seems then that one's

---

(4), might seem different from the typical characterization. Indeed, since "intrinsic value" explicitly *involves* "being beyond duty," the difference might appear to lie there. However, in defining "supererogatory act" Heyd should be defining what it is for an act to be "beyond duty."

Moreover, acts of duty can be done voluntarily and from an altruistic motive (as Kagan 1984 points out). For other examples of what's essentially the typical characterization, see Chisholm 1964, especially pp. 147 and 153 (including n. 16); and O'Neill 1975, p. 96.

<sup>6</sup> Cf. the excellent discussions of this point in Mellema 1991, chapter 2, section 4, and Clark 1979.

<sup>7</sup> This characterization does not seem adequate for the class of more modest supererogatory acts, including simple kindnesses and generosity, and so on. Those acts

own happiness (and its risk or loss) is a consideration with weight that is independent of the mandatoriness of those ends, and whose self-sacrifice for those ends can contribute an especial value to their promotion—a moral value (I presume), inasmuch as it promotes a moral end or moral ideal.

### §3 Distinguishing the main argument from that of recent literature

There has been a lot of important literature, since Urmson's article, responding to familiar objections to Kant's moral theory—responses to Urmson's objection, or objections akin to the one he raised in that article. I mean objections which maintain that, in one way or another, Kant distorts or eliminates many values, including moral values, because of the enormous value he places upon duty or the moral law. I agree with much of the literature, sympathetic to Kant, responding to this sort of objection; nevertheless I think I can send Urmson's point home.

It is important then for me to say what I do *not* think. I do not think that Kant has no category for the "morally good but not required," nor that he thinks moral worth consists in filling one's life with duties (in order to be doing always more for duty's sake), nor that the particular duties generated by his theory are (on the whole) extraordinarily burdensome by comparison with the duties of other theories, or of commonsense, nor that duty must be the driving force of every morally significant act (as though, for example, one had to marry, or pick one's friends, from a sense of duty alone). It is not even *simply* that, as I have said, Kant makes moral worth consist in the motive of duty, proper regard for duty. Put just like that, the point is unilluminating. Put just like that, it *does* show that supererogatory acts will have to be such that they are done from the motive of duty, which is an awkward consequence. Still, this consequence leaves room to maneuver, as should be clear presently. And it is an unilluminating consequence, unless it prompts consideration of *why* Kant conceives of moral worth as consisting in the motive of duty, or respect for the law.

What is crucial, for my main thesis, is how that conception is forged from his theory of value. As noted, there is a stark separation in his theory of value between the value of one's own happiness and moral value, a separation which is itself implied by his idea of proper respect for morality: namely the unconditional subordination of happiness or self-love to it.

This is a separation and subordination required by our recognition of the special sort of incomparable and unconditional value of the moral law, proper

---

do not involve any great self-sacrifice. I believe that intuitions differ about whether to include this class at all. However, I need not settle the issue in order to establish my main claim. It does seem that any supererogatory act must fit the more general description: namely, morally good not simply in virtue of the moral goodness which is in virtue of the properly motivated fulfillment of duty.

regard for which constitutes the motive of duty—the moral worth-giving motive. I pursue the question of supererogation in order to isolate what has seemed traditionally to be so difficult about this stark cleavage and subordination in Kant's theory of value, and to show that it is particularly difficult when, like Kant, we take it to be implicit in ordinary moral judgment.<sup>8</sup>

For, in order to account for the commonly admitted supererogatory character of certain acts, a moral theory must be able to grant to self-love the proper sort of value or weight in relation to the foundational moral values and principles that might require one to sacrifice it. So, for example, a theory must let those foundational norms, and their authority to constrain us, be constituted at least in part by considerations of our own happiness as such, as in eudaimonistic theories. In that case, there could not be the sort of radical difference between morality (moral value, motivation, source of requirements, and so forth) and happiness that Kant insists upon. Likewise, it seems that there must be a prerogative over one's own life and happiness, to risk or lose it, free of the respect we are bound to give to any moral norm, and whose exercise in self-sacrifice can contribute moral value to an act (or purpose or commitment), beyond whatever moral value there may be in acts done from the respect we are bound to give to any moral norm.

These things are completely at odds with Kant's conception of morality. They are at odds with his conception of the proper commitment to morality: i.e. they are at odds with his notion of respect for the moral law (necessarily a principle of duty, moral constraint, for our will), and proper commitment to the mandatory ends which emerge from the law.<sup>9</sup> I can show this by attending to the most general features of the law: viz., its role as a supreme value and expression of an ideal of pure rational agency, set against all that concerns the (contingent) embodiment of the law by human rational agency, and its motivational role in that setting. I need not, for example, consider the various formulations of the categorical imperative. Kant himself, in the application of the law and in the casuistry of specific cases, almost always relies on a more or less intuitive grasp of the ideal of rational agency, or the related ideals of autonomy, respect and dignity, for example.

---

<sup>8</sup> Cf. the passages cited in note 3, above (especially those from G and MM), cf. also C2 105.

<sup>9</sup> Again, cf. the texts at note 3. Also, as I have mentioned, quite apart from the specific topic of supererogation, the main misstep may be the one that insists on separating moral value from the value of happiness (one's own or anyone else's in general), in the way Kant claims we must. For the sake of argument, I shall assume that we can separate them, contrary to some eudaimonistic theories. And, anyway, it is not obvious that the traditional eudaimonistic theories will contribute to the solution of the difficulties I am raising for Kant. Cf. note 49.

#### §4 Imperfect duty

Almost all arguments on this issue turn on Kant's notion of imperfect duty, in their effort to show either (i) that Kant's theory *can* account for supererogation, or (ii) that it *cannot* and *need* not.

It is interesting that the arguments for (i) and (ii) sometimes proceed from a common basis: namely that Kant has a concept of *imperfect* duty, which preserves whatever seems worth preserving of our intuitions about supererogation. In service of the first claim, (i), this means that the notion of imperfect duty preserves pretty much everything the friend of supererogation wants; in service of the second claim, (ii), it means that not much is preserved nor worth preserving. I shall argue against both claims—(i) and (ii)—by showing that Kant's notion of imperfect duty cannot be used to support either one, and that the real point of interest, for the question of supererogation, lies rather more precisely with the theory of value behind Kant's general conception of duty, from which all particular duties emerge. Moreover, although I agree with part of the second claim, (ii), as I've mentioned, I do not accept some of the arguments others offer to support that part either.

I shall explain this last remark in a moment (next section), but first I should say a word about Kant's distinction between imperfect and perfect duty, and show how his notion of imperfect duty might seem capable of supporting *either* claim (i) or (ii). For the most part, I need not be too precise about the distinction between these two types of duty in Kant's theory. And for the moment I can rely on typical examples.<sup>10</sup>

*A propos* of claim (i), consider the imperfect duty of beneficence: the duty to promote the welfare of others, or as Kant would say, to make the happiness of others my end (i.e. one of my ends). This duty is quite indeterminate by comparison with a perfect duty, like the duty to pay back a loan on time. There is a wide range of acts, types of act, or patterns of act, which can fulfill the duty of beneficence, and some are more self-sacrificing than others. So, on the one hand, it might seem that some of the self-sacrificing fulfillments are, as it were, over-fulfillments and thus supererogatory, even by Kant's own lights—for all his emphasis upon duty and upon the doctrine that the moral worth of an act consists in its being done from the motive of duty. Perhaps in this respect there is no fundamental difference between perfect and imperfect duty. After all, I can go out of my way to pay you back in rare bills for your collection, or with interest you didn't require, or the like. In any case, it looks as though a certain indeterminacy in duty, especially imperfect duty,

---

<sup>10</sup> In Kant's view the distinction is related to (and *more or less* correlated with) duties of virtue and duties of right, or ethicality and mere legality, broad and narrow duties, etc. But it is not easy to lay all these distinctions and correlations out. I follow Hill 1971 pretty much. Cf. Gregor 1963, and Baron 1995, chapter 3.

allows for supererogation. This is a first approximation of the sort of argument that leads to claim (i).

On the other hand, as regards claim (ii), it might be argued that the acts are after all *fulfillments of duty*, because they all fall within the wide net cast by the duty of beneficence. They are all instances of acts which promote the relevant mandatory end (the happiness of others). Moreover, on Kant's view they must be done from the motive of duty if they are to have moral worth. Also, among the Kantian imperfect duties is the duty to perfect oneself (one's talents and one's moral character). Hence, any seeming over-fulfillment of imperfect duty is, as it were, absorbed by the apparently inexhaustible scope of imperfect duty, in particular by the duty to perfect oneself. Imperfect duties are never done, as their name suggests: they are not perfectible, not completeable. They cannot be dispatched like a debt. Some wish to argue that therefore Kant *needn't* and shouldn't account for the putative supererogatory character even of especially self-sacrificing fulfillments of imperfect duty.<sup>11</sup>

### §5 Transition to the secondary literature

Now, there is some truth to both of these arguments I've just sketched, but I hope to show that neither one is on the whole successful.

Also, claim (ii) holds that Kant *cannot* account for supererogation. Accordingly, those who support it offer certain arguments against the claim that he *can*. Yet I must reject some of these arguments, even though I agree that Kant cannot account for supererogation. For example, I shall maintain that it is a mistake to hold that the incompatibility with supererogation has its source in Kant's wide, apparently inexhaustible imperfect duty—and furthermore that it does not lie, more particularly, in the imperfect duty to perfect oneself.<sup>12</sup>

Rather, as I have said, the source of the incompatibility is with the motive of duty—with Kant's special conception of the proper regard for duty or the law, and in the theory of value involved in this conception.

Again, I want to avoid arguments which do not illuminate the main point as I see it. Thus, as I have said, I do not think the argument amounts simply to this: if Kant were to admit supererogation, then his conception of moral worth and duty would imply that supererogatory acts are done from the motive of duty.

This is indeed implied by his view of moral worth and duty, given that supererogatory acts are morally good and that (as he thinks) the motive of duty is the only moral worth-giving motive. There's no doubt that this would be an awkward consequence. I don't know if it is an entirely unacceptable one, though. It may even be the correct account of some cases of supereroga-

---

<sup>11</sup> Cf. especially Baron 1995, chapter 1.

<sup>12</sup> Pace Baron 1995, pp. 41–42.

tion, cases in which it seems quite plausible that an act (or pattern of acts) was done from a commitment to duty, even although it was nonetheless supererogatory. It seems inevitable that in the ordinary course of life a conscientious person will perform some such acts—i.e. supererogatory acts from a sense of duty, from a sense of basic decency, basic good neighborliness, civic mindedness, or the obligations of friendship, among other things. This claim has an air of paradox, yet it seems inevitable that a sincere embrace of mandatory moral ends is incompatible with too much concern about whether one has done more than enough. Therefore mere conscientiousness tends to be generous; it tends (so to speak) to spill over into the supererogatory in this way, although probably the acts would be more modest examples of supererogation than the heroic or saintly type. Anyway, this seems plausible to me, and I wish to grant it, at least for the sake of the argument.

Even so, it doesn't seem plausible as a general account, especially not for the sort of cases Urmsom adduced, where (if nothing else) the self-sacrifice would tend to concentrate one's mind about what may or may not be done in such circumstances.

In the next section, I give a recapitulation of the recent literature on supererogation. This leads up to the main article I wish to discuss: a sophisticated, plausible, but ultimately unsuccessful, attempt at a Kantian defense of supererogation. Then I try to send home the difficulties for Kant's theory.

### §6 A recapitulation of recent literature

I want to discuss the main themes and dialectic in recent literature by considering some initially plausible (though inadequate) reasons for thinking, on the one hand, that supererogation *is* and, on the other hand, *is not* possible in Kant's theory. This will help prepare the way for an important argument by Thomas Hill, Jr., to which I want to give close attention.<sup>13</sup>

As noted, in Kant's moral theory, there are duties that require a commitment to general ends. For example, the duty of beneficence requires that we adopt or "will," as Kant says, the happiness of others as an end. This is a very general end, and the related duty is quite indeterminate about the acts it requires.<sup>14</sup> Some acts clearly fall within its scope, like giving to charity or visiting the sick (from the proper motive).<sup>15</sup> Indeed, Kant had in mind acts

---

<sup>13</sup> Hill 1992.

<sup>14</sup> There are many good discussions of this aspect of Kant's theory. For discussion relevant to the purposes of this paper see, Hill 1992, and Herman 1993, chapter 3 ("Mutual Aid and Respect for Persons"), and see her "Credits" for original place of publication of this article, and the others of hers I cite.

<sup>15</sup> Herman (1993, chapter 3) includes charity, but might not include visiting the sick, under the duty of beneficence (cf. especially sections 2 & 3, where she discusses the idea of "true needs"). In her account, everything depends upon the meaning of "true needs." She draws a fairly sharp distinction between beneficence, a duty to respond to the "true needs" of others, and benevolence, which includes small acts of kindness, among other

like these in urging the duty upon us. It is impossible, though, to be specific about how much of this sort of thing fulfills the duty.<sup>16</sup> There are countless ways of promoting the happiness of another. The scope of this duty seems limitless. Or the duty might be thought of as an exceedingly large disjunctive duty to do this beneficent act or that one, this set of beneficent acts or that other one. In this respect, there is perhaps a rough analogy with disjunctive duty, an analogy which might seem to eliminate supererogation, because Kant's theory could make all putative supererogatory acts fall within the scope of such broad duties.<sup>17</sup>

Thus it might seem that in Kant's theory the possibility of supererogation would be pre-empted straightaway by imperfect duty. For, to return to the case of mere disjunctive duty, the giving to Oxfam rather than CARE is not supererogatory if there is a disjunctive duty to give to one or the other.

There is an obvious response, though: one might give to *both*. And that, or something like that, is always a possibility and, so to speak, *the* possibility of supererogation. In case of the very broad and indefinite duties, it seems that there could be an act or pattern of acts which played an analogous role to the act of giving to both Oxfam and CARE. It should be clear how, if this is so, it holds forth the possibility of supererogation falling within the category of Kantian imperfect duty.

Yet, the analogy between disjunctive duty and broad imperfect duty is rough at best, and seems strained, really. In the simple case of Oxfam and CARE, there is a determinate way of dispatching the duty, and thus a determinate way of going beyond it. But a very broad duty like beneficence allows so much latitude that nothing determinate can be said about its fulfillment. In this respect the simple disjunctive duty *is*, and beneficence is *not*, like the keeping of a promise or repayment of a debt. Simple disjunctive duty, with

---

related things (which, according to her account, are not duties—but see my next note). I do not think the general thrust of my argument is affected by the distinctions she proposes between beneficence and benevolence (or kindness, etc.).

<sup>16</sup> Cf. MM 393. Since Herman (*ibid.*) understands the duty of beneficence as the duty to respond to “true needs” (cf. my last note), she maintains that we are then obligated to give aid whenever we can if the aid is crucial to the person's true needs (as long as our own true needs are not threatened thereby). So on her view beneficence allows considerably less latitude than I am assuming it does. This is why many small acts of kindness and even some large benevolences or generosity fall outside the scope of the duty of beneficence, in Herman's view. She is careful to add that kindness, benevolence, and so on, have a “moral structure”: they flow from virtues, and perhaps also from some natural inclinations which we have a moral obligation to cultivate (*ibid.*, pp. 71–72). But the relation is fairly subtle and the casuistry is complex. It is not obvious that on her view benevolence (and so on) is supererogatory, for example. On her view, benevolence may not have moral worth or, alternatively, there may be a kind of indirect duty to do it (in which case I suppose it may have moral worth, but not supererogatory worth). Cf. also Herman 1993, chapter 1 (“On the Value of Acting from the Motive of Duty”), p. 17 n. 27.

<sup>17</sup> Cf. Baron 1995, pp. 41–42.

clearly specified disjuncts, is more like perfect duty than the broad imperfect duties such as beneficence.

Moreover, consider that even if the broad duty of beneficence ever does somehow, in some circumstances, come down to so specific an alternative as giving to Oxfam or CARE, then the duty is not necessarily dispatched upon performance of one alternative, or even both alternatives. To take the sort of case which has been discussed in the literature: imagine someone who now appreciates that he has a long history of neglecting beneficence, and wishes to begin or to renew his commitment towards it. Perhaps because of the way his time and resources are tied up (due to the same neglect), the first opportunity (or the first reasonable and obvious one) to begin or to renew his commitment offers itself in the form of a fairly simple alternative, like sending a check to Oxfam or CARE immediately.<sup>18</sup> Even if he gives to both charities (let alone one), it is by no means obvious that his 'generosity' is supererogatory. He has only begun to fulfill his obligation after a long pattern of neglect.<sup>19</sup> It seems that his charity is supererogatory only after there is a well-established pattern of promoting the obligatory end involved. But a "well-established pattern" is vague.

So there is also a *disanalogy* between the case of a simple or fairly specific disjunctive duty and the broad, indefinite duties laid down by imperfect duties like beneficence or other broad mandatory ends. And someone might wish to argue that because of the special indefiniteness of broad duties like beneficence, there isn't any basis for, or principled way of, drawing the distinction between supererogation and duty.<sup>20</sup>

---

<sup>18</sup> Cf. Hill 1992, p. 173, and also Baron 1995, p. 32 n. 25.

<sup>19</sup> But this is sort of a special case. However, it is worth pointing out that on Herman's view (1993, chapter 3, section 3), it is quite possible that giving to both charities (and more) is in general required until it threatens the giver's true needs. I think her view implies that supererogatory beneficence is impossible, since everyone must give aid whenever he can, i.e., until it threatens his own true needs. Her view admits, however, more latitude than some forms of consequentialism do. For example, we are allowed the latitude of inclination to some extent: we are permitted to help someone we love rather than, say, a perfect stranger (as Kant himself says in effect, MM 390). But Herman is quite aware that her account nevertheless makes beneficence more demanding than it is usually thought to be by Kantians (Hill 1993 for example) and others, and of course supererogationists. We can get a sense of *how* demanding when we consider that, if nothing else, the poor are always with us, and it is easy enough for us to write a check at any time. It might be thought that benevolence or kindness or generosity beyond some reasonable commitment to this, would be perhaps in some sense 'supererogatory,' even on the sort of view Herman is proposing. That is, the benevolence, etc. would not be directly required. But in that case it is not clear that it would have moral worth either, for it is not clear that the beneficence or benevolence could be done from the motive of duty, in Kant's sense. Indeed, it might conflict with the duties to oneself.

<sup>20</sup> Baron notes (1995, p. 41) that there is "no clear line of demarcation" between what's obligatory and what's nice but morally optional. She means there is no clear line demarcating such a distinction in Kant's theory. This may not be so: if Herman is right, "true needs" will provide a fairly clear line. Although, as I have said, Herman's account

This is not a successful argument. For although “a well-established pattern of promotion” is vague, it nonetheless provides a desired and useful criterion. As long as Kant can recognize, as it seems he can, that certain patterns of act manifestly show proper commitment to the welfare of others, then it may be possible to say that certain patterns which go well beyond that are supererogatory. There is in Kant’s theory no duty to push the pattern of beneficence to the hilt of self-sacrifice. And vagueness does not of itself prevent him (or anyone else) from acknowledging that some patterns or acts go well beyond the vague line, anymore than the indeterminateness, say, of the concept of generosity prevents knowledge of when someone has been generous, or quite generous, or the like. In this way the distinctions implicit in “supererogatory” are like other perfectly legitimate and useful, but indeterminate ordinary distinctions.<sup>21</sup>

The point is, then, that perhaps Kant can recognize that some heroic acts of goodness fall, or could very well fall, on the supererogatory side of the indefinite line drawn by, “a well-established pattern of promotion.” So, returning to Urmson’s examples: if the doctor or soldier had been in general conscientious about the welfare of others, surely then there is no question about the supererogatory status of their respective acts (all other things also equal). Isn’t this something Kant’s theory can grant?

It depends on whether he can grant that certain patterns of beneficence go well beyond what is involved in the minimal morally motivated commitment to the relevant mandatory end.<sup>22</sup> This is where Kant’s theory does finally meet

---

does not obviously provide a case for supererogation either (see notes 15, 16, 19 above). More important, not too much should be made of the indeterminateness in the distinction between fulfillment of imperfect duty and supererogation, as I go on to explain.

<sup>21</sup> Some of Baron’s arguments (1995, chapter 1) against supererogation attempt to show that there is not in fact any act *type* that falls beyond what could reasonably be required. She argues that not even “(good but) extremely difficult” or “(good but) extremely self-sacrificing” delivers the supererogatory act type. She gives short shrift, though, to the fact that we can nevertheless distinguish circumstances in which the self-sacrifice puts the act squarely in the desired type. Again, consider the analogy with our application of ordinary indeterminate terms, like “generous.” But cf. also Baron’s response to Richard McCarty (Baron 1995, pp. 102–3).

<sup>22</sup> On Herman’s account (1993, chapter 3), the minimal moral commitment to beneficence involves the commitment to respond to the true needs of others, irrespective of how much one has done so in the past (with all provisos mentioned thus far: see notes 15, 16, 19 and 20 above), and also especially if there is *not* much self-sacrifice involved (nor any questionable dependency upon the benefactor). This is quite different from Hill 1992. And I think it will seem counterintuitive and unduly burdensome to many even though they may admire those who keep that standard. Obviously, I myself am inclined to think so, especially of cases where the self-sacrifice is significant, even though I admire many who try to live by that standard.

Nevertheless, aside from this, I think Herman’s arguments (1993, chapter 3, sections 1&2) constitute a persuasive exercise in Kantian casuistry about beneficence, especially as they support her other chief concern: to show that the duty of beneficence does not rest ultimately upon prudential reasoning (in answer to objections from Schopenhauer and Sidgwick, et al.).

a serious obstacle. I hope to show that this is so, in spite of plausible considerations to the contrary developed by Hill, which I turn to now.

### §7 Hill's argument

I have entertained a rough analogy between broad duties and disjunctive duties in order to recapitulate some arguments and counter-arguments in recent literature, and also in order to show that the notion of imperfect duty does not necessarily crowd out supererogatory acts.

I should note, however, that it is more accurate to maintain that in Kant's view beneficent acts, or patterns of act, *manifest or express*, rather than *fulfill* the inner commitment to the relevant mandatory end. Strictly speaking the fulfillment is in the inner commitment to the end. Certainly, no merely outward act of beneficence or pattern of acts fulfills the duty of beneficence.<sup>23</sup> If they are properly motivated, then they are expressions of the inner act of willing the happiness of others as an end (out of respect for them as persons).

Even with that in mind it might nevertheless seem that Kant could recognize that certain patterns of beneficence flow from an inner commitment that goes well beyond anything involved in the proper commitment to the relevant end. After all, as I have said, there is no requirement in Kant's theory that we take beneficent acts to the hilt of selflessness. Indeed, there are obligations to oneself which prohibit one's doing so. Moreover, the welfare of others is only one mandatory end among others. And it is Kant's express view that we are allowed to subordinate one end to another (e.g., love of one's neighbor to love of one's friends), in deciding how we shall fulfill our imperfect duties.

Does anything prevent Kant's admitting that minimal moral commitment to mandatory ends allows some latitude about how much we must sacrifice? And does anything prevent him from deeming the great self-sacrifices as especially good, and not deeming them to be so *simply* in virtue of their being from a properly motivated fulfillment of (imperfect) duty? It might seem that nothing does.

Thus, consider the following case, offered by Hill, to illustrate the point. Hill says,

Suppose a man, who had conscientiously helped others and given to charities all his life, buys for a neighborhood child a treat that is neither needed nor solicited but is quite welcome. The

---

<sup>23</sup> Even on Herman's account (1993, chapter 3), which may be less latitudinarian than Hill's (1992), we are allowed the latitude of subordinating one maxim of duty to another (e.g., love of one's neighbor to love of one's parents), according to inclination. Cf. also Herman 1993, p. 63 n. 27.

Also, I am simply describing Kant's view, and thus setting aside difficulties about the inner and outer. In this case, they are the particularly vexing difficulties of the noumenal/phenomenal distinction, since the moral willing of the mandatory end is a noumenal act.

giver, a naturally cold person, is not prompted by kindly or sympathetic feelings but a strong desire to do the sort of act commended by moral principles. Far from rejecting Kant's limited principle of beneficence, he is disposed out of a (perhaps misguided) respect for it to do more of the sort of act it directs than it strictly requires. Imagine, too, that other moral principles are irrelevant in this case. He is not, for example, contributing to the gluttony of a greedy and obese child, and there is nothing else he is obliged to do with the money.<sup>24</sup>

This case is meant to fit a five-part Kantian definition of supererogation which Hill also offers. The definition says that a supererogatory act,

(a) is of a sort commended by a principle of wider imperfect duty, and (b) is motivated by a sense of duty (or, perhaps, respect for moral reasons), (c) is neither forbidden nor required by another, more stringent duty..., (d) is in a context where no alternative is required by more stringent duty and there is at least one alternative that is neither forbidden by more stringent duty nor commended by other principles of wide duty, and (e) is done by an agent who has adopted the relevant principle of wider imperfect duty and has often and continually acted on that principle.<sup>25</sup>

Someone might query the act's being a small act of generosity. But I think this feature should not be considered indispensable to Hill's main point. Hill himself intends that we extrapolate from this case to the heroic type I have been emphasizing.<sup>26</sup> There are various other things that might be queried, but I think the definition and example more or less capture what the friend of supererogation has in mind.<sup>27</sup>

However, what of the idea that the act must be done from the motive of duty?

Perhaps Hill's parenthetical remarks betray that he sensed a difficulty here. In one passing remark he suggests that the man is motivated by a "perhaps misguided" respect for the principle of beneficence, in another (definition part (b)) he notes that it is better to say that the man is guided by "respect for *moral reasons*" rather than by the motive of *duty*. Hill must have sensed some trouble in the idea that a supererogatory act could be done from the motive of *duty*. Presumably, when he suggests that the man in the example is misguided, he has in mind that the man must be under the misapprehension that what he does is a duty, when in fact it is supererogatory. Hill must have thought that otherwise the man could not do it from the motive of

---

<sup>24</sup> Hill 1992, pp. 168–69.

<sup>25</sup> *Ibid.*, pp. 168–69.

<sup>26</sup> *Cf. Ibid.*, p. 170.

<sup>27</sup> Someone might wonder whether Hill intends "a naturally cold temperament," or the like, to be essential to the case. I think that he does not; rather he wants us to be clear that no *mere* natural sympathy is involved (*cf. G* 398).

Also, by wide or wider imperfect duty Hill means duties like beneficence. My discussion emphasizes (a) and (b). But, I should note that (c) and (d) are meant to ensure the permissibility of the act (and its omission), and (e) is meant to ensure that the agent has, so to speak, done his fair share of beneficent acts.

duty.<sup>28</sup> But is it plausible that supererogation involves this sort of mistake? It is generally implausible, especially in heroic cases, where if nothing else the self-sacrifice would tend to concentrate one's mind about what may or may not be done in such circumstances. Perhaps considerations like these prompted Hill to make the further suggestion that it is better to characterize the motive as: respect for "moral reasons" rather than *duty*.

In any case, Hill's considered view is as follows:

...we might suggest that it is not an unreasonable extension of Kant's position to say that what is required for moral worth is not a motive to do one's duty but a motive to do what is demanded or encouraged by moral considerations. Kant was preoccupied with duty in much of his ethical thinking, but his main point, surely, was that moral worth depends upon one's disposition to live by whatever demands and ideals are implicit in a rational, moral life. If doing a certain favor for a person is not morally obligatory, then Kant would not want him to do it in the mistaken belief that it is. The most that he could expect of a moral-minded man is that he does the favor because he wants to do the sort of thing encouraged by moral principles even if on this occasion they do not demand it of him.<sup>29</sup>

### §8 The motive of duty and supererogation: C2, Part II ("Methodology of Pure Practical Reason")

As reasonable as Hill's remarks may be (and *will* be to anyone wishing to grant the possibility of supererogation), they nevertheless propose a serious revision in Kant's views. That is what I wish to establish in this and the following section, through a careful reading of some passages in the second Critique. It should be clear, in the end, that Kant rejects the possibility of supererogation. More important, it should be clear that the friendly amendment Hill proposes is incompatible with Kant's theory of value, the value exemplified in action from the motive of duty.

However, first let me say that it is surely true that, as Hill notes, Kant does not think that moral worth flows from a preoccupation with duty—some easily discredited obsession with duty. Still, Kant resists any account of *moral* motivation (moral worth-giving motivation) that does not make the motive out to be respect for the law or proper regard for duty. This is not a quibble about terms. For, reasons or motives are *moral*, according to Kant, only insofar as they are instances, or exemplifications, of *the* moral reason: the moral law. And the moral law is necessarily a principle of duty for us, and the form exemplified by all specific principles of duty (moreover, Kant believes that all of this is implicit in the *ordinary notion of duty*).<sup>30</sup> These views are reflected in Kant's explicit discussion of various examples of putative supererogation and of a putative special worth. I want then to turn to

---

<sup>28</sup> Cf. Baron, pp. 35f.

<sup>29</sup> Hill, p. 172.

<sup>30</sup> See, e.g., C2 81–86, especially 85. Cf. C2 32, 74–75, and G 397, 400–401 & 405, MM 379, 383.

these and a few other related texts, in order to send home the argument against the possibility of supererogation in Kant's theory.<sup>31</sup>

The whole of the *Critique of Practical Reason*, Part II: "Methodology of Pure Practical Reason," treats of the sort of acts of heroic goodness Urmson adduced: "noble (super-meritorious) actions," as Kant calls them.<sup>32</sup> And Kant sets the tone against the possibility of supererogation, early on, by going so far as to say that "one need only to reflect a little to find an indebtedness which the vaunted hero has in some way incurred to the human race...which will prevent the thought of duty from being repressed by the self-complacent imagination of merit."<sup>33</sup> And he adds, "even if it be only that, by the inequality of men under the civil constitution, [the hero] enjoys advantages on account of which others must be lacking to just that extent."

If this claim is meant to be quite general, I dare say that it will strike most everyone as a stretch. I trust that examples like Urmson's make that clear, even for someone sympathetic to the idea that the well off are, as we say, "indebted to society" in certain ways (even though they may be well off

---

<sup>31</sup> Cf. Baron's quite different treatment of these same texts, (1995, pp. 36f).

<sup>32</sup> C2 155. Note the term "super-meritorious," which is as close as Kant comes to using the term "supererogatory" here (cf. Baron 1995, p. 47 n. 49).

It is worth noting this term because MM 227–28 & 389–91 are sometimes thought to contain a use of "meritorious" which is synonymous with at least some sense of "supererogatory." He says, for example, "What anyone does in accordance with duty beyond what he can be compelled to do by the law is meritorious (*meritum*); what he does only just in accordance with the law is debt owed (*debitum*)." (MM 227).

Richard McCarty (1989, p. 43) suggests that Kant may have the concept of supererogation in mind when he speaks here of doing what is *beyond* (compulsion by) the law. And, perhaps, this text was influencing O'Neill too (1975, p. 96).

But by "what one can be compelled to do by the law," Kant means, "what one can be externally constrained to do" by some external force: the courts, for example. In this section of *Metaphysics of Morals*, "compulsion" is always "external compulsion." In other words this text is just another way of saying what he has been saying throughout this part of the *Metaphysics of Morals*: namely that morality, and thus moral worth (merit) come into view only when we get beyond mere external duties, and beyond the *non*-moral motives for doing one's duty, that is, when we get to the doing of duty for the right reasons, or, in other words, when one, as he says, "keep[s] one's promise [for example]" even when "no constraint is to be feared." (MM 220). Cf. MM 218ff and also, MM 389–90 and note how what's "owed" is closely tied to what can be externally compelled. This should be distinguished from the sense in which we "owe" the law the sort of meritorious respect he describes here, i.e. a respect we are bound to give, and no more optional than the keeping of a promise.

What he opposes in the section of C2 under discussion is "super-merit," i.e. merit beyond the moral merit in respect for the law.

<sup>33</sup> C2 155n: "*Man darf nur ein wenig nachsinnen, man wird immer eine Schuld finden, die er sich irgendwodurch in Ansehung des Menschengeschlechts aufgeladen hat...um durch die eigenliebige Einbildung des Verdienstlichen den Gedanken an Pflicht nicht zu verdrängen.*" I think by "merit" here, he must mean the "super-merit" he is querying.

Cf. also *Religion*, p. 44, where Kant says that all the good a man ever does is *his simple duty*. Cf. Hill 1992, p. 171 n. 14.

through, so to speak, no fault of their own).<sup>34</sup> And of course many saints and heroes are of humble origins. But someone might say Kant went farther than he needed. And, it does seem that at times he is willing to grant the possibility of supererogation. Yet it will be evident from the way he qualifies his remarks about such a possibility that his considered view is always against it.

The context of discussion in the section under consideration is Kant's plea to moral educators that they spare us the sort of hero which fills the pages of "sentimental writings."<sup>35</sup> He is concerned about the danger of our otherwise losing sight of the quotidian, the warp and woof of our obligations, where the will is rarely stirred by the thought of heroic greatness.<sup>36</sup> Now, this plea or caveat is not inconsistent with the idea of supererogation. But whenever Kant does entertain heroic cases his purpose is to call attention to the supposed fact that only the motive of duty is the genuinely moral motive.

Thus, as in his first example, suppose that the king compels you to lie for some evil end, but that you stand firm against all the terror he can command.<sup>37</sup> Now Kant does say "virtue here is worth so much only because it costs so much," which suggests that degrees of self-sacrifice increase moral worth. However, Kant is inclined towards the example because it exhibits firmness at all cost for the sake of *perfect duty* (refusing to lie for some evil end). This is because in general his point is that heroic examples best set forth ideals of *moral* worth (and thus ideals most useful for *moral* instruction) only to the extent that they are examples of perfect *duty*, or else approximating perfect duty. Kant is interested in such cases insofar as they illustrate that the likely source of action was the thought of duty alone. For, they thereby illustrate duty's unconditional worth and appeal. Examples of heroic perfect duty do this as well as any example can, because of the way they limit the plausible options about the source of motivation. They make it implausible that happiness or self-love was the motive (although we can never know for sure, even in the best cases, cf. G 407). This is one theme of the section, and part of a more general theme that morality provides a greater motive than happiness. More important, by contrast with examples of wide imperfect duty, examples of heroic *perfect* duty also make it plausible that there could not have been an especial, meritorious opting for the heroic fulfillment of duty (over a more modest choice). For, the heroism is *required*; it is—in Kant's view—one of those circumstances in which morality demands the greatest self-sacrifices.<sup>38</sup>

---

<sup>34</sup> Cf. a claim in MM 454 along these lines.

<sup>35</sup> C2 155; cf. C2 157.

<sup>36</sup> Cf. C2 157: "the dry and earnest idea of duty which is more fitting to human imperfection and progress in goodness," as compared with the "soft-hearted feelings or high-flying, puffed-up pretensions" of sentimental idealism, "which wither instead of strengthening the heart."

<sup>37</sup> C2 156.

<sup>38</sup> And I agree.

In sum, Kant offers the example of heroically holding out against the king as an example of *perfect* duty, and thus as an example of the worth which is in virtue of the accomplishment of duty, and not as indicating any supererogatory worth. When he says it costs so much and is thus worth so much, he means that the moral worth or virtue is *shown* to be so great in our eyes by our acknowledgment that it is worth the greatest self-sacrifices. These sorts of cases confirm for us our recognition of the far greater worth of morality (the law in us), than the worth of any value there might be in the satisfaction of inclination. On Kant's view, it is important for moral educators to illustrate these things because of the universal temptation to think of our own happiness as a consideration of overwhelmingly great weight, indeed as something of unconditional value on a par with, or even greater than, the law.

It might be said, however, that Kant is not concerned in these passages to criticize the idea of supererogation except as it distracts from these essential moral and pedagogical points. Indeed, he seems at times to grant that there is a praiseworthiness and merit beyond duty, in some sense. He says, for instance, that "it is proper to extol actions which display a great, unselfish, and sympathetic disposition and humanity." He is referring to the so-called noble and super-meritorious acts. Yet in a characteristic hedge, he hastens to add that "in them we must attend not so much to the elevation of soul, which is very fleeting and ephemeral, as to the subjection of the heart to duty, from which a more lasting impression can be expected..."<sup>39</sup> Moreover he concludes these remarks with the one I introduced in the beginning, saying "one need only to reflect a little to find an indebtedness which the vaunted hero has in some way incurred to the human race...which will prevent the thought of duty from being repressed..."

His chief purpose here, and throughout, is to establish the motivational power of the pure motive of duty as compared with certain other motives that might seem to compete with it. He is concerned with whether there is, as he says, "more subjective moving force of a drive in the thought of an action as noble and magnanimous than when the action is thought of merely as duty in

---

Cf. Gregor 1963, chapter 7, for a discussion of how the distinction between perfect and imperfect duty developed gradually. Moreover, even in relatively clearcut form, the application of the distinction might involve casuistical questions: e.g., Did the heroic saving of others involve something like a suicidal disregard of one's own life?

I am assuming that Kant takes the shipwreck case as imperfect or as less clearly perfect than the perfect duty to preserve one's own life and well-being. The thought that the one is less clearly perfect than the other is a reasonable thought, insofar as the one appears to fall under beneficence. But I suppose the cases raise a casuistical question in the end, with certain cases putting pressure on the distinction between, say, imperfect beneficence and a stricter duty, akin to justice, to preserve human life. (What if I only have to reach over the side to save you?)

<sup>39</sup> C2 155n. Cf. G 398, where Kant makes essentially the same point in his remark about the natural inclination to beneficence or honor in some persons.

relation to the solemn moral law.” He is out to show the latter. And it is with that in mind that he introduces two more heroic cases.

The first one is of someone who loses his life in the attempt to save others from a shipwreck. He says that although the person’s action “will indeed be counted...as duty,” it will also be counted “even more as a meritorious action...” Yet he hedges immediately about the notion of merit, when he adds that “our esteem for [the act] will be weakened very much by the concept of [the hero’s] duty to himself, which here seems to have been infringed.” Then he drops the case and turns straightaway to the “more decisive” case of the “magnanimous sacrifice of [one’s own] life for the preservation of his country.” I think Kant takes it as obvious that this is indeed a more decisive case for illustrating his main point about the thought of duty. Yet he hedges again, for it is not entirely decisive as an example of duty; thus he observes, “there still remain some scruples as to whether it is so perfect a duty to devote oneself spontaneously and unbidden to this purpose, and the action itself does not have the full force of a model and impulse to imitation.”<sup>40</sup>

This is supposed to be in contrast with cases of “inexorable duty, transgression against which violates of itself the moral law without respect to human welfare and...tramples on its holiness...[In the case of inexorable duty] we give our most perfect esteem to pursuing it and sacrificing to it everything that ever had value for our dearest inclinations; and we find our soul strengthened and elevated by such an example when we convince ourselves, by contemplating it, that human nature is capable of such an elevation above everything that nature can present as a drive in opposition to it.” I presume this is meant to describe the case of holding out against the king. And thus we’ve come full circle: Kant’s chief concern is always with eschewing all that is not clearly an example of the worth and motivating force of duty for duty’s sake.

But why does he say of the heroic act of losing one’s life in saving others from a shipwreck that it will be counted as duty but “even more as a meritorious action?” I am primarily interested in the remark about meritoriousness.

But, first, why does he say that the act will be counted as duty? How can he, when the act seems to him to conflict with a duty to oneself? Is it that he thinks of the act as a *prima facie* duty insofar as the act falls within the wide scope of beneficence? And does he therefore think of it as overruled by the (presumably) stricter duty to oneself? And what specific duty to oneself does he have in mind? The (indirect) duty to preserve one’s own life and well being? I think so, and he does explicitly say (elsewhere) that the duty to oneself is the stricter sort of duty, that it trumps beneficence.

---

<sup>40</sup> Cf. C2 158.

But if the heroic act is in fact morally impermissible, then how could it be counted *even more* a meritorious action? How could it be meritorious at all? I suppose Kant could mean that it is meritorious when regarded *prima facie* as an instance of beneficence. It is an extraordinarily noble and selfless beneficence, and thus to be regarded even more as a meritorious action than, say, an unremarkable beneficence; as compared with an unremarkable beneficence, it is harder (though not impossible) to regard it as flowing from anything but the highest and purest motives.

This seems to me to be the right reading, given the overarching theme of the section: the power and purity of the distinctly moral motive. And the hedgings, the qualifications, are part of Kantian casuistry, according to which there may not be a definite answer in every case, and according to which each one of us must use his or her own judgment with what guidance general principles provide.

In general, whenever Kant considers the details of any difficult moral case, his evaluation is guided by the general parameters set by his fundamental ideals, and a more or less intuitive grasp of those ideals—i.e. rational agency, autonomy, ends in themselves, dignity and respect. I mean his thinking about the details of any case is lead by these ideals rather than the application of explicit rules, including the categorical imperative (which is hardly mentioned in the applied ethics of the *Metaphysics of Morals*).<sup>41</sup> He was well aware of the indispensable role of judgment in the application of any rule-governed procedure in ethics (or any other field). Principles and rules can guide us, setting more or less specific presumptions in favor of this or that type of act (or forbidding it) in various contexts. But the role of judgment is ineliminable. And good judgment requires good training or education, and there is no algorithm or strict method for producing that.<sup>42</sup>

### §9 The supreme and central value of duty

So for Kant, the first and foremost thing to be done in moral education, and for the development of virtue, is to distinguish as clearly and explicitly as possible that value which we call moral. For he thinks that when our ordinary notion of moral value is properly explicated and distinguished from all potential competing values (whether happiness, great aesthetic value, great talent and utility, or even heroic altruism or so-called super-meritorious nobility), it is seen to be supreme over them all, and the condition of their

---

<sup>41</sup> Cf. *The Metaphysics of Morals*, the various casuistical questions in the “Doctrine of Virtue.”

<sup>42</sup> All the points in this paragraph are now commonly accepted, contrary to traditional interpretations. Cf., e.g., Herman 1993, chapters 4 &7, for particularly good discussion.

value (G, ch. 1, especially 393 & 396).<sup>43</sup> But all of them tend also to *compete* or to become mixed up with the idea of morality, i.e., they tend to compete as sources of motivation for moral requirement (and the development of character and so on), even more than they tend to *conflict* with moral requirements. The supreme and unconditional good, the ideal of rational agency expressed in the law, is a constraint on this ever-present sort of competition and potential conflict.<sup>44</sup> And this is why the supreme value is duly represented by the notion of *duty*, and why proper regard for the supreme value is exhibited in proper regard for duty. It is represented for us by a categorical ought constraining our will against the deep and natural propensity to ground, at least in part, all of our actions and purposes in a principle of self-love. This is a categorical ought that, in Kant's special sense, is based on an ideal of pure rational nature that need not link up with happiness, or anything else, for its authority to set requirements, unlike the categorical oughts that might emerge from other theories.

This is enough to establish why, *pace* Hill, Kant's insistence upon the term *duty* is entirely central and apt—it represents the essential core of his conception of morality. It represents a value at the heart of his conception of moral motivation: the idea of a purely rational, practical principle unconditionally constraining our will. This represents the one indefeasible value or consideration guiding the casuistry of the details of any cases.

Now the law, and our grasp of its ideal of autonomous, rational agency, allows plenty of latitude for nuanced and sophisticated casuistry about particular duties. And it might be said that, unless we accept everything in Kant's moral psychology, we needn't emphasize the idea of *constraint* so much. For, although Kantian morality must not, need not, look to considerations of happiness to support its authority to constrain our will, it is far from being entirely aloof from happiness in application. It must look to happiness in application, for that is the primary contingency of human life to be taken into account in the application of the law to the only species of rational nature that we know of. The duty of beneficence itself originates from the application of the law to this contingency. Moreover, the natural concern for one's own happiness is generally automatic and firm, and yet there is even an indirect duty to attend to one's own happiness, especially in those special circumstances when the natural concern can be eroded.

And besides, as has been discussed, the law grants plenty of latitude about which sacrifices we are to make in the promotion of mandatory ends. This latitude is the species-specific embodied expression of our autonomy, and

---

<sup>43</sup> Among many other texts. It is in some sense even supposed to be the source of the value of them all. This does not reduce them to instrumental value nor deny them "intrinsic value," in perhaps a certain sense of that term. But see Korsgaard 1983 and 1986a.

<sup>44</sup> I discuss this point at length in *Kant's Theory of Moral Motivation* (forthcoming, Westview Press).

gives expression to that side of the law which exalts, as contrasted with that side of it which constrains and humbles, our attachment to self-love. Kant's idea is of embodied autonomy as self-constraint over one's acts and purposes, a constraint accomplished through unconditional identification with that aspect of ourselves which gives those purposes their value and meaning.

But when someone opts to risk everything for a good cause, Kant must say that the agent's act (or purpose or character) is not *morally* motivated and thus not *morally* good unless it is done from the unconditional subordination necessarily involved in this identification. Kantian morality is essentially, and in the first place, the recognition of a value the only proper response to which is an unconditional subordination of anything else that might be of value. Moral worth consists in acting from this evaluation. No act or purpose or character is even partly moral until it is influenced by this evaluation—and thus no beneficent self-sacrifice is either, however great or small (as good or right as it might be otherwise).

Kantian moral worth flows from the recognition that our good acts are an expression of something we are bound to do, even to the detriment of all our inclinations (G 401). It probably won't come to that, especially not with a very broad duty like beneficence—moreover, especially with the counter-vailing duty to oneself to care for one's "true needs." But the rider about the potential ultimate cost can never be completely idle in the moral motivation of an act. It indicates the unconditional subordination involved in respect for the law. Circumstances forcing the ultimate cost in willing the happiness of others as an end would have to be extreme. Yet I suppose that Kant thinks we must have some sense of that possible cost, and I take it he thinks this comes from our grasp of the incommensurable worth of the law that binds us to beneficence (or any other duty). And we would not have appreciated this worth if the improbability of high cost was influencing our thought of the worth of committing ourselves to the law. Moreover, the rider is not simply the (analytic?) claim that duty is what must be done, come what may. It marks the unconditional subordination we are bound to make to an ideal, an ideal which of itself has nothing to do with happiness (e.g., it applies to non-sensible rational beings).

But, it might seem Hill's thought still has some force.

Inasmuch as the greatest self-sacrifices are not usually required, why couldn't someone who had nevertheless made the unconditional subordination, opt for the greatest self-sacrifices, with the latitude allowed to us all in the fulfillment of our broad duties to others? The sacrifice would be for a moral end, but not specifically required. Thus, on the one hand, we would have some modest self-sacrifices which were done from the recognition that beneficence is the sort of thing we are bound to do out of unconditional respect for other rational agents, but on the other hand, we would have *great* self-sacrifices that would be from the *same* recognition. If the modest ones

could be done from that recognition, surely the greatest ones could be too, although none of the instances, modest or great, would be specifically required. Indeed, in a way, the heroic self-sacrifice is perhaps more of a concrete expression of the sort of subordination involved in respect for the law, since it actually does, in one fell swoop, express the subordination. Couldn't there be supererogatory worth in opting for that mode of expression, even on Kant's view?

Kant can't see his way clear to this conclusion. For, he sees that a duty to oneself encroaches on the permissibility of the self-sacrifice. One's life is never one's own to risk or to give, not even for a great and noble end. One could only be constrained by a countervailing, presumably perfect, duty to risk or to give it. In either case, the respect we are bound to give to the supreme value in us is at work. And the "latitude granted to all" could never, in any circumstances, grant a prerogative over one's life free from that unconditional bind. It could never, in any circumstances, involve a prerogative to risk or threaten that—or anything else so crucial to the proper functioning of rational agency of our type.

So the conflict with the notion of supererogation persists.

Inasmuch as we think highly of the sort of acts Urmson proposed, we admire the self-sacrifice of a life, or of so much held dear, for a great and noble end. For we think that there is, in some circumstances, a prerogative to pay that cost for a great end, when it would have been perfectly appropriate to hang back too, and care for what we all often reasonably and appropriately hang back to care for, in spite of the ever-present invitation to us all to make a similar commitment to such ends. For isn't it open to us all to adopt great ends and pursue them with the greatest selflessness, whether it means only transforming one's commitment to his present ends, or adopting different ones?

Perhaps someone will say that we are all at fault either for lack of nerve or imagination, or for other weakness or plain selfishness in not doing so, wherever we are, whatever our station, and that there is nothing wrong with a moral theory that indicts us for this because it insists upon the highest ideals. Or else perhaps someone will say that only a countervailing duty could justify and make good the heroic or saintly acts and commitments in the first place.

### §10 Exhibiting the strength of the *prima facie* case

I think that with these sort of objections we must return to the *prima facie* case, and to the burden of proof it seems to establish against the elimination of all good candidate cases of supererogation. Let me exhibit the strength of the *prima facie* case by running through the basic possibilities.

If the acts Urmson adduced are not supererogatory, then it seems they must be either morally forbidden, morally required or morally neutral.

Could such acts be morally neutral, like having ice-cream instead of cake? It seems not. Lives are saved and great suffering is averted through someone's freely risking or sacrificing his or her own life. The view that such things could be morally neutral looks like a non-starter. Could such things be forbidden—or even questionable, or in some other way less than good? Perhaps they could, if some standing obligations were flouted, or the motives were bad or doubtful, and so on.<sup>45</sup> But clearly these qualifications will apply only to a limited number of cases. If nothing else, all are agreed that such acts can be entirely praiseworthy when they are duties.

Could they be good only in virtue of being duties? Is that the anti-supererogationist claim? If so, then it seems easy to construct counter-examples from cases all are agreed on. To stick with Urmson's war motif, I assume all sides are agreed that it can be morally good to take up arms for a good cause. Then, suppose we have taken up arms with some compatriots against a fascist take over of our country. Assume, for the sake of argument, that we are correct in judging this to be our duty. With this assumption, then, all sides can agree that the risks and self-sacrifices are praiseworthy. But then it seems easy to construct a case of supererogation from this common ground. For example, couldn't a foreign volunteer to our cause perform all the good that *we* might perform in such circumstances? (Did not many foreigners do so in joining with republican forces in the Spanish Civil War, for instance?)<sup>46</sup> Take a volunteer medic if the element of violence detracts.

Say a gifted, unattached, foreign medical student takes a year's leave and serves bravely in our cause. This sort of selflessness is typically admired—to put it mildly. It is just the sort of case people ordinarily have in mind when they maintain that some acts are supererogatory or “beyond duty,” as it might ordinarily be put. It is also the sort of case many moral theorists wish to account for as supererogatory.

Now there are interesting philosophical arguments which attempt to show that nevertheless this sort of case is an example of duty after all. Of course, such arguments must be able to exploit initially plausible ideas too. For example, they may exploit the utilitarian idea that what's good is good entirely in virtue of its fulfillment of the duty to maximize utility overall. According to this view the selflessly heroic act is good inasmuch as the act is a fulfillment of the overarching utilitarian duty. Otherwise we must maintain

---

<sup>45</sup> Cf. Baron (1995, pp. 38ff.) who shows how putative supererogatory acts might involve repugnant motives, or be associated with chauvinistic or hypocritical customs and practices. Most of Baron's examples are well taken. Nevertheless, she seems to me to present an imbalanced diet of examples in her effort to disparage the idea of supererogation and supplant it with the notion of Kantian imperfect duty. It should be borne in mind, against the arguments she adduces, that it is easy to think of repugnant things associated with the notion of *duty* as well, in spite of its role in clear and unexceptionable cases.

<sup>46</sup> Cf. George Orwell, *Homage to Catalonia*.

that it was forbidden, unless it is a toss up about which act maximizes utility overall.

But of course this strategy is not open to Kantians. Instead, Kantians must try to show that the act can be appropriately understood as an *imperfect* duty, a duty with wide latitude about the particular acts which fulfill it. Thus, the heroically selfless act may fulfill the broad duty of beneficence. Yet, is it plausible that the act we are considering is on a par with all other beneficences in respect of being morally good for fulfilling a duty?

For example, compare our volunteer medic and her similarly situated (gifted and unattached, etc.) classmate who opts to send money to our cause instead. We can imagine this case falling along a whole range of cases of varying degrees of self-sacrifice. No doubt intuitions will vary somewhat about what acts and patterns of act constitute a reasonable commitment to the welfare of others. Admittedly, some acts or patterns of act will, on any view, do no more than exemplify such a commitment. Moreover some of those acts will exemplify a more or less minimal commitment, i.e. a commitment that anyone, in more or less ordinary circumstances, will meet if he is sincerely committed to beneficence at all. Now I think that such acts are morally good entirely in virtue of their fulfillment of what morality requires with regard to the happiness of others. But the question is: is it plausible that *all* morally good self-sacrifices for the sake of others are morally good entirely in virtue of the fulfillment of that requirement, or of the duty to support democracy around the world, or to promote any other meritorious end? It seems quite implausible when we take account of the enormous differences in self-sacrifice at one end of the range.

Sometimes it is said that saints and heroes typically regard what they do as duty, or as nothing especially good. *A propos* of this I have also heard it said that Kant's theory is a "first person" theory. The idea in both points seems to be that this shows that the heroic or saintly acts *are* duty. The idea is that we should give considerable weight to how the agent regards the act, perhaps especially so in Kant's theory because of how Kant makes the agent's intention (maxim) the principle object of moral evaluation.

Now, I don't think anyone wishes to maintain that even an exceedingly virtuous person is incorrigible about duty. So I take it the idea is that what a virtuous person regards as duty, *probably* is duty. Perhaps this has merit as a general rule. But the conclusion to be drawn from it (against supererogation) rests on an empirical claim about how acts of putative supererogation are regarded by those who perform them. And the testimony of saints and heroes is not an entirely uniform and straightforward matter. It is certainly not so straightforwardly clear as to settle all cases of putative supererogation nor, I

think, so clear as to settle even some promising cases for the anti-supererogationist view.<sup>47</sup>

It is true that Kant's theory looks primarily to the agent's intention (*maxim*, in Kant's somewhat technical sense of that term) rather than to the outward act. Perhaps in that sense it is a "first person" theory. But against any special Kantian emphasis on first person conceptions and assessments of morally significant choice and action, recall that the moral law itself holds for all rational agents as such. Every interpretation must cope with this notorious abstraction from historical and culture context. It is true that the *categorical imperative* is a form of the moral law which takes account of the contingencies of our embodied rational agency, the contingencies of our finiteness. But we know Kant himself thought he could nevertheless derive absolute or nearly absolute prohibitions against lying and suicide, for example. Thus, I think that the evidence from first-person assessments against the possibility of supererogation must get its force from the point already dealt with, namely the testimony of saints and heroes. And the anti-supererogationist consequences of that testimony are disputable at best.

Perhaps someone will say the objection I am trying to raise against Kant's theory is dubious inasmuch as I must rely on these extreme cases in order to exploit whatever force the objection may have against a theory which delivers the right answer in the bulk of the cases. But all major theories give more or less the same results with the bulk of the cases. Their true test lies in their foundations, and in the ultimate values and sources of motivation that they propose, and in how the ultimate values guide us in the difficult cases,

---

<sup>47</sup> Remarkably good persons often speak modestly, or even disparagingly, of their putative supererogatory acts, as though the acts were not especially good, or as though they were mere duty, as noted by Susan Hale (1991, pp. 273ff.) I think this point deserves serious consideration, more than I can give in this paper. But notice that even Hale's treatment shows that the testimony of saints and heroes cannot be taken at face value. Thus, in one extraordinarily selfless case of resistance to the Nazi's that Hale cites, the persons who heroically resisted are even reluctant to call their extraordinary actions good. But of course the acts *are* good, as everyone, including Hale, concedes. Thus not even she takes the testimony entirely at face value. It does not seem unreasonable, then, for anyone who has the intuition that the acts were obviously supererogatory too, to treat all of the testimony as unduly modest, including the testimony that the acts were on a par with duty. Or at least, it is not unreasonable to be suspicious about what such testimony can be used to establish against the putative supererogatory character of the acts.

Moreover, (at least) saints often see themselves as having a special calling, through divine revelation. They therefore see themselves as responding to *special* obligations. And, even with that aside, I doubt one could find much evidence of saints or heroes ever thinking of what they did as perfect duty (to adapt a point from Urmson 1958, p. 212). Moreover, they tend to be quite reserved about expecting from anyone else the same level of dedication they routinely show (in their pursuit of broad ends), in contrast to the way we all might expect everyone to keep his promises or refrain from theft or violence or whatnot. When they do impose their standard upon others, we admire them less (I think of, for example, Gandhi's at times oppressive expectations of his wife and family).

and in how all of this can make sense of our pre-philosophical view of the bulk of the cases.

Or perhaps, (again) echoing Hill's suggestion, someone will still want say that the objection amounts to a quibble about terms. After all, Kant acknowledges non-moral values, even proto-moral ones: beauty, honor, natural sympathy, among others. So, perhaps the appropriate non-moral value could substitute well enough for the intuition of supererogatory *moral* value. However, this suggestion is blocked by the *impermissibility* of the self-sacrifice, or else by its mandatoriness when and if it is permissible.

The main difficulty really is one about the ultimate Kantian value. Perhaps Kantians could temper its claim on us, and stay true to the spirit of the theory.<sup>48</sup> Perhaps they could reject, as a philosophical extravagance or excess, the identification of *morality* (and duty and worth) with a law of pure practical reason, which serves as the norm for all species of rational beings alike. But I don't think the basic intuition that I have been defending could be preserved

---

<sup>48</sup> I do *not* mean to say that the particular duties generated by Kant's theory are on the whole extraordinarily burdensome by comparison with the duties of other theories or of commonsense. But it is worth commenting on the notorious exceptions: lying and suicide. The counter-intuitive results of the casuistry of each, point up the same difficulties I have tried to raise with the issue of supererogation: the elimination of a certain kind of prerogative.

There's not much casuistical room to maneuver with deceit, since it is such a direct violation (manipulation or perversion) of rational agency. Thus Kant seems forced to bite the bullet (as he notoriously does) and lay down an absolute prohibition. There may be some playroom with the distinction between an outright lie (*impermissible*) and a withholding of the whole truth (*permissible*, and perhaps even required, sometimes), or else with the permissible deceit constitutive of some games, or the untruths of (at least some) polite speech—as in the etiquette of correspondence—"no one is fooled by it," Kant says (MM 429f.) Or else perhaps even to save a man's pride (when, catching us off guard, he asks for an opinion of his work)—although Kant leaves this one unanswered. And here surely the only reason Kant could possibly see any room to maneuver is because of the arguably close connection to respect for the person. Think of the poignant contrast with his views about the murderer at the door (i.e. how could Kant waver here from concern with preventing mere feeling of embarrassment or the like, in view of that other infamous case?). The one thing that's clear in the casuistry of lying or suicide (or anything else so central to the existence and purpose of a rational agent) is that no amount of suffering of itself, for example, can tip the casuistical judgment in favor of deceit or suicide (or of their permissibility at least).

Prolonged suffering, with no end in sight, might permit suicide, to turn now to the other case, if the suffering threatened rational agency. But no amount of, as it were, lucid prolonged suffering, could.

I can imagine examples of holding out, in such circumstances, for the sake of ideals implicated in one's personhood (as Kant conceives of it), which I would admire. And I can imagine circumstances in which perhaps it would be wrong or shameful not too hold out. But, surely, in some circumstances (including some of those in which if the person held out, I might duly admire her), mere considerations of one's own happiness (inasmuch as those considerations can be distinguished from Kant's ideal, or any moral ideal) could justify euthanasia (self-imposed, and probably) otherwise.

For arguments sympathetic to Kant's stricter prohibition (at least against lying), see Korsgaard 1986b and 1988; and also Herman 1993, chap. 7, section 5.

unless they were to grant to one's own happiness the right sort of independent value or weight, one that could give rise to the right sort of prerogative. Kantians must let those foundational norms, and their authority to constrain us, be constituted at least in part by the value of our own happiness, as in eudaimonistic theories. In that case there could not be the sort of radical difference between morality (the *pure* moral value, motivation, source of requirements, and so forth) and happiness that Kant insists upon.<sup>49</sup>

### §11 Summary and concluding remarks

My conclusion has nothing to do with claiming that Kant must think of all morally good acts as specifically required. Anyway, except for theories which, as Kant says, strew our every step with duty,<sup>50</sup> any theory allows enough indefiniteness in duty for an act's being "not required but morally good." Sheer indefiniteness or latitude is a meager step towards securing the possibility of supererogation in a moral theory.

Also, Kant's denial of supererogation has nothing to do with reducing to purely instrumental worth other motives besides the motive of duty, since after all he does not do this (although in some sense he does think their value is conditioned upon the moral value).<sup>51</sup>

---

<sup>49</sup> Bernard Williams gives perhaps the most trenchant contemporary argument against this stark separation, in terms of what Kant would call a heteronomous notion of personal integrity. Williams's idea is of an integrity which involves the deep natural attachment to one's personal projects and commitments (see the works cited by Herman 1993, p. 23 n. 4 and her response to Williams, in the same chapter).

Also, my claims in this paper echo some objections which emerge from the related agent-centered and agent-neutral controversy. But the precise relationship of my claims to that issue would take some time to discuss. Cf. Scheffler 1982. It is worth noting that Scheffler explicitly defends deontological agent-centered prerogatives in order to account for supererogation, among other things. His view is a hybrid of utilitarianism and the prerogatives of non-consequentialist principles (of the sort sometimes associated with Kant).

Classical utilitarianism gives, of course, direct weight to happiness as such in constructing its moral ideal and the principles which can require self-sacrifice. But the goal of optimistic action seems to many to eliminate the idea of supererogation (cf. note 2, above). This is one reason Scheffler rejects it for a hybrid consequentialist/deontological theory. Perhaps for anyone who wishes to stay closer to Kant's own theory than that, Aristotle's form of eudaimonism gives the best alternative account of the intuitions I am defending.

Yet, it does seem to me that Urmson was right to see in *all* traditional theories a tendency that is incompatible with the proper account of those intuitions. What all traditional theories seem to do is to exalt an ideal or supreme value to the point of making the value of the more or less ordinary happiness of an individual shrink practically to nothing by comparison (the subliming of ordinary morality, as it were, in the attempt to account for it). *A propos* of Aristotle, compare the well known difficulties of reconciling what Aristotle says in book ten of the *Nicomachean Ethics* about contemplation with what he says about *eudaimonia* in the previous books.

<sup>50</sup> MM 409: "But that man can be called fantastically virtuous who allows *nothing to be morally indifferent (adiaphora)* and strews all his steps with duties," etc.

<sup>51</sup> Cf. note 43 above.

Also Kant can distinguish degrees of strenuousness in our self-sacrifices for the good, the degrees of difficulty in overcoming natural and legitimate fears, etc. But that norm of comparison is psychological and empirical, and thus amoral at best.<sup>52</sup> Such a comparative difficulty could not, on his view, make a difference to the *moral* quality of an act, except perhaps as it involved a violation of a duty to oneself. On Kant's view, it *might* involve a violation of the duty not to jeopardize one's own true needs, even though the act was for the good of another, for example.<sup>53</sup> In such a case the self-sacrifice is morally ruled out because it fails to show proper regard for the unconditional value of oneself. Obviously, though, none of this brings the theory any closer to the possibility of representing our ordinary view of supererogation. In fact, it presents its own obstacle to that possibility by morally forbidding self-sacrifices which might ordinarily be thought to be especially praiseworthy.

It is important to note, as well, that the Kantian foundational norm (the law) does not necessarily impose a form of alienation from oneself, from the things one might come to care most about. The requirements of the law may dovetail with one's own happiness by and large. Indeed, they probably do, as Kant himself thinks. Moreover, we can find great meaning and happiness in unconditional identification with a transcendent ideal.<sup>54</sup> Many saints and heroes do. That they do is often seen to be to their greater credit. But that this is so is not at all a matter of their fulfilling moral requirement. We may want to think of their acts as a manifestation of greater virtue.<sup>55</sup> And no doubt such acts often are a sign of greater virtue. But sometimes people do such things as Urmson described who haven't shown much evidence of virtue. And even when supererogatory acts *are* a manifestation of great virtue, there remains the question of whether we must be that virtuous.

Likewise, the impossibility of supererogation (or at least of a certain paradigmatic type) in Kant's theory is not a matter of the requirement of moral perfection in his theory. It might seem to be. After all, given such a requirement we never exhaust the demand on how good we should be. Marcia Baron argues that this requirement duly eliminates the need for a category of supererogation. That is, the duty of moral perfection eliminates the need for a line that demarcates the obligatory and the morally nice, as she puts it.<sup>56</sup> However, it should be noted that even for Kant, the duty of perfection is broad. It permits, say, relaxing at the park once in a while for no particular reason. Thus, it leaves it open what we should say about those who, with the latitude permitted to us all, do what is heroic and extraordinarily self-

---

<sup>52</sup> Cf. MM 391, where Kant distinguishes different "subjective" senses of "merit."

<sup>53</sup> Cf. MM 393, C2 158 (the shipwreck example discussed above).

<sup>54</sup> As Herman notes (1993, pp. 38f) in her answer to Bernard Williams's Gaugin case.

<sup>55</sup> Cf. Baron 1995, pp. 54f.

<sup>56</sup> See *Ibid.*, pp. 41ff. and 55.

sacrificing for the good. Typically we think of such persons as showing great virtue or extraordinary commitment to the ideal of moral perfection, a virtue or commitment that is supererogatory.

There is no clear line demarcating the distinction we make, but that fact of itself provides no reason for supplanting the notion of supererogation with the notion of Kantian imperfect duty, even the imperfect duty of perfection. To do so would be to ignore a moral distinction as significant as the one between, say, 'generous' and 'helpful,' on grounds that there is no clear line of demarcation between them (cf. p. 9 above).

Although there is a noble appeal in the idea that we can transcend self-love for a great ideal (an ideal which of itself has nothing to do with happiness), and although we admire those who do, it is nonetheless hard to understand why we *must* subordinate self-love unconditionally to any such ideal, the way Kant thinks the morality of an action requires in reverence for the ideal of the law. Who can reasonably expect such a subordination of anyone and everyone? Those that do it, i.e. the saints and heroes?<sup>57</sup> On what grounds?

The Kantian answer is that each one of us asks it, or even commands it, of himself or herself, inasmuch as we are rational, reasonable agents. It is hard to stand by this idea when confronted with such cases as Urmson describes or when the idea is claimed to be at the back of our ordinary moral judgments about the bulk of the cases. It is natural to think instead just what Kant would have us not think, just what he argues so strenuously against throughout the "Methodology of Pure Practical Reason": namely that we are, as it were, morally on our own, beyond duty, "beyond the worth which a man can and must give himself in his own eyes through the consciousness of not having transgressed his duty" (C2 155).<sup>58</sup>

---

<sup>57</sup> Again, many of the ones we really admire rarely do, and they expect it of themselves often on the basis of special revelation. Kant of course rejects that as a basis for morality. Although I think part of what drives his view is the attempt to give us, in the "holy law," something immanent, and yet with all the transcendent power of a divine command. God can demand or inspire the sort of subordination or reverence in question. But then if God really exists, perhaps there is no "ordinary" morality. Kant would be the first to say that, if it were possible, knowledge of God (a mystical vision say), other than practically based "rational faith" that God exists, would change everything.

<sup>58</sup> Many, many people were kind enough to comment upon various drafts of this paper. I wish to record my thanks for all their help. Robert M. Adams and Warren Quinn were profoundly helpful, through countless drafts. I regret that Professor Quinn did not live to see the final piece. I also wish to give special thanks to Gavin Lawrence, Michael Griffith, Torin Alter, Roger Florka, Carol Voeller, Marilyn Adams, Andrews Reath, Barbara Herman, Christine Korsgaard. Thanks also to Jerome Neu, Richard Wasserstrom, Holly Thomas, Ellen Suckiel, David Hoy and David Chalmers, who helped with various drafts and who have given a lot of support and encouragement beyond that. Also, portions of this paper were written while on course relief provided by a Junior Faculty Development Award at UC Santa Cruz, and generous matching support from the department of philosophy. These and the support of a UC Santa Cruz Senate Grant are gratefully acknowledged.

## REFERENCES

- Baron, Marcia W. 1995: *Kantian Ethics Almost Without Apology*. Ithaca: Cornell University Press.
- Chisholm, Roderick. 1964: "The Ethics of Requirement." *American Philosophical Quarterly* 1, 147–53.
- Clark, Michael. 1979: "The Meritorious and the Mandatory." *Proceedings of the Aristotelian Society* 79, 23–33.
- Gregor, Mary J. 1963: *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik der Sitten*. New York: Barnes & Noble.
- Hale, Susan C. 1991: "Against Supererogation." *American Philosophical Quarterly* 28, 273–85.
- Hancock, Roger. 1975: "Mill, Saints and Heroes." *The Mill Newsletter*, 10, 13–15.
- Herman, Barbara. 1993: *The Practice of Moral Judgment*. Cambridge: Harvard University Press.
- Heyd, David. 1982: *Supererogation: Its Status in Ethical Theory*. Cambridge: Cambridge University Press.
- Hill, Thomas E., Jr. 1992: *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca: Cornell University Press.
- Kagan, Shelly. 1989: *The Limits of Morality*. Oxford, Clarendon Press.
- Kagan, Shelly. 1984: "Does Consequentialism Demand Too Much? Recent Work on the Limits of Obligation." *Philosophy and Public Affairs* 13, 239–54.
- Korsgaard, Christine. 1988: "Two Arguments Against Lying." *Argumentation* 2, 27–49.
- Korsgaard, Christine. 1986a: "Aristotle and Kant on the Source of Value." *Ethics* 96, 486–505.
- Korsgaard, Christine. 1986b: "The Right to Lie: Kant on Dealing with Evil." *Philosophy and Public Affairs* 15, 325–49.
- Korsgaard, Christine. 1983: "Two Distinctions in Goodness." *The Philosophical Review* 92, 169–95.
- McCarty, Richard. 1989: "The Limits of Kantian Duty and Beyond." *American Philosophical Quarterly* 26, 43–52.
- Mellema, Gregory. 1991: *Beyond the Call of Duty: Supererogation, Obligation and Offense*. Albany: State University of New York Press.
- O'Neill (Nell), Onora. 1975: *Acting On Principle*. New York: Columbia University Press.
- Scheffler, Samuel. 1982: *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. New York: Oxford University Press.

Urmson, J. O. 1958: "Saints and Heroes," in A. I. Melden, ed., *Essays in Moral Philosophy*. Seattle: University of Washington Press, 198–216.