



## Above and Beyond: Kant and Supererogation

Andrew Delunas<sup>1</sup>

*1 Adjunct Professor, Department of Philosophy, San Jose State University, San Jose, United States*

Email- [andrew.delunas@sjsu.edu](mailto:andrew.delunas@sjsu.edu)

Received: 18/05/2020

Accepted: 21/05/2020

Published: 10/06/2020

### Abstract

*The soldier who throws herself upon a grenade in order to save the lives of her comrades is surely to be commended for her selfless act. While most moral theories do not require an agent to make such a sacrifice, nearly every ethical theorist lauds such behaviour as a paradigm case of beneficence. Daniel Guevara, by contrast, argues that Kant cannot account for the moral worth of supererogatory acts. In this essay, I shall examine Guevara's criticism of Kant's seeming inability to account for the beneficence of supererogatory acts, and argue that, despite surface appearances, Kantian deontology does not categorically dismiss the moral worth of every supererogatory act.*

**Keywords:** Supererogatory, Kantian casuistry, Kant's theory, Guevara's criticism

---

The soldier who throws herself upon a grenade in order to save the lives of her comrades is surely to be commended for her selfless act. But what establishes the moral worth of such acts? While no moral theory requires an agent to make such a sacrifice, nearly every ethical theorist lauds such behavior as a paradigm case of beneficence. One notable exception is Kant's deontological theory, with its strict emphasis on duty. Indeed, Daniel Guevara is quick to point out that this inability to account for the moral worth of supererogatory acts is deeply rooted in Kantian casuistry. Given Kant's seeming unwillingness to account for the moral worth of such acts, Guevara recommends the rejection of Kantian ethics. In this essay, I shall examine Guevara's criticism of Kant's seeming inability to account for the beneficence of supererogatory acts, and argue that, despite surface appearances, Kantian deontology does not categorically dismiss the moral worth of every supererogatory act.

"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>1</sup> Given that the soldier's act of throwing herself on the grenade (or indeed any "heroic" or "saintly" supererogatory act) was not performed out of respect for the moral law (nor with any seeming mind to moral duty whatsoever) Guevara contends that Kant would categorically dismiss any suggestion that the act had moral worth. While Guevara acknowledges that supererogatory acts can be rooted in a respect for moral duty (and indeed, they may not be *possible* without a regard for said duty) he nevertheless asserts that their "moral worth cannot *consist* in

---

<sup>1</sup> Daniel Guevara, "The Impossibility of Supererogation in Kant's Moral Theory". *Philosophy and Phenomenological Research*, Vol. 59, No. 3. September, 1999. p. 594

proper regard for duty, or properly motivated fulfillment of duty”.<sup>2</sup>

The important point is that in acknowledging an act to be of a supererogatory nature, we tend to hold such an act in higher esteem than an act performed in deference to “everyday” imperfect moral duties<sup>3</sup>. Moreover, given Kant’s insistence that moral duty necessarily trumps considerations of one’s own happiness, it becomes *impossible* for Kant to account for the moral praiseworthiness of supererogatory acts on Guevara’s interpretation. This insistence is made clear in the *Grounding for the Metaphysics of Morals*:

Such is the case not merely because this principle is false [i.e., happiness should serve as the grounds for morality] and because experience contradicts the supposition that well-being is always proportional to well-doing, nor yet merely because this principle contributes nothing to the establishment of morality, inasmuch as making a man happy is quite different from making him good and making him prudent and sharp-sighted for his own advantage quite different from making him virtuous. Rather, such is the case because this principle of one’s own happiness bases morality upon incentives that undermine it rather than establish it and that totally destroy its sublimity, inasmuch as motives to virtue are put in the same category as motives to vice and inasmuch as such incentives merely teach one to be better at calculation, while the specific difference between virtue and vice is entirely eliminated.<sup>4</sup>

As we shall see, the often-requisite sacrifice of one’s happiness (and our acknowledgment of that sacrifice) that is a significant part of supererogatory acts will become a major stumbling block to any Kantian acknowledgment of the moral worth of such acts, precisely because the sacrifice required of heroic or saintly acts of supererogation is beyond the reasonable bounds of duty (despite the fact that the acts are incorporated under Kant’s notion of mandatory ends). Taking the work of J.O. Urmson (1958) as his point of departure, Guevara cites Urmson’s examples of the soldier and the grenade as well as the doctor who volunteers her services in a foreign city stricken with plague as paradigm cases of supererogatory acts. Urmson argues “no major moral theory could account for the supererogatory nature of such acts, and that therefore none could account for ‘all the facts of morality’”.<sup>5</sup> And while Urmson notes that in cases of extreme duress, self-sacrifice may even be required by duty, “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”.<sup>6</sup> Indeed, Guevara maintains that he is certain that this is a fairly evident and widely held view. But the failure to account for ‘all the facts of morality’ has profound implications for Kantian moral theory; “I shall proceed on the assumption that a theory is discredited if in fact it makes supererogation impossible”<sup>7</sup>.

It may prove prudent to here discuss just what constitutes a “supererogatory” act. The common definition of “supererogatory” i.e., that of acts that are “morally good but not morally required”, is for Guevara wholly unsatisfactory owing to the fact that “It does not distinguish supererogatory acts from acts which merely fulfill certain kinds of duties: so-called disjunctive

---

<sup>2</sup>Guevara 1999. p. 594

<sup>3</sup>It is important to note that this is one of Kant’s primary arguments against the worth of supererogatory acts: the everyday person’s consistent moral performance is ignored while the hero (who may only have risen to the occasion only in this one instance) is lauded.

<sup>4</sup>Immanuel Kant, *Grounding for the Metaphysics of Morals*. James W. Ellington, trans. Hackett Publishing Company, Inc. 1981. P. 46.

<sup>5</sup>Ibid.

<sup>6</sup>Guevara 1999. p. 595

<sup>7</sup>Guevara 1999. p. 59

duties”<sup>8</sup>. Disjunctive duties are a concatenation of several imperfect duties linked over time, such as my donating my time to a specific soup kitchen as opposed to say, a homeless animal shelter. Both acts are fulfillments of my imperfect duty to be beneficent, but Guevara argues (correctly I suggest) that the fulfillment of our imperfect duties is not adequate to merit the designation “supererogatory”. What *is* required (and is indeed indispensable) for an act to be deemed supererogatory is something beyond the accrual of moral worth that all requisite acts of fulfillment of duty ought to garner. Indeed, in the “heroic” or “saintly” cases of supererogation, Guevara maintains that the defining characteristic of a supererogatory act 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”<sup>9</sup>. It is not simply the sacrificing of one’s life (or exposing oneself to mortal danger) that makes an act supererogatory then, but rather the corresponding loss of *one’s own happiness* that affords an especial value on said act. Such a weighty sacrifice lends an even greater esteem to an already laudable act of beneficence.

With an adequate definition of supererogation in hand, Guevara moves to discard various criticisms of Kant’s theory that he considers non-starters:

“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.<sup>10</sup>”

This last point Guevara notes, “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”<sup>11</sup>. The crucial point for Guevara is the inevitable consequence of the relationship between moral worth and duty, namely the forsaking of one’s happiness in deference to the moral law. Guevara notes, “...there is a stark separation in [Kant’s] 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*”<sup>12</sup>. This enduring difficulty remains the Achilles’ heel of Kantian deontology, thinks Guevara, and indeed, it severely curtails any attempt to account for heroic or saintly acts of supererogation.

It is not that self-love or happiness should be the *decisive* factor in moral considerations for Guevara; rather he seems to suggest that Kant does not offer *enough* consideration of our happiness in moral decisions, “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”<sup>13</sup>. I

---

<sup>8</sup>Guevara 1999. p. 596

<sup>9</sup>Ibid

<sup>10</sup>Guevara 1999. p. 597

<sup>11</sup>Ibid

<sup>12</sup>Ibid

<sup>13</sup>Ibid

will take issue with this point later in this paper, but for the moment I want to continue with the exposition of Guevara's criticisms of Kant's inability to account for the happiness of the rational agent in making moral decisions. Citing Aristotle's virtue ethics as an example of a theory that does afford the proper weight to issues of the agent's happiness Guevara notes, "a [moral] 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"<sup>14</sup>. Moreover, there must be "a prerogative over one's own life and happiness" that stands apart from any moral value that considerations of one's duty might impart on a moral act<sup>15</sup>.

This criticism seems to suggest that in order for any theory of morality to be taken seriously it must coalesce with our pre-philosophical *intuitions*. And we have seen that Guevara is quite adamant that any moral theory that cannot account for supererogation must be wholly rejected. It is well apparent that Kant cannot afford the proper weight to considerations of one's happiness in his casuistry (specifically, Kant cannot account for the significance of the loss of our happiness in the performance of our moral duty) and this, thinks Guevara, is a damning criticism of Kant's theory (as well as a complete impediment to an account of supererogation within Kantian theory). Guevara contends, "These things [i.e., proper consideration of one's happiness and the role it plays in moral decision making] are completely at odds with Kant's conception of morality"<sup>16</sup>. Again, the point is clear: to the extent that Kant places such a premium on a moral act's being performed in deference to the moral law, he cannot afford the proper consideration to the role one's happiness *must play* (on Guevara's account) in moral decision making.

Let us now consider some possible defenses of Kantian theory in light of Urmson's and Guevara's criticisms. One possible scenario I will consider involves the idea that supererogation is not (or should not be) a fundamental aspect of morality and moral worth. Consider the case of the young German soldier in Hitler's army. Let us suppose that he is completely taken with Nazi propaganda, and has eagerly enlisted his services to the fascist regime. Let us further suppose that he is confronted with the exact situation that Guevara considers a paradigm case of "heroic or saintly" supererogation; i.e., he must throw himself upon a grenade in order to save his comrades. Now let us imagine this young soldier's career in the German army prior to this point: he has been a party to the burning of several villages; he has murdered innocents in cold blood; he has helped to round up scores of Jews for repatriation to the death camps. And now he is faced with what seems to be the very epitome of a moral conundrum. He makes his decision and immolates himself, thereby saving his comrades. Is his selfless act teeming with moral worth? Initially it might seem that the answer is "Yes".

But consider the cause that the young soldier chose to align himself with; can we say honestly say that anyone devoted to the Nazi cause was worthy of the appellation "honorable"? It seems that one can be heroic or saintly in the service of an ignoble cause (indeed Arthur Schopenhauer makes precisely this criticism of the "virtue" of courage). Moreover, even if we assume that the cause is just, that does not rule out the possibility that atrocities can be committed in the name of this just cause (witness the torture of Iraqi prisoners of war by "heroic" American soldiers during Operation Iraqi Freedom). We may well be justified in asking whether we have been mistaken in considering supererogation a morally praiseworthy act at all; but this wholesale rejection is too strong. Despite the

---

<sup>14</sup>Guevara 1999. p. 598

<sup>15</sup>Ibid

<sup>16</sup>Guevara 1999. p. 598

potential for abuse, we do not want to “throw the baby out with the bath water”; if there is any potential moral worth in acts of supererogation (and it seems clear that there is) we should be able to account for it.

Is it the case then, that supererogatory acts are deemed exceptional owing to a failure of morality on behalf of humanity writ large? The world seems to abound with examples of moral failures in international affairs (e.g., the genocides in Germany, Rwanda, and Sudan) as well as in everyday life (e.g., widespread domestic violence). It is hardly surprising that acts that go above and beyond the call of duty often garner hearty praise. But it is interesting to note that Kant vehemently rejects the glorification of the hero as a moral paradigm as decidedly detrimental to morality; “When we can bring any flattering thought of merit into our action, then the incentive is already somewhat mixed with self-love and thus has some assistance from the side of sensibility”<sup>17</sup>. In other words, it is extremely difficult to know if the hero’s act was done *in deference to* the moral law or merely *in accordance with* the moral law.<sup>18</sup> Unless we adopt a strict utilitarian view, and consider only the consequences of one’s actions, forsaking motivation entirely, it is nigh impossible to uphold a publicity seeking daredevil as a paradigm of moral virtue. Moreover, it seems far too extreme to suggest that it is the hero or the saint alone who is truly moral; there is surely moral virtue to be found in everyday occurrences, amongst people who are clearly neither saints nor heroes. Nevertheless, we still acknowledge something exceptional about instances of supererogation. But it seems highly dubious to attribute the exceptional moral worth of such acts to the exceptional moral failures of the bulk of humanity.

Guevara’s criticism of the seeming inability of Kant’s theory to account for acts of supererogation seems to turn on the assumption that Kant cannot afford supererogatory acts their proper due, for the reasons considered above. However, in the *Critique of Practical Reason* Kant offers the caveat that infallible moral guidance may not be forthcoming in each and every instance of moral dilemma, and that in such instances the best anyone can do is to employ his or her judgment as best accords with certain broad guidelines.<sup>19</sup> Moreover, the moral law compels us to attempt to induce in others a feeling of happiness, as an unavoidable consequence of our beneficent actions. But if I cannot raise this feeling in someone whom I have attempted to benefit it may prove impossible to fulfill my particular duty in this instance. This consideration brings us to an important point I mentioned in passing earlier: it is not so very important for morality that Kant takes considerations of our happiness into account. What *is* of great importance is that Kant (and Kantians) *acknowledge the sacrifice* of our happiness in the fulfillment of certain moral acts. This acknowledgement is essential to the possibility of an account of supererogation in Kantian moral theory.

How then might a Kantian account for the moral worth of a supererogatory act? The answer is that she may not be able to offer a *direct* account of such worth. In the *Critique of Practical Reason* Kant notes that our moral decision making may be guided by a more or less intuitive understanding of certain broad moral guidelines. These guidelines (e.g., rational agency, autonomy, ends in themselves, dignity and respect, etc.) are routinely employed in moral decision making. Kant is

---

<sup>17</sup> Kant, Immanuel. *Critique of Practical Reason*. Cambridge University Press. Mary Gregor, trans. Cf.5:159. 1997

<sup>18</sup> Although it should be noted that this is the case in every moral act, according to Kant

<sup>19</sup> Kant, Immanuel. *Critique of Practical Reason*. Cambridge: Cambridge University Press. Mary Gregor, trans. Cf. 5: 132. 1997

adamant about the need to develop good moral judgment (and indeed it must be *cultivated* in Kant's view; it is *not* innate):

“Aiming at the highest good [i.e., happiness to everyone in proportion to their moral virtue], made necessary by respect for the moral law, and the presupposition flowing from this of its objective reality lead through the postulates of practical reason to concepts that speculative reason could indeed present as problems *but could never solve*”<sup>20</sup>.”

It may prove to be the case that in such extreme circumstances as the grenade example that consideration of these “broad moral guidelines” is not even possible owing to temporal constraints or general uncertainty as to one's duty in such a case. In this particular example the only broad moral guideline at our disposal would seem to be our own good moral judgment to preserve life to the extent that we are able. And indeed, Guevara recognizes the difficulty of appealing to a sense of duty in such circumstances, “Principles and rules can guide us, setting more or less specific presumptions in favor of this or that type of act...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>21</sup>. Insofar as a moral agent acted in deference to these broad moral guidelines (and by extension, in deference to duty), it is difficult to imagine that Kant could deny the moral worth of such an act. To be sure, he is adamant that such an act must not be considered a paradigm case of morality, but he would certainly acknowledge the magnanimous sacrifice of one's own life for the preservation of one's comrades. Despite the fact that the perfect duty to preserve oneself trumps any imperfect duty we owe to others, I contend that it is not completely out of the question that Kant would allow for something like supererogation in such extreme circumstances. Indeed, in such an instance as the doctor's volunteering in a plague-stricken city I contend that Kant would *not* consider such actions to constitute an example of an imperfect duty; the doctor has made the commitment to develop her talents and now seems obligated to use those talents and skills to preserve human life.

While it is certainly the case that we are obligated under Kantian moral theory to abide by our perfect duties where these conflict with our imperfect duties, I maintain that Kant does allow for the *possibility* that there may be situations of extreme duress which do not allow for the careful consideration of the fulfillment of our perfect moral duties. In such instances, in which the time for deliberation may only consist of a few seconds (or less), or in situations in which it is not at all clear just what our moral duty *is*, the supposedly inflexible nature of Kantian moral theory shows itself to be surprisingly practical and adaptable.

---

<sup>20</sup> Kant, Immanuel. *Critique of Practical Reason*. Cambridge: Cambridge University Press. Mary Gregor, trans. Cf. 5: 132. 1997. Emphasis mine.

<sup>21</sup> Guevara 1999. p. 612

**REFERENCES**

Cambridge: Cambridge University Press. 1997.

Guevara, Daniel. 1999. "The Impossibility of Supererogation in Kant's Moral Theory."  
*Philosophy and Phenomenological Research* 59, 593-624.

Hackett Publishing Company, Inc. 1981.

Kant, Immanuel. 1788. *Critique of Practical Reason*. Gregor, Mary J. Gregor, trans.

Kant, Immanuel. *Grounding for the Metaphysics of Morals*. James W.