Supererogation
Thomas E. Hill, Jr. and Adam Cureton

“Supererogation” is now a technical term in philosophy for a range of ideas expressed by terms such as “good but not required,” “beyond the call of duty,” “praiseworthy but not obligatory,” and “good to do but not bad not to do” (see Duty and Obligation; Intrinsic Value). Examples often cited are extremely generous acts of charity, heroic self-sacrifice, extraordinary service to morally worthy causes, and sometimes forgiveness and minor favors. These concepts are familiar in institutional contexts, for example, when teachers give points for “extra credit” work, corporations give “bonuses” for profitable leadership, and armies award medals for extraordinary service and valor. Moral philosophers, however, have generally focused on whether or not some of these terms refer to a fundamental moral category and, if so, how the category should be defined, which acts the category includes, and whether various moral theories adequately acknowledge it. The idea that acts can be good and praiseworthy but beyond duty has seemed puzzling for several reasons. For example, it seems that “good and praiseworthy” appeals to scalar standards of (“more or less”) value and virtue whereas “duty,” “required,” and “obligatory” invoke nonscalar norms that make (“all or nothing”) demands. This raises the theoretical question whether morality presupposes two distinct conceptual schemes, and, if so, to what extent and how these can be unified in a consistent and coherent moral theory.

Historically the idea of supererogation was developed within the Roman Catholic tradition to acknowledge that some lives achieve a higher level of moral perfection than can be demanded of everyone. Saints and martyrs, for example, displayed deeds of extraordinary self-sacrifice or exceptional commitment to charity, piety, and chastity that were deemed beyond the strictest duty, and perhaps even beyond...
the capacity of most other people. According to Aquinas, anyone would be blameworthy for violating the fundamental moral precepts, such as the prohibitions of murder and adultery, but counsels of supererogation present ideals of virtue that are meritorious to fulfill but not similarly blameworthy or vicious not to fulfill (see Aquinas, Saint Thomas). The distinction addressed an apparent tension between the strict commandments of the Old Testament and the call to a higher ideal found in the New Testament, but the relevant passages, Church doctrines, and examples remain controversial. Historically the most prominent controversy was over the idea that the supererogatory acts of saints produced a superabundance of “merit” that could be stored and distributed by the Church as “indulgences” to penitents to reduce the penance required for their sins. Protestants, such as Martin Luther, objected not only to the corrupt practice of selling indulgences but also to the very ideas that the merits of individuals can be stored and transferred and that human beings can achieve virtue or salvation by good works. In this Protestant view, human beings are so corrupt, our deepest motives so opaque, and duty so strict and far above how we actually live, that human morality has no place for supererogatory acts.

Apart from these historical controversies, questions about supererogation were raised again for contemporary philosophers by J. O. Urmson’s essay “Saints and Heroes” (1958). Urmson argued that utilitarianism and Kantian ethics have no place for the commonly acknowledged fact that some actions are praiseworthy but beyond duty (see utilitarianism; kant, immanuel; kantian practical ethics). Urmson noted that in certain cases we call a person saintly or heroic for doing what duty requires in such difficult circumstances that few others would do so. He argued, however, that in some cases we regard a person as a saint or hero in the further sense that the person does actions “far beyond the bounds of duty,” effortlessly or by extraordinary self-control.

Urmson’s challenge to utilitarianism was directed to the forms of “simple” utilitarianism prominent at the time of his essay, for example, the hedonistic utilitarianism of Jeremy Bentham (see bentham, jeremy) and the “ideal utilitarianism” of G. E. Moore (see moore, g. e.). Simple versions of hedonistic utilitarianism measure utility by pleasure and pain and classify an act as obligatory or right if it is the only one that maximizes (actual or expected) utility, wrong if it results in less than the maximum, and optional if and only if it would maximize utility and result in exactly the same utility as another act that would also maximize utility. Moore’s (1993) “ideal utilitarianism,” identifying utility with nonnatural intuited “intrinsic value,” also regards acts as objectively duty, wrong, or (rarely) indifferent depending on whether or not, compared to the agent’s other options, they actually result in the most intrinsic value possible. Thus, as Urmson says, these theories imply that there cannot be heroic or saintly deeds that are good to do but “beyond duty,” for these would be, strictly speaking, not duty, not wrong, and yet not indifferent. Hedonistic and ideal utilitarians could respond that, even if the agents’ particular acts fall short of what is objectively right, we should praise those who act from motives that are generally utility maximizing because praising such well-motivated acts tends to
promote the best results. For example, heroic but failed rescue attempts could be called “praiseworthy” in a sense even when, due to unforeseeable circumstances, they do not promote the most possible utility. Similarly, many kind and helpful acts could be praiseworthy but not strictly speaking duty simply because the agent could have done something even more utility promoting. Moore acknowledged that we cannot know with confidence what particular choices will result in the most possible intrinsic value, and so he thought that utilitarianism recommends that we generally follow the common moral norms of society. If our society’s moral code classifies some acts as “praiseworthy but not required,” then, Moore’s utilitarianism could endorse our use of this description, for practical purposes, even though it may lead us to praise acts that are actually not the best but, strictly speaking, indifferent or wrong. Many will probably find these concessions unsatisfying.

J. S. Mill endorsed Bentham’s general description of the aim of utilitarianism as “the greatest happiness,” but he added significant qualifications that seem to allow more room for acts that are “good to do but not required” (see MILL, JOHN STUART). In Chapter 2 of his Utilitarianism he says that the utility principle “holds that actions are right in proportion as they tend to promote happiness, and wrong as they tend to produce the reverse of happiness” (Mill 1998: 2.2). In Chapter 5, however, he adds qualifications that make clear that not all acts that utilitarianism commend as contributing to the greatest happiness are matters of duty that one would be wrong not to fulfill. “We do not call anything wrong unless we mean to imply that a person ought to be punished in some way for doing it – if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience…. Duty is a thing which may be exacted from a person, as one exacts a debt” (Mill 1998: 5.14). Whether people ought to be punished for their acts or omissions must be determined by the utility principle. The question for Mill, then, is whether it tends to promote the greatest happiness for society to use legal coercion, social pressure, and cultivated guilt feelings to get people to behave as they should. Making all failures to maximize utility punishable by law, opinion, or conscience would presumably not promote the greatest happiness. For example, although the benefits generally outweigh the costs of punishing murder by law, extreme rudeness by social sanctions, and neglect of charity by a nagging conscience, it does not seem similarly beneficial to punish people for not making heroic and saintly sacrifices whenever such sacrifices would contribute, even in a small way, to the greatest happiness. Mill did not explicitly address questions about supererogation, and scholars disagree about how he understood the role of rules in his theory. Nevertheless, it seems that by adopting distinctions like those that Mill proposed, utilitarians could affirm that some acts are “beyond duty” – not wrong, not indifferent, but “good to do” and praiseworthy.

Kant’s moral theory, as Urmson says, also seems to be in conflict with common moral opinion about praiseworthy acts “beyond duty,” but, as in the case of utilitarianism, the issues are more complex than they first appear. Kant, who was raised in a Protestant environment, objected to the use of examples of superheroic deeds in moral education. Instead he emphasized the difficulty and importance of doing one’s duty even in ordinary cases. Only acts done from duty or respect for the moral
law have “moral worth.” Even “heroic” and “saintly” acts would be lacking in moral worth if they were motivated by sympathetic feelings. What is “good” to do is what reason prescribes, and what is unconditionally good to do is categorically imperative for us as imperfectly rational beings. In Kant’s theory ethics does not propose, prior to the standards of rational willing, any independent substantive goal (such as the greatest happiness or most intrinsically valuable universe) that could determine what is duty or “good to do beyond duty” (Kant 2002).

Despite these anti-supererogation starting points, Kant’s moral theory is complex in ways that may partially accommodate the common opinion that some acts are praiseworthy though not required. In his late work *The Metaphysics of Morals* (1996), Kant lays out a system of rational ethical principles more fully and specifically than he did in his earlier *Groundwork for the Metaphysics of Morals* (2002). All the principles, apart from law and justice, prescribe ends that we must adopt. The substantive ends of one’s own perfection and the happiness of others are based on “humanity” as an end in itself. To promote the happiness of others is a “wide imperfect duty” that does not determine specifically what, when, or how much one is to do for others but leaves some “playroom” for free choice (see imperfect duties). This seems to imply that often (though not always) when a properly beneficent person does something charitable, the particular act on that occasion was not required though praiseworthy (because motivated by commitment to the general end required by duty). Acts can be “meritorious,” Kant says, when a person does “more in the way of duty than he can be constrained by law to do” (Kant 1996: 6.228). The more obvious point is that the laws of the state cannot compel one to fulfill duties of virtue, for example, making it one’s end to perfect oneself and promote the happiness of others. Arguably, however, Kant’s theory allows that extraordinary commitment and efforts to further those moral ends in especially trying circumstances are meritorious in the further sense that these are “more in the way of duty” than the moral law categorically requires. If the wide imperfect duty of beneficence does not specify the extent of what must be done and leaves room for the pursuit of one’s own happiness, then at least some extraordinary sacrifices for others’ happiness would be more toward a moral end than necessitated by the (obligatory) sincere and well-motivated adoption of that end. Even so, Kant implies that all virtuous acts must be ultimately rooted in the agent’s dutiful respect for moral law, and for this reason they will not count as supererogatory on some interpretations.

Utilitarians and Kantians have struggled in various ways to interpret and make room for the commonsense notion of acts that are good or praiseworthy to do but not required. Moral theories that admit supererogatory acts, however, face a further challenge, raised by Susan Wolf in her paper “Moral Saints” (1982). Wolf argues that even though these traditional moral theories do not require us to work tirelessly for the benefit of others, for example, they nonetheless imply that we would be better, more excellent people, or following the best reasons, if we were always going “above and beyond” the call of duty. Once we identify moral considerations that justify praising or admiring a saint or hero, or warrant regarding her permissible acts as
good, haven't we thereby admitted that our own imperfect lives do not "measure up" to that moral ideal? If there is substantial room for moral improvement even among those of us who are doing our duty, it seems to follow that we have most moral reason to strive toward heroism or saintliness. Devoting ourselves to these moral ideals, Wolf worries, will inevitably come at the expense of other "perfectly wonderful" nonmoral virtues, projects, commitments, ideals, and aspirations. Wolf finds this transcendent and overriding conception of morality, which she detects even in sophisticated utilitarian and Kantian views that admit supererogation, objectionably imperialistic and demanding, but the primary target of her criticism may be the common idea, which is arguably one that Kantians do not accept, that good or praiseworthy acts are such that the more of them we do the better or the more praiseworthy we are.

More recent discussions of supererogation often take place against the background of a simple moral framework that shares this basic feature. On this view, there are moral and nonmoral reasons that are distinguished intuitively, all reasons are commensurable with one another, the moral ones are always overriding, and what we are required to do is a matter of what we have most reason to do (see reasons for action, morality and). If supererogatory acts are good to do because they are supported by moral reasons then, from this moral perspective that takes moral reasons to be overriding, doesn't the balance of reasons overall actually favor performing those acts, making them required rather than optional? The worry is that, on this simple moral framework, there is no room for supererogation. If an act truly is heroic or saintly, for example, then it would seem that refusing to do it would be wrong because we would be failing to act on the best reasons that apply to us. Responses to this puzzle have tended to focus on whether we have a "personal prerogative" or "agent-centered option" (Scheffler 1982: 3) to pursue our own interests when doing so is not favored by the balance of moral reasons and, if we have such a permission, what form it should take (see agent-centered options). Douglas Portmore (2003), for example, gives up the idea that moral reasons are always overriding and allows that the balance of reasons may sometimes favor self-interest over morality. According to him, it is true that we have most moral reason to perform supererogatory acts, but we may not be required to act in those ways all-things-considered because of sufficiently strong nonmoral reasons to pursue our own plans, projects, and interests. If we nonetheless choose in such cases to sacrifice ourselves for the greater good then our actions would be supported by morality, but they would nonetheless be irrational because the overall balance of reasons in such cases favors protecting our own self-interest. Joseph Raz (1975) suggests another possibility that aims to avoid this implication. Supererogatory acts, on his view, are favored by the overall balance of reasons, not just the moral ones, but he rejects the idea that we must always act on the best reasons. According to him, we sometimes have a permission, which we can choose to exercise or not, to refuse to do what we have most reason to do. Having this permission allows that we have most reason to perform supererogatory acts even though it can be rational for us not to do so. Jonathan Dancy (1993: 139) takes a slightly different approach
by examining the nature of a certain class of nonmoral reasons, namely those that are grounded only in harm to ourselves. The strength of these reasons, according to Dancy, can be in part a matter of our own choosing, so the overall balance of reasons will partially depend on how much we discount harm to ourselves. We typically count great personal sacrifice as a strong consideration against doing what morality requires of us on a particular occasion, but such acts are supererogatory because we may discount those costs and so change the balance of reasons in favor of performing the act.

A different approach is to question the simple model more directly. One way of doing this is by making room within morality itself for personal projects, values, ideals, and other aspects of reasonable self-interest. Another is to interpret the idea that morality is overriding as the claim that our all-things-considered duties place unconditional demands on us, even though particular moral reasons that are grounded in those duties may not always take priority over reasons that stem from self-interest. We may also appeal to John Rawls’ (1999: 100) view that certain types of acts, such as bringing about a great good, are generally required of us, but if performing one of those acts on a particular occasion would bring excessive risk or harm to the agent who did it then that person is released from his duty to perform the now supererogatory act (see Rawls, John).

Once we have settled on a suitable conception of supererogation, we face further questions about how that notion fits in with our duties, obligations, rights, virtues, and other aspects of our moral lives. Frances Kamm (1985) argues for the paradoxical claim that we are sometimes permitted to violate our duties in the name of supererogation. One of her examples is of someone who promises to attend a lunch meeting but on her way encounters an accident victim who immediately requires a kidney that she is willing and able to give even though doing so would cause her to miss the appointment. Kamm’s analysis of the case is that, while the woman would admittedly be exempted from her fiduciary obligation by the more stringent duty to relieve great harm at little cost to oneself (common sense says that it would not be wrong for the woman to miss her meeting in order to provide cardiopulmonary resuscitation, for example), there is no corresponding exception to her fiduciary obligation for supererogatory acts. So, according to Kamm, her obligation to attend the meeting remains intact, but she is nonetheless allowed to violate that requirement in order to perform the supererogatory act of donating a kidney. A different way of approaching the example, one suggested by Alan Donagan, is to say that it is always impermissible to provide benefits to others by performing “any action that is impermissible in itself” – many of us would not even consider killing a rich benefactor, says Donagan, to secure the needed funding for a free clinic in a poor community (1977: 86). Donagan thinks that breaking promises to perform permissible acts is impermissible in itself (1977: 92), so his view implies that the woman in Kamm’s example can certainly work to be released from her obligation to attend the meeting, but if she is not successful at doing so, she should keep her promise and only perform acts of supererogation when doing so is within the bounds of her duty. Rawls (1999: 298–9), however,
suggests that matters may be more complicated than both of these views. He thinks that *presumptive* duties about mutual aid, killing, torture, and so on must be interpreted and arranged into a consistent and coherent whole that specifies what our duties are. The moral category of the supererogatory must also be interpreted and explained. Although Rawls famously claims that the “right” always takes priority over the “good” – we should never enslave or torture someone to further the mere personal projects of others – he also thinks that our presumptive duties and obligations are not always prior to supererogatory acts. We must try to work out systematically, then, how our duties and obligations are to be balanced against supererogatory acts when conflicts arise, and we must do so without the benefit of obvious priorities among these things. Although Rawls’ views on supererogation are sketchy, his view may imply that, in stopping to donate her kidney, the woman is making use of an exception that should be built into our rules about promising alongside other recognized ones.

When discussing supererogation, there are several important distinctions we should make that are not always recognized and respected. (1) In ordinary language, we sometimes say that something is a moral duty, social obligation, military duty, family obligation, legal obligation, or duty to king and country (Brandt 1964). These more informal notions of “duty” and “obligation” may have some hold on our intuitions about what it takes to go “above and beyond the call of duty” even though they tend not to figure in abstract philosophy, they are not necessarily binding on us, and they can come into conflict with one another. To be clear, we should ask whether the supererogatory acts under discussion are supposed to be *not morally required all-things-considered* or merely beyond duty and obligation in one of these narrower senses. (2) Acts have been considered required, or not, from different perspectives, for example, the point of view of morality, rationality, self-interest, and perfection of other sorts. We need to ask, then, whether the requirements under discussion (that supererogatory acts are regarded as exceeding) are meant to be moral duties, requirements of reason, necessary means to personal happiness, the demands of a different kind of ideal, or some combination of these. (3) When we judge that something is good, we may not always imply that the more of it there is the better – value may not always aggregate. Value judgments can also be about the reasons we have to respect, honor, and cherish the thing itself, without always implying that we have reason to bring more of it about. We should ask, then, whether or not, when we think of supererogatory acts as “good to do but not required,” we mean that they are of some specifiable kind (for example, acts that promote happiness or acts of unusual charity) that we think, though not always obligatory, always count positively on a commensurable scale of value that determines what is best to do or (potentially) most praiseworthy. (4) T. M. Scanlon (2008) distinguishes between the *moral permissibility* of acts and the *meaning* acts have for our relationships with others and so our judgments about how *blameworthy* we are for performing them. The permissibility of an act, according to Scanlon, depends on what reasons there in fact are for and against doing it, which may be independent of the agent’s attitudes, intentions, and reasons in performing it,
whereas the meaning of an act does depend on how the agent sees the situation, what reasons she takes there to be, what intentions she is acting on, and so on. If we accept this distinction, then, we need to ask whether or not the supererogatory acts under discussion are regarded as both permissible independently of the agent’s intentions and motives and praiseworthy when these factors are taken into account. (5) Finally, returning to the religious roots of supererogation, there is a significant difference between what it is about a person that Protestants and Catholics of the kinds described earlier are praising when they praise her for performing good acts. These Protestants tend to see good acts as manifestations and evidence of something good that is in her, which she has in virtue of God’s grace, her faith, her commitment to God, or (according to Kant) her good will. These Catholics, on the other hand, tend to regard good acts as points in her favor, merits that figure in the overall record of her life that determines what rewards, penance, and punishment she deserves now or later.

SEE ALSO: AGENT-CENTERED OPTIONS; AQUINAS, SAINT THOMAS; BENTHAM, JEREMY; DUTY AND OBLIGATION; IMPERFECT DUTIES; INTRINSIC VALUE; KANT, IMMANUEL; KANTIAN PRACTICAL ETHICS; MILL, JOHN STUART; MOORE, G. E.; RAWLS, JOHN; REASONS FOR ACTION, MORALITY AND; UTILITARIANISM

REFERENCES

SUGGESTED READINGS

Aquinas, Thomas. *Summa Theologica*, esp. I, q. 21; I, II, q. 108; II, II, q. 31, q. 32, q. 106, q. 184; and suppl., q. 25.


