Supererogation and Doing the Best One Can

By: Michael J. Zimmerman

Zimmerman, Michael J. Supererogation and doing the nest one can. American Philosophical Quarterly 30(4), October 1993.

Published by the University of Illinois Press on behalf of North American Philosophical Publications

The original publication is available at http://www.press.uillinois.edu/journals/apq.html

***Note: Figures may be missing from this format of the document

Article:

A thorough philosophical inquiry into supererogation would need to deal with the following four questions. What is the nature of supererogation? Are supererogatory acts possible? Do they actually occur? Are those acts that we commonly call supererogatory in fact supererogatory? There is a certain progression to these questions. Establishing the nature of supererogation does not ensure that supererogatory acts are possible. (Compare inquiring into the nature of a squircle and establishing that it is a square circle.) Establishing the possibility of supererogatory acts does not ensure their actuality. Establishing their actuality does not ensure that those acts that we commonly call supererogatory are in fact so.

These four important questions will be addressed only obliquely here. The question of direct concern is this. Does a theory according to which what one ought to do is the best that one can do imply that supererogatory acts are impossible? The reason for raising this question is that many philosophers accept such a theory and many philosophers also believe that such a theory does indeed have this implication. Yet it will be argued here that the implication does not hold. It will be argued, that is, that such a theory can in fact accommodate supererogation.

Let us call a theory of the sort in question a maximizing theory. According to it, what one ought to do is maximize some value. This may seem to rule out the possibility of supererogatory acts, inasmuch as such acts would seem at once both good and optional. If one ought to do the best one can, how can any good act be optional? Yet it is a commonplace that maximizing theories do allow for the possibility that some good acts be optional. According to such theories, an act is morally wrong (the agent ought not to perform it) if and only if some alternative act has a superior value; an act is morally right if and only if no alternative act has a superior value; and an act is morally obligatory (the agent ought to perform it) if and only if every alternative act has an inferior value. Thus an act is optional (in that it is neither wrong nor obligatory to perform it) if and only if no alternative act has a superior value but also not every alternative act has an inferior value. An act is optional, that is, if and only if it is right and at least one alternative is also right.

Yet it is a commonplace that maximizing theories do allow for the possibility that some good acts be fully optional. According to such theories, an act is morally wrong (the agent ought not to perform it) if and only if some alternative act has a superior value; an act is morally right if and only if no alternative act has a superior value; and an act is morally obligatory (the agent ought to perform it) if and only if every alternative act has an inferior value. Thus an act is optional (in that it is neither wrong nor obligatory to perform it) if and only if no alternative act has a superior value but also not every alternative act has an inferior value. An act is optional, that is, if and only if it is right and at least one alternative is also right.

Here it might be objected that, while maximizing theories do admittedly imply that some acts can be optional in this way, they don't allow for the possibility that acts be fully optional, whereas supererogatory acts are not just optional but fully optional. An act is fully optional only if it is not just not obligatory but also such that its performance does not constitute a way to fulfill some obligation; it is, as it were, wholly beyond the reaches of obligation. Now it is true that a maximizing theory is hardly likely to classify an optional act as fully optional. Consider a case that has this form (over-simplified for the sake of illustration): the agent has four and only four alternatives, A, B, C, and D, whose values are, respectively, 10, 10, 5, and 0. In such a case, A and B are optional, while C and D are wrong. But neither A nor B is fully optional, since there is still an obligation lurking not too far in the background, namely, the obligation to do either A or B. Thus,
although the agent is not obligated to do A and is not obligated to do B, performance of either will fulfill an obligation.

But here two points are to be noted, one minor, the other not. The minor point is this; a maximizing theory can in principle allow for the possibility of fully optional acts; just consider a revised version of the case, where the values of not just A and B but of C and D also are 10. The much more important point is this: it is not true that supererogatory acts must be fully optional. Consider the original case. Why is it that it is obligatory to do either A or B? The answer is, of course, that C and D—that is, all the agent's other alternatives—are wrong. In doing either A or B, the agent is thus fulfilling the obligation not to do something wrong. It cannot be accepted that this alone renders both A and B non-supererogatory. After all, it is surely possible that a supererogatory act have an alternative that is wrong, so that, if the supererogatory act is performed, that very performance will constitute the fulfillment of an obligation, to wit, the obligation not to perform the wrong act.

Why is it, then, that maximizing theories have been thought not to accommodate supererogation? Well, consider the classic maximizing theory of act utilitarianism, which is traditionally seen to be an amalgamation of hedonism (the view that the only thing that is intrinsically good [bad] is a state of pleasure [pain]) and consequentialism (the view that one ought to produce the most favorable balance of intrinsic goodness over intrinsic badness that one can). It is often said that act utilitarianism rules out supererogatory action, but this is questionable. What it certainly does do is imply that many of the acts that we would normally call supererogatory are in fact not so, since it implies that these acts are either wrong or obligatory. But this is not the same as implying that there are no supererogatory acts, let alone the same as implying that there can be no such acts. Indeed, the consequentialist aspect of act utilitarianism (which is what makes it a maximizing theory) rules in the possibility of optional acts (as long as ties in intrinsic "bestness" are in principle possible, and there seems to be no reason to deny this). In view of this it might seem that it is hedonism, and not act utilitarianism's maximizing aspect, that is the source of trouble here and that, were consequentialism conjoined with some other, less "single-minded" theory of value, it would be seen that supererogation can easily be accommodated by a maximizing theory.

But actually it is not that easy. If we turn from the monism of act utilitarianism to the pluralism of G. E. Moore, we find that Moore's theory, too, has a hard time accounting for supererogation. While Moore's theory (which allies consequentialism with a theory of value according to which not just states of pleasure and pain but also other states—knowledge, moral qualities, beauty—have non-neutral intrinsic value) of course allows for the possibility of acts' being optional—as does any consequentialist or, more broadly, maximizing theory—it seems not to square with our common intuitions concerning supererogation any better than act utilitarianism does. For example, just as act utilitarianism implies that many instances of self sacrifice that we would normally call supererogatory are in fact either wrong or obligatory, so too does Moore's theory imply this. If one can maximize intrinsic value only by way of extreme self-sacrifice, then Moore's theory implies that one must, one is obligated to, undertake such sacrifice.

So it is not just hedonism, but consequentialism too, that would appear inimical to our common intuitions concerning supererogation, and it has already been noted that it is their consequentialist aspect that renders both act utilitarianism and Moore's theory maximizing theories. Must we conclude that all maximizing theories are similarly inimical? Not at all. It could be that the problem lies with consequentialism in particular rather than maximizing theories in general. While consequentialism allows for acts' being optional, its focus on the promotion of intrinsic value neglects the fact that it is in the area of personal interest that we commonly look for a moral theory to give the agent some "slack," some latitude as to how to proceed. What, then, if a maximizing theory were to admit into the final determination of right and wrong certain agent-relative values? What if, in the final computation of what is best and what not, an agent's own projects and commitments were given a weight disproportionate to that which their intrinsic value alone supplies? Will we then have found a way to reconcile maximization with supererogation?
Again, it is not that easy. Once more, there are two points to be made, one minor, the other not. The minor point is that it would be a mistake to think that supererogation essentially concerns agent-relative values. Some theory might claim that, while it is obligatory to help human beings whenever possible, it is supererogatory to help non-human animals. This is intelligible, even though what is said to be supererogatory has no essential tie to agent-relative values. The much more important point is this: even though a maximizing theory that incorporates agent-relative values in the way indicated allows, as do all maximizing theories, for certain acts to be optional, it does not appear to allow for certain acts to be discretionary in the way that supererogation requires. Consider a case where $A$ and $B$ are each optional, according to the maximizing theory in question, and where their values are, respectively, predominantly agent-neutral and predominantly agent-relative. For example, $A$ might be a charitable act of some sort (in which the agent has no vested personal interest), while $B$ is the purchase of some high-priced compact disk player (in which the agent has considerable vested personal interest). Now consider a very slightly modified case, where $B$'s value marginally overrides $A$'s because the agent's vested personal interest is marginally increased. The maximizing theory would imply that in this modified case neither $A$ nor $B$ is optional; on the contrary, $A$ is wrong and $B$ obligatory! Similarly, if $A$'s value were marginally to override $B$'s, then the maximizing theory would once again imply that neither $A$ nor $B$ is optional; on the contrary, $A$ would be obligatory and $B$ wrong. In such cases, engaging in the sort of self-sacrifice that $A$ would entail is clearly not left to the agent's discretion.\(^7\)

It is probably at this point that many will diagnose a fatal flaw in all maximizing theories. It is not, as it may at first have seemed, that such theories cannot admit that certain acts are optional; they can. It is rather that such theories cannot admit that certain acts are discretionary, in the sense that it is left to the agent's discretion which of two (or more) acts to perform when these acts are of different values. This, it might be claimed, is something that is essential to supererogation, but something that maximizing theories cannot accommodate.

The point might be put this way.\(^8\) In those cases where it is left to the agent's discretion whether to do $A$ or $B$, where $A$ would be supererogatory and $B$ not, $A$ is superior in value to $B$. In such cases, maximizing theories imply that neither $A$ nor $B$ is optional; on the contrary, $A$ is obligatory (in the absence of any third alternative of equal or superior value) and $B$ wrong. Thus maximizing theories imply that supererogatory acts are impossible.

(An analogous argument can be made concerning what some have called offensive acts\(^9\) but which might be better called suberogatory acts. Just as a supererogatory act $A$ is superior to some alternative $B$ and yet each is optional, so too a suberogatory act $C$ is inferior to some alternative $D$ and yet each is optional. Just as maximizing theories would appear to imply that $A$ is obligatory and $B$ wrong, so too they would appear to imply that $C$ is wrong and $D$ obligatory.)

Once the point is put this way, though, the solution to the puzzle becomes clear. Somehow a maximizing theory must endeavor both to have its cake and to eat it—that is, endeavor both to say that $A$ is superior in value to $B$ and yet to say that $A$ and $B$ are of the same value. This can be achieved by, and only by, declaring that there is more than one set of values pertinent to the moral evaluation of an act.\(^10\) One set of values—call it the deontic set—must be said to be pertinent to the determination of right and wrong and obligation, the other—call it the non-deontic set—not. Where $A$ is supererogatory but it is left to the agent's discretion whether to do $B$ instead, a maximizing theory must declare $A$ to be non-deontically superior to $B$ but deontically equivalent to $B$.\(^11\)

Some may find such a division of values to be a *deus ex machina*, yet appeal to it is routine. When someone is told that he should or ought to do some act although he doesn't have to,\(^12\) what's meant is that it is good in a way that doesn't render it obligatory. (Thus "ought" here does not express obligation, contrary to its use earlier in this paper.) Why should this common appeal to a double standard\(^13\) be deemed unavailable to a
maximizing theory? The problem with most maximizing theories is that they don't make such an appeal, not that they can't, and this is why they cannot (as they stand) account for supererogation.

This is not to say that it is easy to see just how a maximizing theory should invoke such a division of values. It is perhaps easier to see how this might be done when the non-deontic values are not said to supply a reason for acting, harder when they are said to supply such a reason. Whether or not what's good about a supererogatory act is something that provides a reason to perform it is a matter of some dispute (having to do with just what the nature of supererogation is). On some accounts of supererogation, what's good about a supererogatory act is that it or the agent is praiseworthy; on some accounts of praiseworthiness, an act or agent is rendered praiseworthy by the motive underlying the act; on some accounts of motivation, the motive underlying an act is beyond the agent's control; and, on some accounts of reasons, what is beyond the agent's control cannot feature in a reason to perform an act. Thus some accounts will declare a supererogatory act A and a non-supererogatory alternative act B to be wholly on a par with respect to the reasons that exist for doing them; the non-deontic superiority of A to B will be explained without appeal to reasons for action.

But other accounts of supererogation will declare the superiority of A to B to constitute a reason for doing A rather than B. If this is essential to supererogation, then a maximizing theory must declare not only that there are two sets of moral values but that there are two sets of moral reasons for acting. But, while it may be difficult to see just how such an account should go, it is clearly not without precedent. On the contrary, the common invocation of what is ideal rather than obligatory seems to constitute an appeal to a non-deontic set of reasons for acting. At any rate, despite whatever difficulty there may be in working out the details of such an account, my point here is simply this: the abstract strategy of dividing moral values (and perhaps, thereby, moral reasons for acting) is in principle available to a maximizing theory; and this strategy (unless there are further pertinent restrictions on supererogation that have been overlooked in this paper) allows such a theory to accommodate supererogation.

Even if this point is accepted, though, it might be thought that such a strategy is too high a price for any theory to pay, and that a non-maximizing theory would come more cheaply. But this is not so, for two reasons. First, there is reason to think that there is no such thing as a non-maximizing theory of obligation. That is, there is reason to think that every theory of obligation can be appropriately cast in a maximizing mold. Certainly, in addition to act utilitarianism and Moore's theory, rule utilitarianism ("It is obligatory to do what is required by maximization of rule-utility") and Ross's theory ("It is obligatory to do what is required by maximization of the stringency of prima facie duties") can be seen in this light, and perhaps Kant's can also ("It is obligatory to do what is required by maximization of compliance with the Categorical Imperative") as well as other prominent theories, such as (one version of) the divine command theory ("It is obligatory to do what is required by maximization of compliance with God's commands"). The fact that all but one act may in some sense be a maximal failure to do what is morally required (as, e.g., when only A complies in any way with the Categorical Imperative, or only A complies in any way with God's commands) does not detract from the fact that the one act in question is, in the relevant sense, best. Granted, representing some moral theories as maximizing theories may be in some way "gimmicky," but this does not mean that it is inaccurate. And if it is accurate to represent every theory in this way, then any theory that accommodates supererogation can do so only in the way just stipulated for maximizing theories.

The second point, though, is this. Even if it is false that all theories of obligation can be accurately represented as maximizing theories, it would still seem that no theory can accommodate supererogation without invoking a division of values of the sort mentioned. If an act is to be good in a way that does not render it obligatory, its goodness must, it seems (but clearly more needs to be said here), reflect a value that its deontic status does not. Thus maximizing theories would appear to be no worse off in this regard than any other theories (if there are any) with respect to the accommodation of supererogation. If this is so, then the fact that a theory is a maximizing theory is to this extent no strike against it when it comes to such accommodation, since the strategy that it must adopt is the strategy that any theory must adopt.
This is not, of course, to say that every maximizing theory can (as it stands) accommodate supererogation. Those theories that deny any division of moral values cannot. Others that do allow for such division may nonetheless undermine themselves. For example, any theory that says that something's being good from the ideal point of view is itself a deontic reason for trying to achieve it, would appear to have undercut any advantage that the division of values initially afforded it with respect to the accommodation of supererogation. Ross's theory seems to have just this difficulty. In distancing himself from consequentialism by divorcing the right from the good, Ross may at first seem to have proposed a theory that can account for supererogation in a way that Moore's theory cannot. For the special prima facie duties of which Ross speaks that vie for the status of duty proper do not owe their deontic status to the fact (if it is a fact) that the acts in question, or what they bring about, are in themselves desirable. Their deontic ground lies elsewhere, and this would appear to allow Ross to declare that the desirability of an act, or of what the act brings about, provides at most a non-deontic reason for its performance. But Ross does not say this. Rather, he declares the desirability of an act, or of what it brings about, itself to be a deontic reason for performing the act; there is, he says, a non-special, general duty to maximize intrinsic value. He thus seems no better able than Moore to accommodate supererogation.

The general point is again simply this: if a theory is to accommodate supererogation, it must divide values. This does not mean that there cannot be both a deontic and a non-deontic reason for performing one and the same act; but it does mean that these reasons must stem from distinct grounds. For example, some theory might claim that I have a deontic reason to help you, stemming from my promise to do so, and also a non-deontic reason to do so, stemming from the deontically irrelevant goodness of your welfare. Such a theory could then declare my helping you beyond the extent to which I have committed myself supererogatory. But if the theory were to claim that your welfare is deontically relevant, as Ross's does, it becomes hard to see how my helping you could be supererogatory, whether or not I go beyond the extent to which I have committed myself (unless there is yet a third value—a non-deontic one—that is promoted by my helping you).

Finally, two cautionary notes.

That maximizing theories can in principle allow for the possibility of supererogation is one thing; that it is plausible to declare any acts to be in fact supererogatory is quite another. For all that has been said here, it may be that dividing values in the way that supererogation requires is unjustifiable. Certainly, one would like to see a detailed justification for doing so.

And even if such a division were given and were justified, it might of course not imply that those acts that we commonly call supererogatory in fact are so. Nothing that has been said here shows our common opinions on this matter to be acceptable. On the contrary, there is very good reason to believe that they are not.

NOTES

1. G. E. Moore is perhaps the prime exemplar, in Ethics (Oxford: Oxford University Press, 1965), while many others who embrace some form of utilitarianism—such as Jeremy Bentham, in An Introduction to Morals and Legislation (New York: Methuen, 1982); John Stuart Mill, in Utilitarianism (Indianapolis: Bobbs-Merrill, 1957); J. J. C. Smart, in "An Outline of a System of Utilitarian Ethics", in Utilitarianism: For and Against by J. J. C. Smart and Bernard Williams (Cambridge: Cambridge University Press, 1973); and Fred Feldman, in Doing the Best We Can (Dordrecht: D. Reidel, 1986)—are also naturally viewed as accepting such a theory. As will be suggested below, however, it may be appropriate to view many non-utilitarians in this way too.

2. Included in this number are some who accept and also some who reject such a theory. See, for example, in addition to those cited in note 5 below, the following, all of whom appear essentially to hold the opinion in question: Moore, op. cit., p.78; David Heyd, Supererogation (Cambridge: Cambridge University Press, 1982), p. 78; Samuel Scheffler, The Rejection of Consequentialism, pp.
It is assumed here that a necessary condition of an act's being supererogatory is that it be optional *all* morally relevant things considered. (Cf. Heyd, *op. cit.*, p.115.) Thus it is not enough that the act simply "go beyond" some narrowly specified, role-related moral duty or responsibility (cf. Christopher New, "Saints, Heroes, and Utilitarians," *Philosophy*, vol. 49 (1974), pp. 179-89, p. 179ff.). Thus, too, there can be no all-things-considered moral duty to perform it of any sort, not even what Kant calls an imperfect duty (cf. Heyd, *op. cit.*, pp. 58, 121-22, 125; Marcia Baron, "Kantian Ethics and Supererogation," *The Journal of Philosophy*, vol. 84 (1987), pp. 237-62, p. 244).

*Cf.* Slote, *op. cit.*, p. 3: "...consequentialism, precisely because it insists that one do the best one can, leaves no room for moral supererogation..."


10. Note: *moral* evaluation. Supererogatory acts are in some way *morally* superior to their non-supererogatory alternatives. Simon seems to miss this in her interesting paper, "Supererogation: Puzzle or Pseudo-Problem?", forthcoming.

11. Note that the claim here is simply that *A* is superior in some way to *B*. This is not to say that *all* supererogatory acts are superior to *all* non-supererogatory acts. Suppose that fidelity is a deontic value and that beneficence is a non-deontic value. In Situation 1, I have two optional alternatives: contributing to your welfare to a small degree (act *C*), and not contributing to your welfare at all (act *D*). On the assumption that beneficence is a non-deontic value, we can imagine that *C* is supererogatory and *D* not, because *C* is superior to *D* with respect to beneficence. In Situation 2, I have two alternatives: contributing to your welfare to a great degree (act *E*), as I have promised to do, and not contributing to your welfare at all (act *F*). On the assumption that fidelity is a deontic value, we can imagine that *E* is obligatory and *F* not, because *E* is superior to *F* with respect to fidelity. Notice that *C* is not superior to *E* with respect to beneficence; on the contrary, *E* is superior to *C* in this regard. Nonetheless, *C* is supererogatory while *E* is obligatory. This is because *C* and *E* are not alternatives to one another; they pertain to different situations.


16. It may appear that a pertinent restriction has been overlooked, namely, one that Heyd calls the "condition of continuity." In *op. cit.*, p. 5, Heyd says: "Not every non-obligatory act is supererogatory. It has to be morally good, its value being of the same type that makes obligatory action good and valuable... [T]here is a common and continuous scale of values shared by supererogation and duty." Two questions arise here. First, is Heyd correct? Secondly, is Heyd's emphasis on *continuity* inconsistent with the strategy of *dividing* moral values that has been adopted in this paper? The answer to both questions would seem to depend on just what it is that Heyd is
claiming, although on no interpretation does it seem that the answer to both is "Yes." There appear to be three main interpretations. Heyd might simply be claiming that supererogatory acts must exhibit some moral value, just as obligatory acts must exhibit some moral value. Far from being inconsistent with this claim, the strategy adopted here presupposes its truth. On the other hand, Heyd might be claiming that supererogatory acts and obligatory acts can exhibit only one sort of moral value, the former simply exhibiting more of this value than the latter. This claim is indeed inconsistent with the strategy adopted here, but there would appear to be no reason to think it true (although it is true that at some level supererogatory acts and obligatory acts may reflect a common value—see note 21 below). Another possibility, which seems to be suggested on p. 5, ibid., when Heyd says that supererogatory acts that are cases of "oversubscription" (such as paying more than is owed) can easily be seen to satisfy the condition of continuity, is this: that (a) supererogatory acts and obligatory acts must both be morally valuable, and (b) a supererogatory act must be an instance of a nonmoral type of which a certain non-supererogatory, obligation-satisfying alternative act (which the supererogatory act "goes beyond") is also an instance. It has been stated here that (a) is clearly true while (b) is problematic. Still, nothing that has been asserted in this paper is inconsistent with (b), as is shown by the illustration given below of my helping you beyond the extent to which I have committed myself.

19. Certainly no proof has been given here that all theories can be so represented. The general idea underlying this view, though, is simply this: an act's being morally obligatory consists in its being in some way uniquely valuable relative to its alternatives, and this uniqueness can be represented by ascribing it a number superior to any numbers ascribed to its alternatives.
21. At least this is so in terms of what may be called the "direct" reflection of values. Terrance McConnell has pointed out to me that a certain version of rule utilitarianism might state that an act is morally obligatory if and only if it conforms with a correct moral rule that prescribes performing it, and that a rule is correct if and only if it (in some way) promotes welfare better than any of its alternatives. On such a theory, it might happen that a certain act is supererogatory because of the welfare it promotes, even though the determination of what is obligatory also turns on the promotion of welfare. It might thus seem that the value reflected by a supererogatory act's non-deontic status can be exactly the same value as that reflected by its deontic status. Now, it must be granted that this sort of case shows that an act that is supererogatory may owe its non-deontic status directly to a certain value to which it indirectly owes its deontic status. Nonetheless, it would seem to remain impossible that an act that is supererogatory should owe its non-deontic status directly to a certain value to which it directly owes its deontic status. In the case at hand, it is not the promotion of welfare that directly determines the deontic status of an act; rather, whether an act is morally obligatory, wrong, or optional is directly determined by and only by its conformity, or lack of conformity, with a correct moral rule.
25. I am grateful for help from Neera Badhwar, Terrance McConnell, Gregory Mellema, Gregory Velazco y Trianosky, an anonymous referee, and certain members of the audience who attended the presentation of an earlier version of this paper at the Central Division meeting, held in April, 1992, of the American Philosophical Association.