TOWARDS A RULE-UTILITARIAN THEORY OF SUPEREROGATION

Craig K. Ihara

August 18, 1980
ABSTRACT

This article seeks to make a contribution toward the eventual construction of an adequate formulation of rule-utilitarianism by explaining some of the difficulties the notion of supererogation poses for such an enterprise, and by describing a rule-utilitarian theory of supererogation which would resolve those difficulties. The first difficulty that the notion of supererogation raises for rule-utilitarianism is that beliefs concerning acts "beyond the call of duty" are not an insignificant part of many persons' considered moral opinions and any theory, rule-utilitarian or not, which does not provide a place for them will be found lacking on that account. Secondly, I shall argue that without an adequate theory of supererogation it is extremely unlikely that the rule-utilitarian will succeed in arriving at a moral code which will maximize utility. Thirdly, without an adequate theory of supererogation, rule-utilitarian accounts of basic moral concepts such as "right", "obligatory", and "wrong" will be unsatisfactory, even from a utilitarian point of view. I begin in section I by considering how best to define 'supererogation.' I propose a reforming definition at the end of that section. In section II I claim that rule-utilitarians must recognize a distinction between two types of rules if they hope to avoid the difficulties mentioned above. In section III I attempt to explain and to resolve the difficulties supererogation poses for rule-utilitarianism.
Towards a Rule-Utilitarian Theory of Supererogation

This article seeks to make a contribution toward the eventual construction of an adequate formulation of rule-utilitarianism by explaining some of the difficulties the notion of supererogation poses for such an enterprise, and by describing a rule-utilitarian theory of supererogation which would resolve those difficulties. Of course the first difficulty that the notion of supererogation raises for rule-utilitarianism is that beliefs concerning acts "beyond the call of duty" are not an insignificant part of many persons' considered moral opinions and any theory, rule-utilitarian or not, which does not provide a place for them will be found lacking on that account. Secondly, I shall argue that without an adequate theory of supererogation it is extremely unlikely that the rule-utilitarian will be able to formulate a moral code which will succeed in maximizing utility. And finally without an adequate theory of supererogation rule-utilitarian accounts of basic moral concepts such as "right," "obligatory," "wrong," will be deficient on utilitarian grounds. I shall elaborate on these last two difficulties in section IV, after setting out in sections I and II my proposal for a rule-utilitarian theory of supererogation.

I. Defining "supererogation"

Acts such as donating all of one's wealth to the poor, or dedicating one's life to the alleviation of suffering, or dying in order to preserve the lives of others are considered to be paradigms of supererogation. Such acts are what J. O. Urmson had in mind when he made the claim that traditional ethical theories did not do justice to at least some of the extra ordinary acts of saints and heroes, such acts being neither morally wrong, morally indifferent, nor morally obligatory. Ordinary use of the term will hardly provide any significant clues since the term is one which even the


2 Urmson, "Saints and Heroes". p 438.
well educated will not recognize, much less use, in daily conversation. But although the term "supererogation" is relatively unknown, the idea of "exceeding duty" or "doing more than one's duty" constitutes an important part of most people's pre-theoretical moral judgments. For example, few people have difficulty understanding that a condition for receiving the Congressional Medal of Honor is that it be for action which is "above and beyond the call of duty."

However, such common intuitions not withstanding, it seems to me that even a complete analysis of such pre-theoretical judgments will not provide the kind of definition which would be desirable from the point of view of ethics for at least two reasons. For one, our intuitions on this are at best quite vague and ambiguous. For example, paradigms of supererogation involve extraordinary self-sacrifice. When does self-sacrifice become extraordinary? What counts as self-sacrifice? What of self-sacrifice of a less exalted sort? It is also thought that supererogatory acts are highly praiseworthy. But this surely does not distinguish supererogatory acts from the performance of obligatory acts, for clearly the performance of one's duty can also be extremely praiseworthy. Furthermore supererogatory acts are thought to be "above and beyond duty" or "more than one's duty", but in which respect or respects are such acts "more" or "beyond" duty? My own conclusions on this latter question is that there is no helpful or informative way to spell out what could be meant by "more" or "beyond".

Secondly ethics throughout most of its history has been primarily concerned with articulating principles which would enable us to decide what acts are obligatory, wrong, or merely permissible. Anyone convinced by Urmson that such a classification is inadequate and that room must be made for acts of saints or heroes will also be concerned to make room for less exalted acts of generosity, kindness, forgiveness, etcetera which are also overlooked in more traditional ethical theories since they too are at least frequently neither obligatory, wrong, nor merely morally permissible. The fact that such acts are more common and mundane than those performed by saints and heroes does not make them unimportant from the point of view of the aims of traditional ethics. Identifying such acts is crucial to any ethical theory which purports to give us direction as to what to choose. After all whether helping someone in a particular case is one's duty or more than one's duty is important information in making a decision for it sets in an important way the parameters that need to be taken into account. Whether one calls minor acts of kindness or generosity supererogatory is relatively unimportant from this point of view. What is important is to make room for such acts by recognizing a fourth
classification of moral acts which is compatible with the categories of obligatory, wrong, and merely permissable (i.e. morally indifferent or neutral).  

In defining "supererogation" I propose to take this concern of recognizing a fourth category of moral acts as primary. My objective, therefore, is not to arrive at the best reportive definition possible based on our ordinary moral intuitions, but rather to capture that part of the notion which, I think, is most important for traditional ethical theory. I shall continue to use the term "supererogation" in what follows, but if this offends I would not be unhappy to use another term.

In the light of the above I propose the following as a reforming definition of "supererogation":

Definition 1: an act x is supererogatory if and only if x is an act that is not morally obligatory, but is morally desirable.

An equivalent formulation would be the following:

Definition 2: an act x is supererogatory if and only if x is an act that is morally permissible to do or not to do, but is morally desirable.

Both these definitions are intended to exclude praise-worthiness as a necessary condition for being a supererogatory act. The distinction being invoked is the old one made.

---


5 This definition comes close to Rawls' definition of a "good act" in *A Theory of Justice*. (Cambridge, Mass: Harvard University Press, 1971). 438. "We can now define a good act (in the sense of a beneficent act) as one which we are at liberty to do or not to do, that is, no requirements of natural duty or obligation constrain us either to do or not to do it, and which advances and is intended to advance another's good (his rational life plan.)" This definition is unfortunate since it logically implies that obligatory acts are necessarily not good acts and that no good act is obligatory.
by William David Ross between an act and an action. The definitions given above are intended to define a supererogatory act, not a supererogatory action, where the latter would be in part a description of the agent's praiseworthy motivation, and the former being solely a description of the "thing done, the initiation of change." Supererogatory acts in the sense stipulated above would include generous, kindly, and forgiving acts in the sense in which we sometimes speak of the kind or generous thing to do when advising someone as to a choice between alternatives. Certainly from the point of view of arriving at a satisfactory set of substantive moral principles which can guide our behavior the notion of a supererogatory act is of great importance, we want and need to know what good deeds are and are not morally required of us, although the notion of a supererogatory action is also worthy of serious investigation.

These proposed definitions also omit any explicit reference to the notion of self-sacrifice, either of a substantial or of a trivial sort. One reason for this is that, aside from its being evidence of strong motivations, reference to self-sacrifice, or cost, is relevant, given our ordinary moral intuitions, to judgments of what is and what is not obligatory. For example, most people would judge that very great self-sacrifice is usually not morally required. But what this shows is that self-sacrifice, cost, and the closely related notion of "reasonable self-interest" should already be taken into consideration in any plausible account of obligation. If they are, they need not be included again in definitions of what is supererogatory. Their being "not obligatory" will already include such considerations.

Furthermore it is worth noting that although cost and self-sacrifice are plausibly included in any formulation of a theory based on our ordinary moral intuitions, there are many ethical theories, actual and possible, in which cost and self-sacrifice are not relevant factors, for example, in any ethical theory which does not allow any circumstances to override or nullify one's duties. And yet such a theory might still include acts of extraordinary virtue in those cases where what one does, though extraordinary, is not demanded by any obligation. What this shows is that including notions such as "self-sacrifice" in the definition of "supererogation" already injects a substantive element of some specific ethical theories, rather than remaining on the meta-theoretical level.

6 W.D. Ross, The Right and the Good (London: Oxford University Press, 1930). p 7. "I would further suggest that additional clearness would be gained if we used 'act' of the thing done, the initiation of change, and 'action' of the doing of it, the initiation of change, from a certain motive."
As I have suggested these definitions are intended to be quite general, and neutral between different sorts of ethical theories. How then is one to know when a specific act is or is not supererogatory? How is one to know when an act is not morally obligatory, or is morally desirable? These are the kinds of questions specific, substantive moral theories must tell us. So, for example, some theories will tell us that one is not obligated to do x when the cost or self-sacrifice exceeds a certain level, or that one's act is supererogatory if it entails a certain amount of cost. In what follows I will consider how a rule-utilitarian ought to deal with such questions.

II. A Rule-Utilitarian Theory of Supererogation

For a rule-utilitarian (RU) a plausible definition of supererogation, or at least a first approximation of one, would be something like the following:

An act x is supererogatory if and only if it maximizes utility to encourage but not to require or prohibit acts of type x by means of (internal or external) sanctions.

This definition is meant to include not only acts of great utility, but also acts of relatively minor utility. Moderate donations of time and money, for example, will be included. This is in contrast with our everyday usage in which only rare and extraordinary acts are considered supererogation. I depart from common use here because from a RU point of view any type of act that is productive to include in the moral code is of some importance. If we opted for retaining "supererogatory" only for those acts which were exceptional, we would have to invent another term which would include both the exceptional and the not so exceptional acts which are useful to encourage but not to require.

This is not to say that acts which produce trivial increases in utility will also be included in the RU moral code. It is probably that at least most will not be included for two reasons. For one it is doubtful that the good achieved in such cases would outweigh the costs of teaching behavior of that sort. This is not to suggest that there are large disutilities involved in teaching moral principles, nor that teaching is a joyless or onerous endeavor. However learning involves some effort both on the part of the teacher and on the part of the pupil, and a utilitarian must consider whether what is being taught is worth that effort, or whether it might not be of greater utility to teach something else. For example, it seems unlikely that it would be worth the effort...
to encourage people to be attentive to small babies, even though some small contribution might thereby be made to the happiness and development of infants, because most people are already naturally inclined to enjoy and attend to them. A further reason why trivial increases in utility will not count as supererogatory is that such trivial increases would probably not outweigh the costs of having an overly complex system of rules. A rule-utilitarian, in other words, has a utilitarian reason for wanting to keep his moral code simple enough to be managable.

It is also worth noting that the utilitarian definition does not imply that maximizing one's own utility is supererogatory, since here, too, people already have, or tend to have every strong tendencies in the direction of self-interest. Given such strong, natural tendencies it isn't necessary to provide additional moral encouragement in the direction of promoting one's own utility. Of course there are apparent exceptions to this, people who seem to have a very difficult time thinking or doing things for themselves. But even if such people were not in the minority, for the most part all that even they would need is a sense that it is all right to think of their own welfare and their own desires and to put themselves first at least some of the time. In other words reducing their sense of guilt, their sense of what they have a moral obligation to do, is all that is probably necessary for them to act in self-regarding ways that are desirable from a utilitarian point of view.

On this definition of supererogation it is probable that the paradigm case of supererogation, great self-sacrifice, if done for a greater good, will be supererogatory. It also conforms with our intuitions that there are many acts which are good or desirable to do but which are not our duty to do, for example, to forgive others who have wronged you. In other words the definition captures a significant part of our ordinary intuitions about such matters.

On the other hand a number of acts which someone or other might have thought of as "more than one's duty" will clearly not count as supererogatory on this definition. For example, well-intentioned but harmful acts of self-sacrifice, such as devoting one's life to one's child to such a degree that the child never becomes emotionally mature or independent. Similarly other acts of self-sacrifice in which the sacrifice made is not outweighed by the good achieved, will not count as supererogatory, for example, laying down one's life to provide someone with a trivial pleasure or to save some small animal from minor injury. Some acts of this sort may be quite admirable in that they reveal a character of extraordinary
compassion, selflessness, et cetera, yet they are hardly acts which would be useful to encourage, and hence not supererogatory in the utilitarian sense.

One might ask, if supererogatory acts are morally desirable and if it is worthwhile to encourage such acts, why not make all such acts mandatory? After all requiring desirable behavior is bound to increase compliance and therefore increase utility in this regard. But this is a one-sided view of the situation which neglects the disutilities of requiring certain sorts of behavior. First of all it might be that if the act-types demanded are so difficult (e.g. acts of saints and heroes) such that people in general are unable to comply, they may overgeneralize and become discouraged with morality as a whole. This, in turn, could well lead to a diminution in one's motivation to keep other, more realistic, principles of obligation. If so, incidence of the desired behavior may not go up, it may even go down, and incidence of other types of desired behavior may go down with it (e.g. results that were approximately achieved during the Prohibition). Furthermore the negative reinforcement which must be applied in cases of violations are themselves of negative utility. Of course penalties will vary in degree, but it is possible that the level necessary to effect conformity will in and of itself outweigh the utility achieved. Even more serious are the psychological costs. Negative reinforcement restricts freedom and therefore may well lead to frustration. The more demanding the act, the greater the potential frustration, for example if total selflessness were taught as a moral requirement. There are also costs of anxiety and of deliberation that arise out of fear of failing to do what is morally required. Finally the feelings of guilt and self-recrimination that arise once one has violated a moral rule must also be taken into account. These disutilities may outweigh the utilities of requiring behavior which is desirable in itself. In any case they must be taken into account before one can conclude that making a certain type of desirable act-type morally required will maximize utility.

III. Two Kinds of Moral Principles

The RU principle of supererogation depends on the crucial distinction between encouraging and requiring act-types. Of course, encouraging and requiring are not incompatible activities. We are usually both encouraged and pressured into during what we should when we are young. However, the latter is more clearly characteristic of duty, for even in our ordinary talk about duty and obligation we speak of them as things we must do, that we are bound to do and which are binding, and necessary. Duties and obligations are, in our ordinary
moral sense, required. For a rule-utilitarian this sense that certain acts are required or binding is instilled in us through our early education in which penalties, eg. parental punishment or disapproval, are recognized as the price of disobedience. Furthermore, for a rule-utilitarian, principles that specify our duties are understood as sanctioning negative reinforcements of various sorts, from physical punishment to moral condemnation, for violations of those duties. Violation of these principles, then, makes one liable to undesirable reactions on the part of others. In the spirit of Mill's words, "Duty is a thing which may be exacted from a person, as one exacts a debt." Principles of duty and obligation are, then, what we can call Negatively Reinforcing Principles (NRP); that is, they are principles which require certain sorts of behavior by means of negative reinforcements including moral censure. Such principles not only inform one of one's duties and obligations, they also make it clear that violations are morally wrong and warrant negative responses.

On the other hand since acts of supererogation are characterized as acts which are not required, principles that specify acts of supererogation must be principles of a different sort. They specify which acts are supererogatory but, as principles of supererogation, they merely commend such behavior and do not require it. That is to say, not doing what is supererogatory does not make one liable to negative reinforcement; whereas performance provides a defeasible claim to positive reinforcement. Such principles I shall call

7 John Stuart Mill, Utilitarianism. Chapter 5

8 I understand the difference between a "defeasible claim" and a "prima facie claim" in the following way: A prima facie claim is a claim that can be overridden. For example, in the case of competing obligations the weaker obligation is overridden by the stronger. Even when performing the latter it is nonetheless true that one also has a good, though not a sufficient, reason to do the former. In the case of a defeasible claim the claim is not overridden, but nullified. For example, a promise is defeased if it was made under false pretenses. In such cases the fact that one had promised provides no reason for compliance; conditions under which the promise was made make the promise null and void. Performance of a supererogatory and the non-performance of an obligatory one are defeasible claims to praise and blame respectively. They are defeasible because, if certain other conditions are not met - e.g. relevant knowledge, appropriate motivation, et cetera - there are no grounds for either praise or blame. It is not that they have been overridden.
Positively Reinforcing Principles (PRP). PR principles might also be thought of as counsels of morality, e.g. counsels of poverty in the Catholic Church, in contrast with moral commands. PR principles are just as definitive and convey just as much information as NR principles. Epistemologically the two types of principles are not fundamentally different. It is only the force of the principles which distinguishes them.

It is clear that PR principles will have few of the same drawbacks of NR principles. Failure to comply should not reduce respect for the moral law, that is, NR rules of the system. Negative reinforcement will be either non-existent or minimal where PR principles are concerned. Frustration at having one's freedom curtailed and the anxiety and guilt which attends violation of one's duty will not occur when one does not do what is supererogatory as long as the individual perceives the situation correctly; that is, he does not mistakenly believe that he has a duty to do what is, in fact, supererogatory. There will, of course, be some costs involved in teaching PR principles, but these would also be incurred teaching NR principles, and will probably not be as great since in the case of NR principles we want to insure compliance in a much higher percentage of cases.

An individual who has internalized an RU system of morality which recognizes such a distinction between PR and NR principles will have certain characteristic attitudes toward compliance and non-compliance with those principles. Such an individual will tend to think badly of one who fails, without excuse or justification, to do one's duty. He will feel satisfaction and some degree of approval toward one who acts in accordance with duty. He will also tend to feel that failure to comply tends to justify some kind of public censure and perhaps punishment.

On the other hand non-compliance with PR principles will not entail any moral disapprobation from our hypothetical individual. This is not to deny that it may give rise to feelings of disappointment, depending on his expectations of the person involved. For example, if the person leads him to expect especially generous, or kindly behavior from him/her, say through his boastful self-characterizations, and does not live up to them, disappointment would be appropriate; as well as some criticism for not living up to one's own boasts. Or one might be disappointed if one's child fails to live up to the PR standards given the ideals one has tried to instill. Such disappointment or criticism which arise out of special

circumstances do not provide counterexamples to the general claim that non-compliance with PR principles does not by itself entail moral criticism.

Perceptions of others consciously acting in accord with PR principles will tend to elicit feelings of approval. In some cases of great sacrifice approbative moral feelings of the very strongest sort will be felt. Individuals who consistently comply with principles of supererogation, or comply in a few extraordinary instances, will be looked up to and admired, respected, and in the extreme case, revered. Moral heroes, for example great religious teachers, fall into this category. Such persons provide models of conduct which provide much of the motive force behind people's compliance with principles of supererogation. We want to be like Jesus, Gandhi, or Schweitzer and thus perform acts which imitate theirs to some degree.

With regard to himself the person taught this type of RU moral code will have characteristically a strong motivation against non-compliance with NR principles on each and every occasion in which compliance comes into question. Acting in accord with NR principles will characteristically bring him some satisfaction, or perhaps simply maintain a sense of being at ease with himself. It is worth noting that even though NR principles derive their forcefulness primarily, or essentially from the negative reinforcements which attach to them, that a moral individual in this RU sense may not feel pressured or coerced into doing his duty any more than a law abiding citizen is conscious of, or feels coerced by the threat of legal penalties when he obeys the law. The cases are also parallel in that if either individual fails to comply or is tempted not to comply they will most certainly be aware that there is a requirement, moral or legal, that he do his moral or legal duty. If the person who has internalized the RU system actually fails to comply with NR principles he will feel a discomfort appropriately labeled guilt since it is a discomfort which does not derive immediately from an external source, but simply from recognition of one's unjustified violation of one's obligation.

The attitudes of our individual with regard to PR principles are less easy to characterize. He/she will undoubtedly have some motivation to comply with these principles, and for their own sake, because they promote ends which one has internalized as intrinsically good, e.g. promotion of happiness. When one does comply one will generally have feelings of self-respect and self-worth, as well as feelings of satisfaction and self-approval. However he/she will not feel compelled to comply with any PR principles. Indeed if he/she ever feels compelled to comply with PR principles it will be
because he sees compliance as part of his ego-ideal. That is, an individual may set standards for himself that he would not perceive as required of others. If especially beneficent or self-sacrificing behavior is perceived as part of this self-image the phenomenology of one's situation may seem much the same as that of one faced with a duty. Depending, then, on one's self-image, non-compliance with PR principles will entail potentially strong feelings of disappointment in oneself, which may closely resemble the feelings of guilt one feels for violating NR principles. In such situations one will not feel at ease with oneself. For example, one who perceives himself as being very generous will feel disappointed and uncomfortable with himself if he discovers that he does not act as a generous man would act and these feelings will have a very "moral" quality similar to guilt because he has failed to live up to his own standards.

On the other hand it is possible that someone might never feel compelled to comply with PR principles. This would be the case if no PR principles were incorporated into one's ego-ideal. This is not to say that such an individual would never comply with PR principles, only that he would never feel morally bound to do so. Indeed, as we have said, if the moral system is properly internalized, one will feel motivated to comply at least on some occasions - for one sees that PR principles advance ends one holds to be valuable. For some individuals, then, compliance with PR principles will be done solely from a desire to perform acts of that sort - for example, acts of helping others - and without any feeling of "moral" requirement. Non-compliance for such persons will not involve any emotional discomfort or recriminations of the sort encountered by those who see such behavior as part of their ego-ideal, or by those who violate their moral obligations.

IV. Some Problems and Suggested Solutions

Having sketched the foundations of a rule-utilitarian theory of supererogation, we are now in a position to see some very serious defects in RU which arise out of neglect of the notion of supererogation.

First of all in formulating a RU theory it is extremely important for the rule-utilitarian to recognize that there are alternatives to prohibiting undesirable behavior and requiring desirable behavior. This is because it is quite possible that on occasion greater overall utility will be achieved if one of the alternatives is adopted. One alternative, of
course, is to say nothing about a given sort of conduct, even if it is in itself undesirable. Saying nothing may have overall greater utility just because the only effective prohibition would have so much disutility that more harm than good would result from adopting the prohibition. Another virtue of this tack is that, in addition to completely avoiding the disutilities of negative reinforcement, one also avoids the costs of inculcation. Especially in cases where there already are strong, non-moral tendencies to perform the desired behavior it will probably be the case that having no rule would maximize utility.

Another alternative to negative reinforcement has been discussed on the foregoing pages, that is, to recommend or encourage, but not to require desirable sorts of conduct by means of PR principles. Many rule-utilitarians tend to overlook or underemphasize this possibility when determining the utility maximizing set of moral principles. From the point of view of these rule-utilitarians it seems as if acts must be either forbidden, required, or ignored by the moral code. Put in my terms, acts are, by these philosophers, either to be subsumed under some NR principle or left out of the moral code altogether. A clear example of this comes from Richard B. Brandt's article, "Toward a Credible Form of Utilitarianism":

An act is right if and only if it conforms with that learnable set of rules the recognition of which as morally binding roughly at the time of the act - by everyone in the society of the agent would maximize intrinsic value. (Emphasis is mine)

In such a system it is only the morally binding rules, that is the NR principles, which are relevant. But in this Brandt's view is not atypical of rule-utilitarian theories.

However, if rule-utilitarians think this way they will have a serious problem, because the utility-maximizing set of NR


12 Brandt, in his recent book The Good and the Right (Oxford: Clarendon Press, 1979), mentions and briefly describes a view very similar to the one I propose.
principles will probably not be the set of principles which will maximize utility. This is because in some circumstances it will maximize utility to encourage, but not to require or ignore a certain sort of behavior. In such cases the utility maximizing set of exclusively NR principles will not produce as much utility as the same set with the behavior recommended and not required. Simply put, if the rule-utilitarian restricts his choice to sets of NR principles he will overlook some set of mixed NR and PR principles which will in fact be the set which will maximize utility. The rule-utilitarian can remedy this problem simply by broadening the scope of his investigations to include sets which include PR principles.

The other problem which arises from the rule-utilitarian's tendency to overlook acts of supererogation, and their corresponding principles, lies in his account of basic moral concepts. Actually this difficulty is multi-faceted. First of all consider Brandt's definition of "right act." Since it is defined by reference to the set of NR principles the definition makes it logically impossible for a supererogatory act to be a right act!

Secondly, such a definition of "right act" leads to the following problem: suppose that we have two sets of NR principles, A and B, where set A consists of set B plus NR principles, c. Suppose that the overall utility of requiring c (regarding c as a NR principle) is less than the overall utility of recommending but not requiring c. This is not improbable since there is, as earlier described, an inherent disutility in making conduct morally required. It might still be the case that the total utility of A, which includes a NR version of c, is greater than the utility of B which has no c principle. This, according to the rule-utilitarian definition Brandt gives, implies that only acts conforming to set A are right. However this is not satisfactory on utilitarian grounds for the set of NR principles, B, plus the PR principle c (call this set B') would probably have greater overall utility than A. For example suppose c is the principle: Forgive others when they do you or those you love an injustice. Now it may be that making this a moral requirement will have greater utility than omitting any mention of forgiveness from the moral code. However given the difficulty, perhaps impossibility, of getting most people to comply with such a principle in many cases, and the guilt that would accompany

We cannot simply infer that B + c will have greater utility than A, since the utility of a system of rules is not simply the sum of the utilities of each rule in it. However, it is very likely that B" will have greater utility than A in most cases.
violations of it, it is quite likely that regarding c in this case as a PR principle, rather than a NR principle would maximize utility.

A parallel problem occurs in typical rule-utilitarian definitions of "wrong act." If a "wrong act" is one that fails to conform, or which violates, the utility maximizing set of NR principles, then an act of not conforming to principle c in the example above would be wrong since it fails to conform with A. But surely rule-utilitarians would want to use B' as the basis of the definition of "wrong act" and not set A. For example, suppose that set B' includes a principle which recommends but does not require acts of great self-sacrifice, and that B' is more useful than set A which is identical to B' except that in A self-sacrifice is required. Not sacrificing oneself in a given situation and thus not complying with the principle would be allowable in B' but not in A, and thus would not be wrong in the former but would be wrong in the latter. The latter consequence, given that B' maximizes utility, should be unacceptable to a rule-utilitarian.

Rule-utilitarian difficulties with definitions of "right" and "wrong" can be resolved simply by unambiguously eliminating any explicit or implicit restrictions against using PR principles. For example, the following definitions of "right" and "wrong" avoid the problems referred to above:

An act x is right if and only if it is (1) enjoined by NR principles or (2) recommended by PR principles of a utility maximizing set of NR and PR principles.14

An act x is wrong if and only if x violates the NR principles of a utility maximizing set of NR and PR principles.

Along the same lines we can define an obligatory act in the following manner:

An act x is obligatory if and only if it is required by the NR principles of a utility maximizing set of NR and PR principles.

Finally we can provide a more precise rule-utilitarian account of supererogation than the one given earlier:

14 Or, if all one means by "right act" is that it is morally permissible, we could adopt this definition: An act is right if and only if it is not prohibited by the NR principles of a utility maximizing set of NR and PR principles.
An act is supererogatory if it is recommended by the PR principles, and is neither prohibited nor required by the PR principles, of a utility maximizing set of NR and PR principles.

This definition is different from and preferable to the one suggested previously because it takes into account the utility of the entire set of NR and PR principles and not simply the utility of making a particular act-type recommended but not required. However, the earlier version would be preferable for some sophisticated versions of act-utilitarianism.

What I have attempted to do in this paper is to set out some problems that the notion of supererogation raises for rule-utilitarian theories as they have often been formulated, and also to suggest some solutions to those problems by proposing a definition of supererogation and emphasizing the importance of recognizing two kinds of moral principles. No doubt my suggestions have problems of their own. However, some solutions not unlike the ones I have made must be found if rule-utilitarianism is ever to become a completely plausible ethical theory.

Craig K. Ihara
California State University, Fullerton