

On the Origin of the Utilitarian Maximization Requirement*

1 Introduction

The most important achievement of utilitarian ethics consists in the development of a formal model which allows one to take into account the diverging interests of all individuals, $\{I_1, \dots, I_n\}$, concerned by an action A . Assuming that the degree to which A afflicts the interests of I_j can be “measured” by a numerical function $u_j(A)$ which expresses the *utility* of A for the individual I_j , “classical utilitarianism” holds that the *total value* of A , $U(A) = \sum_{j \leq n} u_j(A)$, is the one and only criterion for the morality of A . Now there are many problems connected with this basic thesis; to mention only the most important points of critique:

- (1) The very existence of the *function* $u_j(A)$ which determines the utility of A for I_j in a numerically precise way, say as a rational number r measured at least on an interval scale, may appear quite doubtful.
- (2) It is even more doubtful whether the different functions $u_j(A)$ ($1 \leq j \leq n$) might ever be constructed in an *intersubjectively comparable* way so that, for each action A and for all individuals I_i, I_j : $u_i(A) > u_j(A)$ if and only if A is better (or more useful) for I_i than for I_j .
- (3) It is not always clear “who counts” in utilitarian considerations, i.e. who exactly must be subsumed into the group of all individuals concerned by action A . One specific question concerns the issue whether only the interests of humans or also those of other *non-human beings* have to be taken into account. Another question concerns the status of “*potential people*”, i.e. unborn children, fetuses, embryos, etc.¹ Yet another problem is raised by the issue if (or how far) we should take care of the interests of *future generations*.
- (4) Furthermore, it is more than a merely technical detail whether the total value of an action A should be calculated by the simple formula $\sum_{j \leq n} u_j(A)$, or whether one should perhaps introduce certain weighing factors w_j ($0 \leq w_j \leq 1$) which reflect the possibility that, in the context of action A , the interests of a certain individual I_i , may “count less” than the interests of another individual I_j . While “classical” utilitarianism is characterized – in accordance with Bentham’s maxim “everybody to count for one,

* I would like to thank Rainer Trapp for intensive and helpful discussions of a former version of this paper.

¹ Cf., e.g., the Symposium on “Possible People” in FEHIGE/WESSELS [1998].

and nobody to count for more than one”² – by the special case $w_1=w_2=\dots=w_n=1$, modern authors have offered a wide variety of non-egalitarian forms of utilitarianism. These, however, shall not be considered here.³

- (5) Finally we will here also ignore the question whether the moral admissibility of an action A is determined *only* by the total utility of A, or whether some non-utilitarian factors such as fundamental rights, subjective intentions, etc. should be taken into account as well.

Thus I take it for granted that in a certain way (at least up to further revisions – see section 6 below) the total value of A, $U(A)$, provides an acceptable basis for the following utilitarian criterion of morality:

(BETTER) Action A is morally better than (or morally superior to) action A' if and only if (or, for short: “iff”) $U(A) > U(A')$.

The main aim of this paper is to investigate in some detail in which way this *comparative* criterion entails (or fails to entail) certain *classificatory* criteria as to when some action A is *morally right* (or wrong) and as to when A is morally *permitted* (or forbidden). The standard view of utilitarian ethics appears to assume, or even to presuppose as self-evident, that an action A is morally *right* if and only if A is the *best* among all available alternatives. Somewhat more exactly:

(RIGHT) Action A is morally *right* iff there is no *better* alternative, i.e. iff there does not exist an alternative A' such that $U(A') > U(A)$.

Unfortunately, the notions or concepts of *right* and *wrong* appear to be rather vague. So I would be willing to admit that there exists a meaningful interpretation of the word ‘right’ according to which, in a situation of choice among alternatives which can be ordered on a comparative scale of ‘worse’ or ‘better’, it is always “right” to chose the best (and “wrong” not to chose the best). Therefore I will formulate my critique not primarily as an attack against principle RIGHT. Rather I want to argue that utilitarian ethics becomes unacceptable as soon as it subscribes to the following criterion

(PERMITTED) Action A is morally *permitted* iff there is no alternative A' such that $U(A') > U(A)$.

Of course, the first “half” of this equivalence stating that the optimality of action A is *sufficient* for A's being morally permitted, seems to be beyond any reasonable doubt.

² Quoted according to SIDGWICK [1847: 417].

³ Cf., e.g., the “Theory of Justice” in RAWLS [1972]; and especially certain variants of a “Gerechtigkeits-Utilitarismus” in TRAPP [1988].

However, it is the converse assumption according to which an action A is morally permitted *only if* there is no A' such that $U(A') > U(A)$, or, in other words, principle:

(FORBIDDEN) Action A is morally *forbidden* if there exists an alternative action A' such that $U(A') > U(A)$

which turns out to be unacceptable. In the subsequent sections 2 and 3 the historical roots of the Utilitarian (mis)conception of what is morally permitted will be investigated. In sections 4 and 5 the key arguments in favor of this conception will be discussed and it will be argued that, at least when applied to actions carried out by *some particular responsible individual* I_j, criterion FORBIDDEN should be replaced by a much more liberal criterion of moral permissibility. In the concluding section 6, the implications of this issue for the general problem of supererogation will be investigated.

2 *The Historical Origin of RIGHT and PERMITTED*

One of the earliest statements of the core idea of classical utilitarianism may be found in J. Hutcheson's *Inquiry Into the Original of Our Ideas of Beauty and Virtue* of 1725. In section 3, § 8 Hutcheson declares that action to be the "best which provides the greatest happiness to the greatest number, and the worst is that which produces a corresponding misery". This characterization immediately entails the adoption of the comparative criterion BETTER, as it is also put forward at some more length in Jeremy Bentham's *Introduction to the Principles of Morals and Legislation* of 1789. The "principle of utility" is defined there as:

"2. [...] that principle which approves or disapproves of every action whatsoever, according to the tendency which it appears to have to augment or diminish the happiness of the party whose interest is in question: or, what is the same thing in other words, to promote or to oppose that happiness. I say of every action whatsoever; and therefore not only of every action of a private individual, but of every measure of government." [BENTHAM 1789: 12]

Two points must be emphasized here. First, Bentham is clearly advocating a *comparative* (or even quasi-metrical) notion or criterion of moral acceptability when he says that actions in general are to be judged as morally good *to the degree* in which they tend to augment the happiness of the entire group. Second, the very use of the words 'augment' and 'diminish' indicates that Bentham probably has the following construction in mind. Start by considering the situation or "state of happiness", say H_1 , of a certain group G *before* action A is carried out. Next imagine the "state of happiness", H_2 , as it would be *after* action A were actually carried out. Now *if* the latter value H_2 is greater than H_1 , then A is likely to augment the happiness of G and hence A is morally good, while, if $H_2 < H_1$, then A has the tendency to diminish the happiness of G and thus A is morally bad.

But how can we exactly determine these values $H_1(A)$ and $H_2(A)$ for arbitrary actions A when, as Bentham mentions elsewhere, one and the same action A may both have a tendency to *promote* the happiness and yet have another tendency to *diminish* the happiness of group G ?⁴ Although Bentham remains silent on this point, it seems plausible to model these diverging tendencies in such a way that A has positive implications for *some* members of G , while at the same time A has negative implications for certain *other* members of that group. If we consider one particular individual I_j , the “state of happiness” of I_j *after* action A is executed, $H_2(I_j)$, should plausibly be equated with the utility of A for I_j , $u_j(A)$. Similarly, the “state of happiness” of I_j *before* A is carried out, $H_1(I_j)$, may be expressed by the value $u_j(\sim A)$ where the negation $\sim A$ is meant to symbolize that action A has not (yet) been executed.⁵ Hence action A *is positive for* I_j , or augments the happiness of the individual I_j , if and only if $H_2(I_j) > H_1(I_j)$, i.e. $u_j(A) > u_j(\sim A)$; and A is *negative for* the individual I_j , or diminishes the happiness of I_j , iff if $H_2(I_j) < H_1(I_j)$, i.e. $u_j(A) < u_j(\sim A)$.

Generalizing this idea from one individual to the entire group G , the value $H_2(A)$ must be equated – in accordance with the standard conception of the total utility of A – with $\sum_{j \leq n} u_j(A)$, i.e. $U(A)$, while $H_1(A)$ corresponds to $\sum_{j \leq n} u_j(\sim A)$, or $U(\sim A)$, for short. In sum, then, action A may be said to be *positive for the entire group* G , or augments the happiness of G , if and only if $U(A) > U(\sim A)$, and A is *negative for the entire group* G , or diminishes the happiness of G , if and only if $U(A) < U(\sim A)$. As is evidenced by the earlier quotations, Bentham not only subscribes to the comparative criterion BETTER, but he also holds the view that each positive action (which has the overall tendency to augment the happiness of G) is *ipso facto* morally *acceptable*. Hence he appears to accept the following liberal criterion of utilitarian permissibility which, unlike PERMITTED, does not require us to do always the (morally) best of all available alternatives:

(PERMITTED_{LIB}) Action A is morally *permitted* iff A has the tendency to augment the happiness of the group, i.e. iff $U(A) > U(\sim A)$.

⁴ Cf. BENTHAM [1789: 12/13]: “6. An action [...] may be said to be conformable to the principle of utility, or, for shortness sake, to utility, (meaning with respect to the community at large) when the tendency it has to augment the happiness of the community is greater than any it has to diminish it.”

⁵ The “negative” action $\sim A$ may alternatively be regarded as the “null-action” which leaves the world entirely as it is, or as it will develop by its own. Since $U(\sim A)$ serves as a threshold for moral permissibility, this conception presupposes that “doing nothing”, $\sim A$, is morally *neutral*, i.e., neither good nor bad. As Rainer Trapp pointed out to me, however, such an assumption may not be warranted in specific situations where omitting to do A entails that people will suffer (great) harm which might easily be prevented by me. Since I admit that such omissions are morally bad, the subsequent considerations shall be confined to “normal” actions for which $U(\sim A) = 0$.

Note incidentally that this condition in turn presupposes that each individual function u_j is “normal” in the sense that $u_j(A) > 0$ iff A is good (or useful, or positive) for I_j , while $u_j(A) < 0$ iff A is bad (or harmful, or negative) for I_j . A discussion of the latter condition of “normality” is found in LENZEN [1980] and in section 0.4 of LENZEN [1999].

That, for Bentham, the moral acceptability of permissibility of an action A should be judged by A's overall tendency to augment the happiness of G in the sense of $U(A) > U(\sim A)$, is also expressed in the following passage:

“10. Of an action that is conformable to the principle of utility, one may always say either that it is one that ought to be done, or at least that it is not one that ought not to be done. One may say also, that it is right it should be done; at least that it is not wrong it should be done; that it is a right action; at least that it is not a wrong action. When thus interpreted, the words *ought*, and *right* and *wrong*, and others of that stamp, have a meaning; when otherwise, they have none.” [BENTHAM 1789: 13]

Thus instead of the former rigorous principle RIGHT we obtain the following weak principle:

(RIGHT_{LIB}) Action A is morally *right* iff A has the tendency to augment the happiness of the group, i.e. iff $U(A) > U(\sim A)$.

This principle of course may be supplemented (or paraphrased) by saying that action A is morally *wrong* iff A has the tendency to diminish the happiness of G, i.e. iff $U(A) < U(\sim A)$.

Next consider John Stuart Mill's views as expounded in his book *Utilitarianism*. In Ch. II entitled “What Utilitarianism is”, one finds the following explanation:

“The creed which accepts as the foundation of morals “utility” or “the greatest happiness principle” holds that actions are right *in proportion as* they tend to promote happiness; wrong as they tend to produce the reverse of happiness.” [MILL 1861: 7; my italics]

Like Bentham, also Mill subscribes to the comparative criterion BETTER when he maintains that the moral quality of an action A is *proportional* to its tendency to promote the happiness (or unhappiness) of the entire group. Furthermore Mill also speaks in favor of the classificatory criterion RIGHT_{LIB}. Or, to put it somewhat more exactly, since comparative forms of ‘*right*’ and ‘*wrong*’ would be ungrammatical, Mill's summary of “the greatest happiness principle” is better expressed by saying that an action A is *good* in proportion as it tends to promote happiness, and *bad* in proportion as it tends to produce the reverse of happiness:

(GOOD) Action A is morally *good* iff A has the tendency to augment the happiness of the group, i.e. iff $U(A) > U(\sim A)$.

(BAD) Action A is morally *bad* iff A has the tendency to diminish the happiness of the group, i.e. iff $U(A) < U(\sim A)$.

Note that on this view actions do not necessarily have to be either good or bad; they may be (morally) “neutral” in the following sense:

(NEUTRAL) Action A is morally *neutral* iff A does not affect (i.e. neither augments nor diminishes) the happiness of the group, i.e. iff $U(A) = U(\sim A)$.

The third (and last) “classic” author to be considered here is Henry Sidgwick. In Ch.1 of Book IV of *The Methods of Ethics*, he defines “The Meaning of Utilitarianism” as the ethical theory which maintains

“[...] that the conduct which, under any given circumstances, is objectively right, is that which will produce the greatest amount of happiness on the whole; that is, taking into account all whose happiness is afflicted by the conduct” [SIDGWICK 1874: 411].

This represents a rather explicit statement of the rigorous “maximization” principle RIGHT according to which an action A is morally *right* only if there is no *better* alternative, i.e. only if there does not exist an alternative A’ such that $U(A') > U(A)$.⁶ Sidgwick refers to this principle as a “Universalistic Hedonism” which must clearly be distinguished from an “Egoistic Hedonism”:

“The difference [...] between the propositions (1) that each ought to seek his own happiness, and (2) that each ought to seek the happiness of all, is so obvious and glaring ...” [SIDGWICK 1874: 411/2]

In Ch. 2 he offers the following “Proof of Utilitarianism” which rests on a familiar principle of generalizability:

“When, however, the Egoist puts forward, implicitly or explicitly, the proposition that his happiness or pleasure is Good, not only *for him* but from the point of view of the Universe [...] it then becomes relevant to point out to him that *his* happiness cannot be a more important part of Good, taken universally, than the equal happiness of any other person. And thus, starting with his own principle, he may be brought to accept Universal happiness or pleasure as that which is absolutely and without qualification Good or Desirable“ [SIDGWICK 1874: 420/1].

This “proof” does not appear to have any immediate consequences for the main topic of this paper. However, the parallel that Sidgwick here draws between the selfish reasons for egoistic actions on the one hand and the unselfish reasons for moral actions on the other hand may be considered as having prepared the ground for the crucial utilitarian misconception of what is morally permitted. This parallel may be spelled out as follows. Just as an egoist agent has to chose that course of action which best accords with his *own* interests, so any moral agent has to chose that course of action which best accords with the interests of the entire group.⁷ Therefore it seems to follow that just as an egoist who does not maximize his own interests acts *irrational*, so a person who does not maximize the interests of the entire group acts

⁶ Towards the end of Ch. 1 SIDGWICK [1874: 416/7] discusses the problem which action should be chosen according to the “Utilitarian criterion of right conduct” in case that there exist several distinct actions “distributing the same [maximal] quantum of happiness among the same number of persons”. This issue shall not be dealt with here.

⁷ One might also add that a “saint”, i.e. a morally perfect altruist has to chose that course of action which best accords with the interests of all others, ignoring his own interests altogether.

immoral. As will be shown below, several 20th century philosophers have explicitly subscribed to this (or quite similar a) conclusion.

3 Some 20th Century's Proponents of RIGHT and PERMITTED

The following set of quotations is not intended as a complete survey of 20th century's textbooks on Utilitarianism. It only aims to show how the requirement of *maximizing* the total utility of actions gradually evolves as a basic tenet of Utilitarianism which eventually is not regarded as standing in need of further justification. Thus, in his 1965 book on the *Forms and Limits of Utilitarianism*, after having distinguished so-called rule-utilitarianism from act-utilitarianism, D. Lyons puts forward the following version of the maximization criterion RIGHT:

“An act is right if, and only if, its total effects are no worse than those of any alternative.” [LYONS 1965: 136]

In his introduction to “An outline of a system of utilitarian ethics” (a 1973 reworking of a 1961 monograph with the same title), J.J.C. Smart characterizes act-utilitarianism, roughly speaking, as:

“[...] the view that the rightness or wrongness of an action depends only on the total goodness or badness of its consequences, i.e. on the effect of the action on the welfare of all human beings (or perhaps all sentient beings).” [SMART 1973: 4]

This passage does not yet contain the crucial maximization requirement and it is therefore perhaps best interpreted as an implicit formulation of the comparative criterion BETTER⁸. In section 4, however, SMART [1973: 27] incidentally introduces the idea of maximizing when dealing with the problem “[...] whether we should try to maximize the *average* happiness of human beings [...] or whether we should try to maximize the *total* happiness or goodness.” Furthermore, in section 6 devoted to the “Rightness and wrongness of actions”, Smart vaguely puts forward the idea of maximizing total utility in terms of the principle that “one ought [always] to do that which will produce the best consequences” [o.c., 30]. Finally at the beginning of section 7 he offers the following straightforward formulation of RIGHT:

“According to the act-utilitarianism, then, the rational way to decide what to do is to decide to perform that one of those alternative actions open to us (including the null-action, the doing of nothing) which is likely to maximize the probable happiness or well-being of humanity as a whole, or more accurately, of all sentient beings. [...] The right action is the action among those which we could do [...] which has the best possible results.” [SMART 1973: 43; 45].

⁸ A more explicit version is given on p. 30: “Let us say, then, that the only reason for performing an action *A* rather than an alternative action *B* is that doing *A* will make mankind (or, perhaps, all sentient beings) happier than will doing *B*.”

Let it be noted in passing that the same characterization of “the right action as that which, of the possible alternatives, maximizes the good” was also adopted by Bernard Williams in “A critique of utilitarianism”⁹. Although Williams raises various objections against Smart’s views, he doesn’t, however, criticize the maximization requirement itself.

By the year 1975 when Otfried Höffe edited (in German) his well-known *Introduction to Utilitarian Ethics*, a collection of classical and contemporary texts, the principle of maximization has become one of the core ideas of Utilitarianism. Thus, in Höffe’s preface the principle of utility is summarized as follows:

„Diejenige Handlung bzw. Handlungsregel ist moralisch richtig, deren Folgen für das Wohlergehen aller Betroffenen optimal sind“ [HÖFFE 1975: 10].

I.e., only that action (or rule of action) is morally right the consequences of which are *optimal* for the welfare of all individuals involved. More recently, SHAW [1999: 10] formulated this basic tenet of Utilitarianism briefly as follows:

“[...] an action is right if and only if it brings about at least as much net happiness as any other action the agent could have performed; otherwise it is wrong. [...] We act rightly only when we bring about as much happiness as it is possible for us to bring about”.

Here we have an unambiguous statement of principle RIGHT, or of its negative counterpart, according to which one acts morally *wrong* if and only if one does not perform the best of all possible actions:

(WRONG) Action A is morally *wrong* iff there is some better alternative, i.e. iff there exists at least one alternative A’ such that $U(A') > U(A)$.

Let us now see which considerations might speak in favor of this principle.

4 Rationality and Morality

I do not claim to have full knowledge of, or even to present an halfway exhausting exposition of, the whole literature of this subject, but it seems to me that there basically exists only *one* idea in support of the crucial principles RIGHT, WRONG, and PERMITTED. This prima facie quite plausible argument draws a parallel between acting *rationally* on the one hand and acting *morally* on the other hand. It may be summarized as follows:

(RATIONAL) If of two alternative actions A₁ and A₂ the former is *better* (for agent X) than the latter, then, if X should decide to do A₂ rather than A₁, he would act irrationally.

⁹ Cf. WILLIAMS [1973: 85/6].

Similarly: If of two alternative actions A_1 and A_2 the former is *morally* better than the latter, then, if X should decide to do A_2 rather than A_1 , he would act immorally. Or, more generally:

(MORAL) If there are n alternative actions $\{A_1, \dots, A_n\}$ such that A_1 is morally better than A_2 and ... and A_{n-1} is morally better than A_n , then, if X should decide to do any of the A_j ($j \neq 1$) rather than A_1 , he would act immorally.

This idea is already anticipated in Bentham's *Introduction to the Principles of Morals and Legislation* where the decisive criterion for the utility of an action A is viewed either in the happiness of an *individual* X, or in the happiness of the *community*:

“3. By utility is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness [...] to the party whose interest is considered: if that party be the community in general, then the happiness of the community: if a particular individual, then the happiness of that individual.

4. [...] The community is a fictitious *body*, composed of the individual persons who are considered as constituting as it were its *members*. The interest of the community then is, what? – the sum of the interests of the several members who compose it” [BENTHAM 1789: 12].

From this point of view it appears to follow that just as in the case where person X has to take only her own interests into account, X acts “rightly”, i.e. rationally, if and only if she maximizes her own happiness, so in case she has to take the interests of the whole community into account, X acts “rightly”, i.e. morally, if and only if she maximizes the happiness of the community. This view is spelled out at greater length in HARSANYI [1982: 42-44] who argues:

“[...] that the emergence of modern decision theory has made ethics into an organic part of the general theory of rational behaviour. [...] It is the task of the normative disciplines of decision theory, game theory, and ethics to help people to act more rationally and to give them a better understanding of what rationality really is. [...]

(1) The theory of *individual* rational behaviour, which itself comprises the theories of rational behaviour

(1A) Under certainty

(1B) Under risk [...]

(2) Game theory, which is a theory of rational interaction between two or more individuals, each of them rationally pursuing his own objectives [...]

(3) Ethics, which is a theory of rational behaviour in the service of the common interests of society, as a whole. [...]

In case (1A), the secondary definition of rationality is *utility maximisation* [...] In cases (1B) and (1C) the secondary definition of rationality is *expected-utility maximisation* [...] Finally, in the case of ethics (case (3)), as we will see, the secondary definition of rationality (or of morality) is in terms of *maximising the average utility level* of all individuals in the society.”

Assuming, then, that “the society”, i.e. the set of all individuals concerned by action A , may be regarded as a *bearer of one common interest*, Harsanyi maintains that just as in the private sphere it is always rational for individual I_j to maximize his own interests, so in the social sphere it is always moral to maximize society's interests. However, as becomes evident, e.g.,

from ch. 1, § 5 of Rawls' "A Theory of Justice", the crucial extension of the principle of rational decision from one individual, I_j , to a "society" $\{I_1, \dots, I_n\}$ will work only if one conflates "all persons into one through the imaginative acts of the impartial sympathetic spectator" [RAWLS 1972: 27]. But even if this view of the interests of the "society" as uniformly representable in the mind of an impartial spectator is accepted, it still doesn't follow that each I_j has the *moral* duty to maximize the "common interest". It only follows that if the "society" $\{I_1, \dots, I_n\}$ were to decide which of several alternative actions A_1, A_2, \dots, A_k to perform, it would be *rational* to chose that action A_j by which the "common interest" is maximized.

5 BETTER Revisited

The proponents of the above quoted attempts to base morality on rationality apparently did not sufficiently distinguish between two different kinds of situations. In what might be called a social or *political* situation, the decision has to be made by an impersonal *institution* (e.g., a government) which either has no interests of its own or, at any rate, the interests of which are not afflicted by action A. In this case the following parallel drawn by Richard B. Brandt certainly is legitimate:

"If a proposed course of action does not raise moral questions, it is generally regarded as rational, and its agent well-advised to perform it, if and only if it will maximize expectable utility for the agent. In a similar vein, it can be argued that society's "choice" of an institution of morality is rational and well-advised, if and only if it will maximize expected social utility" [BRANDT 1967: 42].

Thus I do not object to classical utilitarianism as long as it holds only that a *society* is "rightly ordered, and therefore just, when its major institutions are arranged so as to achieve the greatest net balance of satisfaction summed over all the individuals belonging to it" [RAWLS 1972: 20].

Quite a different situation obtains, however, when action A is to be carried out by a specific, responsible agent I_j who belongs to the group of individuals $\{I_1, \dots, I_n\}$ concerned by that action. Of course, if – in Brandt's words – A "does not raise moral question", then person I_j acts rationally iff she maximizes her own interests. But in everyday situations moral consideration normally do play a role, and the task of any ethical theory is just to define the morally acceptable equilibrium between the agent's own interests and the interests of the others. An extreme (or "ideal") egoist may be characterized schematically as someone who takes only his own interests into account, while an ideal altruist would totally ignore his own interests. The decisive question now is: *Where exactly lies the morally acceptable borderline*

between egoism and altruism? To which degree are we allowed to satisfy our own interests, and to which degree do we have to take the interests of others into account?

The answer given by classical utilitarianism is indirectly contained in Bentham’s slogan “everybody to count as one”. Agent I_j is allowed to take his own interests into account only to the same degree that I_j is prepared to take the interests of others into account as well. If, by maximizing *total* utility, I_j ’s *own* utility is increased too, so good for him; but if, by maximizing total utility, I_j ’s utility is decreased, “tant pis”! As the subsequent examples make clear, however, these conclusions are entirely unacceptable. On the one hand, the utilitarian maxim would forbid, or morally disqualify, an *altruistic* action A_{alt} if the benefit for the others is lower than the harm that I am willing to accept for myself. On the other hand, an *egoistic* action A_{ego} might be allowed as morally right provided only that the harm done to others is outweighed by my own personal benefit!

For the sake of simplification, let’s consider a schematic situation where besides agent X only one other individual Y is involved, and where the only alternative action consists in not doing A, symbolically $\sim A$. Now in a first scenario, assume that a (relatively poor) mother, M, plans to give a certain amount of money to her (relatively rich) son S. Let the payoffs of A and $\sim A$ be determined by the following matrix:¹⁰

	M	S
A	-10	5
$\sim A$	0	0

In this situation the total utility of A equals -5 and is hence smaller than the total utility of $\sim A$, 0. Therefore the utilitarian criterion BETTER disqualifies A as *morally worse* than $\sim A$! In the same vein the idea of maximization as expressed, e.g., in principle RIGHT says that M has to omit the “wrong” action A although, from an intuitive point of view, A is altogether altruistic, i.e. M only intends to further the welfare of her son S.

¹⁰ The following assignment of utilities is not only quite arbitrary but it also neglects an important distinction (explained at more length in Trapp [1988: 321-3]) between “Eigennutzen” and “Effektivnutzen” of an action A. The former expresses the “intrinsic” utility of A for the individual I_j , while the latter also takes into account (i) in which way A afflicts the interests of other individuals, and (ii) to which degree I_j wishes or prefers these interests to be afflicted by A.

In our example the value -10 assumed as the utility of A for M would plausibly have to be conceived of as “Eigennutzen” while the “Effektivnutzen” is likely to be much higher since the mother M was assumed to *want* to give the money to her son. I owe this point to Rainer Trapp.

Similarly, in the parallel situation where a (relatively poor) son S' plans to steal a certain amount of money from his (relatively rich) mother M' , we can assume the payoffs of action A' to be:

	M'	S'
A'	10	-5
$\sim A'$	0	0

Since the total utility of A' now is greater than that of $\sim A'$, utilitarian ethics - either in the form of the comparative criterion BETTER, or in the sense of the classificatory principle RIGHT - recommends the egoistic action A' as being morally better than $\sim A'$, or even as being morally obligatory, although by doing A' the son would do unnecessary harm to his mother.

The upshot of these considerations should be quite clear: the morality of an action A to be executed by a specific responsible agent I_1 must not be measured by the *total* utility of A but rather by – what I would like to call – the (total) “*altruistic value*”, of that action, i.e. the sum of $u_j(A)$ for all I_j *except agent I_1 itself*: $U_{alt}(A) := \sum_{i>1} u_i(A)$. Accordingly one better replaces the comparative criterion BETTER by the following principle:

(BETTER_{MORAL}) Action A of an agent I_1 is *morally better* than action A' iff A augments the happiness of the *other* individuals (I_2, \dots, I_n) to a higher degree than does A' , i.e. iff $U_{alt}(A) > U_{alt}(A')$.

Furthermore, the unacceptable qualitative criteria PERMITTED and/or FORBIDDEN should be replaced by much more liberal principles which, in the traditional terminology of utilitarianism, might be formulated as follows:

(PERMITTED_{MORAL}) Action A of an agent I_1 is *morally permitted* iff A has the tendency to augment the *happiness of the others* (I_2, \dots, I_n), i.e. iff:
 $U_{alt}(A) \geq U_{alt}(\sim A)$.

(FORBIDDEN_{MORAL}) Action A of an agent I_1 is *morally forbidden* iff A has the tendency to diminish the *happiness of the others* (I_2, \dots, I_n), i.e. iff:
 $U_{alt}(A) < U_{alt}(\sim A)$.

Note, incidentally, that these principles may be regarded as precisely refined versions of another traditional ethical doctrine, *Neminem laedere*, which, roughly, tells us that we should never do harm to others.

6 Supererogation

The very existence of the *concept* of supererogation shows that, in a common sense understanding of morality, not everything that is morally *good* will *eo ipso* be morally *obligatory*. Supererogatory actions, i.e. morally praiseworthy actions which go beyond what morality requires, certainly do exist. This fundamental intuition can be expressed by the following abstract principle which is independent of any *material* condition for moral obligations:

(SUPEREROGATION) The inference from the two premises
 (P1) Action A is morally obligatory
 (P2) Action A' is morally better than A
to the conclusion
 (C) Action A' is morally obligatory
is not generally warranted.¹¹

It is somewhat surprising to see that only very few philosophers have noted and explicitly discussed the problem that the utilitarian requirement of maximizing total utility stands in conflict with this principle. Thus, e.g., Franz von Kutschera pointed out that, according to utilitarian ethics, in a situation of choice among actions A_1, \dots, A_n person I is *morally obliged* to carry out the *optimal* action (or one of the optimal actions) A_j , and he goes on to explain such a conception of moral obligations as very doubtful since we ordinarily do not have the *duty* to act morally *optimal* on each occasion. Some (if not many) acts are *supererogatory*, i.e. doing them would be doing something beyond what is morally obligatory¹².

When discussing the question “Is Utilitarianism Too Demanding?”, SHAW [1999: 129] starts with a simple example: “Instead of buying a new car, we could make do with our old one and give the rest of the money to a charitable cause. And so on: our lives are rarely so productive of good that it would be impossible for us to do yet more”. Although he attempts to play down this problem by claiming “that in practice the theory’s implications are not so drastic”, SHAW [1999: 131] is soon forced to admit “that utilitarianism implies that morality

¹¹ For a closer discussion cf. WESSELS [2001].

¹² Cf. KUTSCHERA [1982: 22-23; 25]: „Eine Handlung ist erlaubt genau dann, wenn sie unter den möglichen Alternativen moralisch optimal ist“. „Es gibt supererogatorische Akte, d.h. Handlungen, mit denen man über seine Pflichten hinaus etwas Gutes tut“. Cf. also section III 1.8 of TRAPP [1988].

demands much [...] more than one might have supposed”. However, he does not take this as a *refutation* of utilitarianism – instead, he argues, our moral *intuitions* might simply be wrong:

“No doubt, the proposition that they are morally obliged to give thousands of their hard-earned dollars every year to famine relief would strike most middle-class Americans as preposterous. Yet it remains a fact that, in the larger scheme of things, such a sacrifice would pale into insignificance when compared to the benefits to those whose lives are imperiled by famine or disease” [SHAW 1999: 131].

But such considerations are far from convincing. Even if one is willing to admit that “in the larger scheme of things”, i.e. from an “objective” or unselfish point of view, the sacrifices of prosperous people in the industrial countries would be small or even “insignificant” in relation to the benefits for the poor people in the 3rd world, it still would not follow that such sacrifices are morally *obligatory* – it only follows that such altruistic actions are morally (very) *good*! In application to social scenarios as those envisaged by Shaw, the utilitarian idea of maximizing total utility basically entails an unrealistic and over-demanding requirement of maximizing voluntary financial offerings.

Now it certainly is no fault when philosophers emphasize how *good* such offerings would be from a moral point of view. Yet they should also be realistic enough to see that in their concrete everyday life, most utilitarians:

“[...] will not advocate a moral norm requiring people to give away most of what they have to help those in distress. Instead, utilitarians will uphold [at best] the less demanding norm [...] that we should aid strangers when the benefit to them is great and the cost to ourselves comparatively minor“ [SHAW 1999: 283].

Moreover, philosophers should also be bright enough to see that the specific utilitarian idea of *maximization* simply is theoretically misguided. As has been aptly noted by BRANDT [1967: 39], the core idea of Utilitarianism is just “the thesis that the moral predicates of an act [...] are *functions* in some way, direct or indirect, of consequences for the welfare of sentient creatures, and of nothing else. Utilitarians differ about what precise function they are, and they differ about what constitutes welfare and how it is to be measured. But they agree that all one needs to know, in order to make moral appraisals correctly, is the consequence of certain things for welfare.” Nothing in this basic idea entails that it would be *morally obligatory to maximize* the welfare of all sentient beings.

Literature

BENTHAM, Jeremy [1789]: *An Introduction to the Principles of Morals and Legislation*,
Oxford.

- BRANDT, Richard B. [1967]: "Some Merits of One Form of Rule-Utilitarianism", *University of Colorado Studies*, 39-65; German translation „Einige Vorzüge einer bestimmten Form von Regelutilitarismus“, in HÖFFE 1975, 133-161.
- FEHIGE, Christoph & Ulla WESSELS (eds.) [1998]: *Preferences*, Berlin (de Gruyter).
- HARSANYI, John C. [1982]: „Morality and the theory of rational behaviour“ in Amartya Sen & Bernard Williams (eds.), *Utilitarianism and beyond*, Cambridge (Cambridge University Press), 39-62.
- HÖFFE, Otfried [1975]: *Einführung in die utilitaristische Ethik – Klassische und zeitgenössische Texte*, München (Beck).
- HUTCHESON, Francis [1725]: *An Inquiry Into the Original of Our Ideas of Beauty and and Virtue*, London.
- KUTSCHERA, Franz [1982]: *Grundlagen der Ethik*, Berlin (de Gruyter), here quoted according to the second edition of 1999.
- LENZEN, Wolfgang [1980]: "Ist gut", in *Akten des 5. Internationalen Wittgenstein-Symposiums*, Kirchberg a.W., 165-171.
- LENZEN, Wolfgang [1999]: *Liebe, Leben, Tod – Eine moralphilosophische Studie*, Stuttgart (Reclam).
- LYONS, David [1965]: *The Forms and Limits of Utilitarianism*, Oxford (Clarendon).
- MILL, John Stuart [1861]: *Utilitarianism*, here quoted according to the edition by G. Sher, Indianapolis Cambridge (Hackett), 1979.
- RAWLS, John [1972]: *A Theory of Justice*, Oxford, (Oxford University Press).
- SHAW, William S. [1999]: *Contemporary Ethics – Taking Account of Utilitarianism*, London (Blackwell).
- SIDGWICK, Henry [1874]: *The Methods of Ethics*, MacMillan and Co., 7th ed. London 1907; reprint Bristol, Thoemmes Press, 1996.
- SMART, J. J. C. [1961]: *An Outline of a System of Utilitarian Ethics* (Melbourne University Press).
- SMART, J. J. C. [1973]: "An outline of a system of utilitarian ethics", in J.J.C. Smart & Bernard Williams, *Utilitarianism for and against*, Cambridge (Cambridge University Press).
- TRAPP, Rainer [1988]: »Nicht-klassischer« *Utilitarismus – Eine Theorie der Gerechtigkeit*, Frankfurt a.M. (Klostermann).
- WESSELS, Ulla [2001]: *Die gute Samariterin - Prolegomena zu einer jeden künftigen Theorie der Supererogation*, Habilitationsschrift Universität Leipzig.