

Act utilitarianism

Ben Eggleston

(PhilPapers version, uploaded in January 2017)

citation for published version:

Ben Eggleston, "Act utilitarianism," in *The Cambridge Companion to Utilitarianism*, edited by Ben Eggleston and Dale E. Miller (Cambridge University Press, 2014), pp. 125–145.

abstract:

An overview (about 8,000 words) of act utilitarianism, covering the basic idea of the theory, historical examples, how it differs from rule utilitarianism and motive utilitarianism, supporting arguments, and standard objections. A closing section provides a brief introduction to indirect utilitarianism (i.e., a Hare- or Railton-style view distinguishing between a decision procedure and a criterion of rightness).

The basic idea

The definition of act utilitarianism

In nearly every part of the world, there is moral opprobrium attached to the idea of a doctor ending a patient's life, even if the patient sincerely requests it because he has a terminal, debilitating, and painful disease. Efforts to relieve such patients' pain are widely regarded as humane, but active euthanasia is widely condemned both ethically and legally. Suppose that, despite these prohibitions, a doctor gives a lethal injection to such a patient. Depending on the circumstances, the doctor might be subject to general excoriation, the revocation of her license, and even criminal prosecution. But has she done anything wrong?

Act utilitarianism, like other forms of utilitarianism, approaches questions of this kind by holding that morality is ultimately a matter of overall well-being. What distinguishes act utilitarianism from other, rival, forms of utilitarianism is the

extremely direct and straightforward way in which it specifies this basic utilitarian idea. It holds, quite simply, the following:

Act utilitarianism: An act is right if and only if it results in at least as much overall well-being as any act the agent could have performed.

In other words, in any situation, an agent acts rightly if she maximizes overall well-being, and wrongly if she does not. In the example given above, if the lethal injection promoted overall well-being at least as much as any act the doctor could have performed, then it was right, according to act utilitarianism. And if it did not, it was wrong.

Act utilitarianism, indeed utilitarianism more generally, is both broad and narrow in ways that are sometimes surprising to people when they first encounter the view. It is remarkably broad because of its account of whose well-being matters to the moral value of an act. A natural thought about well-being and morality is that only the well-being of people directly affected by an act can influence the moral value of that act. But act utilitarianism holds that all well-being – experienced by any being (human or otherwise), in any place (near or far), at any time (whether in the present or the remote future) – matters to the moral value of any individual act. For example, to the extent that Plato’s dialogues continue to bring pleasure (or displeasure) to twenty-first-century students of philosophy, the precise moral value of acts committed more than two millennia ago is still evolving. Thus, in the definition of act utilitarianism, in the phrase ‘overall well-being’, the breadth of the adjective cannot be overstated.

Though act utilitarianism is remarkably broad in the way just mentioned, there is another way, also having to do with well-being, in which act utilitarianism is strikingly narrow. Few would question the thought that the moral value of an act depends at least in part on whether it makes people better or worse off: only a truly bizarre moral theory would hold that well-being does not matter, morally. But act utilitarianism goes to the opposite extreme, holding that *only* well-being matters, morally. Whatever other properties a particular act might have – e.g., that it was a felony, or was an instance of disloyalty, or was done from selfish motives – these properties do not have any independent relevance to the moral value of the act.

They might indicate ways in which the act has affected or will affect overall well-being, but they do not matter in and of themselves. This narrowness of act utilitarianism is arguably the most distinctive aspect of the view and, indeed, the utilitarianism tradition generally: that it focuses on this one item that is widely regarded as relevant to the moral values of acts – their effects on overall well-being – and declares that nothing *else* is relevant to the moral values of acts.

Three further clarifications

An artificially schematic but nonetheless useful way to think about act utilitarianism is in terms of the choice situation that a given agent faces at a given time. Suppose that, in a particular choice situation, an agent can choose to perform any of, say, seven acts: $A_1, A_2, A_3, \dots, A_7$. Now suppose that, for any act A_i , we define $W(A_i)$ as the world that would result if act A_i were performed, so that in the case at hand, we have $W(A_1), W(A_2), W(A_3), \dots, W(A_7)$. We then rank these worlds according to the amounts of overall well-being they respectively contain. In the resulting ranking, either there will be just one of these worlds in first place or there will be two or more of these worlds tied for first place. If just one of these worlds is in first place, then the corresponding act is not only right but obligatory, with every other act being wrong. If several of these worlds are tied for first place, then no particular act is obligatory, but it is nonetheless obligatory for the agent to choose only from among the acts corresponding to the tied-for-first-place worlds, and any such act is right. As in the simpler case, every other act – every non-utility-maximizing act – is wrong.

This way of thinking about act utilitarianism is particularly useful in dispelling the objection that act utilitarianism is impractical or incoherent because its goal of utility maximization can never be achieved. This objection is based on the claim that regardless of the act any agent performs at any time, there will always be more work for him and others to do in the future toward the maintenance and production of well-being. Although this claim is obviously true, it does not present a problem for act utilitarianism, since act utilitarianism (like all prominent forms of utilitarianism) is compatible with the idea that promoting well-being is a never-ending enterprise rather than a discrete task that some agent might have the

opportunity to bring to completion. An agent's duty at any given time, according to act utilitarianism, is not to act so that the resulting world has as much overall well-being as a world can have, but just to act so that the resulting world has as much overall well-being as any world that could have resulted from the acts that were among the agent's options at the time of acting. In other words, the idea of maximization that act utilitarianism involves is the idea of maximizing over the agent's set of options, not the idea of maximizing in the sense of leaving no increases to be achieved subsequently.

A second idea meriting further clarification is that of a world's overall well-being. When we think about a world, in all its spatial vastness and the entire duration of its existence, what does its overall well-being consist of? For utilitarianism, it is just the sum of the well-being had by the entities in that world that are capable of having well-being. (In most utilitarian theories, these are the organisms that are capable of feeling pleasure and pain.) Each such entity will have some total, lifetime well-being – positive, it is to be hoped, but possibly negative – and the sum of those, for all of the world's creatures, is that world's overall well-being.

Conceiving of overall well-being in this way helps to pre-empt a misunderstanding than can arise from the phrase 'the greatest happiness for the greatest number', which dates from the eighteenth century¹ and remains, in common parlance, synonymous with utilitarianism. On its face, this phrase suggests that an act should not only produce as much happiness as possible, but should also produce happiness for as many people as possible. That makes this phrase problematic as a criterion of right action, since it is often the case that the most beneficial act is different from the act that will spread the benefit most widely (since, in many choice situations, a small set of people has much more at stake than the rest of humanity does). In contrast, when overall well-being is conceived simply as the sum of individuals' well-being (as explained above), the 'for the greatest

¹ See Hutcheson, *An Inquiry into the Original*, p. 125 (though there the last word of the phrase is plural), and Bentham, *Comment and Fragment*, p. 393.

number' part of the phrase proves otiose.² Maximizing overall well-being might often result from the act that benefits the most people, but even in that case the act is right (according to act utilitarianism) simply because it maximizes overall well-being, not because it benefits the most people. We may conclude, with Russell Hardin, that "No philosopher should ever take the dictum of the greatest good for the greatest number seriously except as a subject in the history of thought."³

Finally, we saw above that according to act utilitarianism, nothing other than overall well-being matters to the moral value of an act. For example, the fact that an act is a crime, or results from a vicious character trait, does not make it wrong; moreover, such a fact does not detract from its moral value at all, according to act utilitarianism. By the same token, act utilitarianism entails that the moral value of an act does not depend, at all, on whether the act complies with any kind of moral rule (other than the act-utilitarian rule of "Maximize well-being"). This is important because the concept of a rule is often regarded as integral to the concept of morality. For example, morality is often understood as the rules for the regulation of behavior that are generally accepted (in the agent's society, typically),⁴ or the rules that are generally accepted that satisfy some ethical criteria,⁵ or the rules that ought to be generally accepted, regardless of whether they are currently accepted.⁶ Potential examples of rules meeting one or more of these criteria are the prohibition against active euthanasia, mentioned above, and the requirement that the owners of pets keep them reasonably comfortable.

According to act utilitarianism, the fact that an act would comply with, or would violate, a rule that meets any criterion such as those just mentioned is irrelevant to its moral value: all that matters is how the act would affect overall well-being, relative to how alternative acts would affect overall well-being. This is not to say, of course, that in practice act utilitarianism is blind to the existence and

² Bentham, *Deontology*, p. 309. This page is in Bentham's "Article on Government" in that volume.

³ Hardin, *Morality within the Limits of Reason*, p. 22.

⁴ Such a possibility is discussed (critically) in Gert, *The Nature of Morality*, p. 119.

⁵ See, for example, R. B. Miller, "Actual Rule Utilitarianism," p. 22; see also p. 7.

⁶ See, for example, Hooker, *Ideal Code, Real World*, p. 32; see also p. 144, n. 3.

potential usefulness of moral rules. The existence of moral rules can affect the way an act benefits or harms people; for example, in a society with a moral rule against the cremation of the bodies of revered elders, such an act would have different consequences than in a society that accepts cremation as a valid practice. Moreover, it is consistent with act utilitarianism to hold that, as a matter of psychological and sociological fact, the existence of certain moral rules, in a given society or throughout the world, can be useful for the promotion of overall well-being, because they are an effective device for the restraint and coordination of behavior. The catch is that such rules would not, according to act utilitarianism, have any actual bearing on the moral value of acts done in that society, or anywhere. Act utilitarianism's simultaneous repudiation of moral rules (as irrelevant to the moral value of acts) and embrace of them (as potentially beneficial tools) is arguably the most subtle and complex aspect of act utilitarianism, and we will return to this topic in the last section of this chapter.

Historical and contemporary context

Act utilitarianism has a long history, having been espoused in landmark utilitarian treatises such as Jeremy Bentham's *An Introduction to the Principles of Morals and Legislation* (1789),⁷ Henry Sidgwick's *The Methods of Ethics* (1st edn., 1874),⁸ and G. E. Moore's *Principia Ethica* (1903).⁹ For some or all of these authors it may have been, to some extent, merely the default form of utilitarianism rather than a conscious choice, since it was not explicitly and influentially formulated as a particular kind of utilitarianism until the 1950s. That decade, however, saw the emergence of rule utilitarianism as a well-defined alternative to act utilitarianism – presented as important both historically, for the interpretation of Mill's

⁷ Bentham, *An Introduction to the Principles of Morals and Legislation*, pp. 12–13 (in chapter 1).

⁸ Sidgwick, *The Methods of Ethics*, p. 411 (in bk. iv, ch. 1, § 1).

⁹ G. E. Moore, *Principia Ethica*, pp. 147–148 (in § 89). Moore's conception of what is to be maximized is not limited to well-being, so his view is not a form of act utilitarianism, strictly speaking. But it is close enough to have been influential in the development of utilitarian thought.

Utilitarianism (1861), and as a substantively plausible view¹⁰ – resulting in a correspondingly heightened precision in the delineation of act utilitarianism as one specific option within the broader utilitarian school of thought. The label ‘act-utilitarianism’ seems to have entered the philosophical literature in 1959,¹¹ and within two years there appeared the unhyphenated variant, which has become the more common term.¹²

To further contextualize act utilitarianism within the utilitarian school of thought, let us define rule utilitarianism more precisely:

Rule utilitarianism: An act is right if and only if it would be permitted by a system of rules whose general acceptance would result in at least as much overall well-being as would the general acceptance of any system of rules.

Rule utilitarianism affirms act utilitarianism’s claim that rightness is conceptually dependent on overall well-being, but denies act utilitarianism’s claim that the dependence is direct, or immediate: instead, it holds that the dependence is indirect, because it is mediated by rules. Now, it is possible to affirm rule utilitarianism’s claim that the dependence is mediated, but deny rule utilitarianism’s claim that rules are what do the mediating. One might, for example, privilege motives instead:

Motive utilitarianism: An act is right if and only if it would result from the motives whose general possession would result in at least as much overall well-being as would the general possession of any motives.¹³

Along these lines one can envision conscience utilitarianism,¹⁴ virtue utilitarianism,¹⁵ and so on. Such views are often labeled “indirect” forms of

¹⁰ See the works mentioned by Smart, “Extreme and Restricted Utilitarianism,” p. 344.

¹¹ Brandt, *Ethical Theory*, p. 380.

¹² Smart, *An Outline of a System of Utilitarian Ethics*, 1961 edn., p. 2.

¹³ This view is adapted from Adams, “Motive Utilitarianism,” p. 470, where the phrase is suggested as a name for a particular view about the moral value of patterns of motivation.

¹⁴ See, e.g., Adams, “Motive Utilitarianism,” p. 479; Brandt, *Facts, Values, and Morality*, p. 145; and Hooker, *Ideal Code, Real World*, p. 2, pp. 90–92, and pp. 131–132.

¹⁵ See, e.g., Crisp, “Utilitarianism and the Life of Virtue,” p. 154.

utilitarianism (though this term is problematic because it is also used to describe a kind of act utilitarianism, as explained below). Rule utilitarianism is the most thoroughly developed and discussed of these, and is the main rival of act utilitarianism within contemporary utilitarian thought. I will discuss rule utilitarianism further below, and Dale Miller discusses it more thoroughly (and sympathetically) in his chapter in this volume (Chapter 7). Meanwhile, it is worth noting that act utilitarianism's claim that rightness is *directly* conceptually dependent on well-being is one of its most important characteristics.

Some theorists who accept the directness of act utilitarianism object to a different component of it: that of *maximization*, in its requirement that acts maximize overall well-being. Most such theorists recommend, instead, the concept of "satisficing," first presented by Michael Slote as holding that "an act might qualify as morally right through having good consequences, even though better consequences could have been produced in the circumstances."¹⁶ (Slote took the term 'satisfice' from the writings of the economist Herbert Simon.¹⁷ Although it is often assumed to be a portmanteau of 'satisfy' and 'suffice', Slote notes that "it is a Scotticism for 'satisfy'.")¹⁸ Proponents of satisficing forms of utilitarianism generally claim that its demands are more reasonable than are those of maximizing forms of utilitarianism; for example, an act that results in a great deal of overall well-being but does not happen to maximize well-being might be right according to a satisficing form of utilitarianism but would, obviously, be wrong according to act utilitarianism.¹⁹

In response to the satisficing proposal, defenders of maximizing sometimes express skepticism that the view can really amount to anything other than a nuanced form of maximizing. For example, Robert Goodin writes that "maximization under constraints of time and information costs" is "the best sense I can make of"

¹⁶ Slote, "Satisficing Consequentialism," part I, p. 140.

¹⁷ Slote, "Satisficing Consequentialism," part I, pp. 141–142.

¹⁸ Slote, "Two Views of Satisficing," p. 27, n. 1.

¹⁹ Sinnott-Armstrong, "Consequentialism," section 6.

the view.²⁰ And when satisficing is construed as a genuine alternative to maximizing, defenders of the latter insist that it cannot be permissible to intentionally choose consequences that one acknowledges to be all-things-considered worse than consequences that one might choose instead. In short, many follow Philip Pettit in claiming “that [a defender of satisficing] is committed to unmotivated sub-maximization and that this is profoundly irrational.”²¹ The relative merits of maximizing and satisficing continue to be debated.²²

The final component of the contemporary context of act utilitarianism concerns the many questions to which there is no definite act-utilitarian answer, so that act utilitarians can (and do) differ among themselves about how best to answer them. Perhaps the most prominent of these questions is about the nature of well-being, which is discussed in the chapters by Chris Heathwood and Ben Bradley (Chapters 10 and 11). Act utilitarianism requires the maximization of well-being, but is compatible with various conceptions of well-being. Similarly, act utilitarianism is compatible with various answers to the question of whether the moral value of an act depends on its actual effects on overall well-being or the effects that could reasonably have been expected when the act was performed. (We often say to people who unwittingly do harm, “You couldn’t have known.”) These possibilities, and others, are discussed in the chapter by Elinor Mason (Chapter 9). As a final example, act utilitarianism is compatible with various answers to the question of whether what is to be maximized is the total quantity of well-being or the average of the levels of well-being had by the entities that are capable of having well-being. The significance of this issue, and the most important arguments that bear on it, are explained in the chapter by Tim Mulgan (Chapter 16). Above, act utilitarianism is presented in terms that explicitly or implicitly refer to an act’s *actual* effects on *total* well-being, but this is just for expository convenience: a sufficiently flexible formulation would be cumbersomely complex.

²⁰ Goodin, *On Settling*, p. 34 and p. 83, n. 15.

²¹ Pettit, “Satisficing Consequentialism,” part II, p. 172.

²² For recent discussions, see the papers collected in Byron, *Satisficing and Maximizing*; and B. Bradley, “Against Satisficing Consequentialism.”

Supporting arguments

The range of arguments that can be given in support of act utilitarianism is remarkably diverse, and many of the arguments are complex. But most of them are elaborations of one of two basic strategies that can be presented here in a stylized form.²³

Respecting individuals' interests

The first strategy for justifying act utilitarianism starts with individuals' interests, and regards morality as primarily concerned with resolving conflicts between those interests. Some such conflicts have obvious resolutions. For example, some individual might have an interest in owning a bank account that someone else owns instead. But some such conflicts are not so one-sided. For example, some commuters might have an interest in the widening of a particular road while nearby residents prefer the status quo. This strategy for justifying act utilitarianism sees an individual's interests as constituting his or her well-being. So, conflicts between individuals' interests are seen as conflicts between individuals' well-being. Thus, morality is primarily concerned with what should be done when increasing one individual's well-being entails decreasing (or just declining to increase) another individual's well-being.

To resolve such conflicts, this strategy for justifying act utilitarianism holds that the strength of individuals' claims to the maintenance and improvement of their well-being is proportional to the *magnitudes* of the *changes* in their well-being that are under consideration in a particular case.²⁴ Thus, in a two-person conflict, if the first person stands to experience a large increase in well-being and the second person's well-being will be reduced only slightly, the first person's claim would have more moral weight. This strategy also holds that, in cases affecting more than two people, the strengths of multiple individuals' claims should be combined. Consider,

²³ My division of arguments into these two kinds, along with some of the examples I cite, is indebted to Kymlicka, *Contemporary Political Philosophy*, pp. 32–37.

²⁴ This is controversial, as explained in the discussion of the Transitional Equity principle in Krister Bykvist's chapter in this volume (Chapter 5).

for example, a case resembling the previous one, except that instead of just one person facing a slight reduction in well-being, there are several. If they are numerous enough, their claims would have more moral weight than the first person's claim, even though none of their claims, taken individually, would have more moral weight than the first person's claim.

Now, it so happens that the foregoing way of resolving conflicts between individuals' well-being is equivalent to the maximization of the total quantity of well-being: if every conflict is resolved in accordance with the claims corresponding to the largest amounts of well-being at stake, no outcome containing less well-being than another possible outcome will ever be chosen. Thus, on this view, the moral imperative of respecting individuals' interests is made more determinate, and summed up, by the act-utilitarian principle of maximizing overall well-being.

Sum-ranking welfarist act consequentialism

The second strategy for justifying act utilitarianism begins with a focus on states of affairs, rather than people (or other individuals), and it holds that some states of affairs are better than others. It then makes this idea more determinate by embracing two additional theses. One of these holds that only well-being contributes to the goodness of a state of affairs:

Welfarism: The value of a state of affairs is positively related to, and determined by nothing other than, the well-being it contains.²⁵

Now, this thesis is compatible with several mutually exclusive theses about how the value of a state of affairs is determined by the valuable things (such as well-being) it contains. One of these is concerned with equality, holding that the value of a state of affairs is positively related to, and determined by nothing other than, how equally the valuable things (whatever they may be) are distributed in that state of affairs.²⁶

²⁵ Sen, "Utilitarianism and Welfarism," provides a thorough discussion of welfarism and an influential critique of it, particularly with a view to its use in utilitarianism. See especially pp. 471–489.

²⁶ Such a view is discussed at length in Dworkin, "What Is Equality? Part 1: Equality of Welfare."

Another is concerned with minimizing disadvantage, holding that the value of a state of affairs is positively related to, and determined by nothing other than, the quantity of valuable things that is enjoyed by the individual who has the smallest quantity of it.²⁷ A third is concerned with maximizing the total quantity of what is valuable. This, of course, is the thesis that contributes to the present strategy for justifying act utilitarianism:

Sum-ranking: The value of a state of affairs is positively related to, and determined by nothing other than, the total quantity of value it contains.²⁸

Combining the theses of welfarism and sum-ranking yields the view that the value of a state of affairs is positively related to, and determined by nothing other than, the total quantity of well-being it contains. This view is, obviously, very close to act utilitarianism.

But this view is only about states of affairs, not acts. So the justificatory strategy under consideration embraces one further thesis, about the way the value of every act depends on its consequences:

Act consequentialism: An act is right if and only if its consequences are at least as good as the consequences of any act the agent could have performed.

This thesis, combined with welfarism and sum-ranking, completes this justification for act utilitarianism. This justification is, then, a *sum-ranking welfarist act-consequentialist* one.

²⁷ This thought has obvious affinities with John Rawls's difference principle (Rawls, *A Theory of Justice*, pp. 75–83). This thought is developed differently in prioritarianism, a much-discussed cousin of utilitarianism for which the most-discussed source is Parfit, "Equality and Priority." Also see the earlier McKerlie, "Equality and Priority."

²⁸ Sen, "Utilitarianism and Welfarism," pp. 468–471. Also see the discussion of sum-ranking welfarism in Krister Bykvist's chapter in this volume (Chapter 5).

Contrast and historical examples

Although the two strategies obviously have much in common, they are fundamentally different. The first strategy takes individuals' interests as fundamental and takes morality to be primarily concerned with resolving conflicts between those interests. The concept of maximization comes in fairly late in the proceedings, almost as a mathematical accident. If this strategy's principle for how to assess the relative strengths of competing claims were altered, the resulting view could fail to be a maximizing one.

In contrast, the second strategy is an essentially maximizing one: it takes morality to be primarily concerned with maximally promoting valuable states of affairs. Correspondingly, individuals, and their interests, come in somewhat later – if not accidentally, then certainly more subordinately. If this strategy's principle for what determines the value of a state of affairs were altered, morality might not have anything to do with individuals, and their interests, at all.

Perhaps because of the second strategy's impersonal, "top down" orientation, the first strategy has been preferred by more of the major figures in the history of utilitarianism. Variations of it are arguably deployed by Bentham ("every individual in the country tells for one; no individual for more than one")²⁹ and Mill ("To do as one would be done by, and to love one's neighbour as oneself, constitute the ideal perfection of utilitarian morality"),³⁰ as well as by the twentieth-century and contemporary theorists John Harsanyi ("a social welfare function ought to be based ... on the utility functions (subjective preferences) of *all* individuals, representing a kind of 'fair compromise' among them"),³¹ R. M. Hare ("We are led to give weight to

²⁹ Bentham, *Rationale of Judicial Evidence*, vol. VII, p. 334. Philip Schofield identifies this as the source of what Mill calls "Bentham's dictum": "everybody to count for one, nobody for more than one" (*Utility and Democracy*, p. 84, n. 25).

³⁰ J. S. Mill, *Utilitarianism, Collected Works*, vol. x, p. 218.

³¹ Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," p. 315. See also Harsanyi, "Morality and the Theory of Rational Behaviour," pp. 44–48.

the preferences of all the affected parties ... in proportion to their strengths”),³² and Peter Singer (“when we make ethical judgments ... we weigh interests”).³³

Although the first strategy has been the more popular one, the second strategy has had its influential exponents as well. For example, Sidgwick is famous for suggesting that the moral point of view is “the point of view ... of the Universe,” and he reports that “it is evident to me that as a rational being I am bound to aim at good generally ... not merely at a particular part of it.”³⁴ Sidgwick’s cosmological aspirations are shared by Moore, who focuses on “the greatest possible amount of good in the Universe” and who writes that “the primary and peculiar business of Ethics” is not, for example, the resolving of people’s conflicts, but “the determination [of] what things have intrinsic value and in what degrees.”³⁵ Moreover, the second strategy may enjoy greater prominence than its historical frequency would suggest because of its neat factoring of utilitarianism into distinct components³⁶ and because of the rise, in recent decades, of consequentialism as a focal point within the discipline of moral philosophy.³⁷

Objections

The basic idea of act utilitarianism has a certain obvious appeal: well-being is a fine thing, and of course folks should have more of it rather than less. There are, however, several important objections to act utilitarianism. These objections have been prominent throughout the history of the view as well as presenting ongoing challenges for contemporary theorists.

³² Hare, “The Structure of Ethics and Morals,” p. 187. See also Hare, *Sorting Out Ethics*, p. 145 (echoing Bentham’s dictum).

³³ Singer, *Practical Ethics*, p. 20.

³⁴ Sidgwick, *The Methods of Ethics*, p. 382. The precise construction of the former phrase, in Sidgwick’s text, is ‘the point of view (if I may say so) of the Universe’.

³⁵ Moore, *Principia Ethica*, p. 147 (in § 89) and p. 26 (in § 17).

³⁶ See, for example, the excellent overview in Scarre, *Utilitarianism*, pp. 4–26.

³⁷ Early visible examples of this trend include Scheffler, *The Rejection of Consequentialism* (1982); and Railton, “Alienation, Consequentialism, and the Demands of Morality” (1984).

Impractical to implement

Perhaps the most straightforward objection to act utilitarianism is that it is impractical to implement in everyday decision-making, due to the difficulty of predicting the effects on well-being of all of the possible acts that make up a given choice situation. Mill anticipates this concern, at least in part, and replies that during “the whole past duration of the human species” people “have been learning by experience the tendencies of actions.”³⁸ For example, experience has shown that people tend to become corrupted by power.³⁹

But tendencies are not enough: act utilitarianism holds that the moral value of an act depends on all of its effects on well-being, however atypical, far-flung, or delayed they may be. And because many of an act’s effects on well-being are unforeseeable, act utilitarianism seems to require, for its competent implementation, an impossible degree of foresight. For example, people aiding strangers risk inadvertently putting those strangers in harm’s way.⁴⁰ And Hitler’s ancestors could not have known, when they engaged in procreation, that their actions would eventually be among the causes of the Holocaust.⁴¹ Such examples suggest that act utilitarianism may be impractical to implement, compromising its initial appeal.

Harmful if implemented

A second objection also focuses on act utilitarianism’s implementation, but moves beyond the issue of bare feasibility to claim that such implementation would have severe consequences for ordinary decision-making, social order, and virtue. When a high-placed government official of eighteenth-century England called the utilitarian principle a “dangerous” one, Bentham playfully pretended to fail to

³⁸ J. S. Mill, *Utilitarianism, Collected Works*, vol. X, p. 224.

³⁹ J. S. Mill, *Considerations on Representative Government, Collected Works*, vol. XIX, p. 445.

⁴⁰ McCloskey, “Utilitarianism: Two Difficulties,” p. 62.

⁴¹ This example is adapted from pp. 344–345 of Lenman, “Consequentialism and Cluelessness,” which offers a sophisticated presentation and discussion of this objection. Notable (and dissimilar) replies to Lenman include Dorsey, “Consequentialism, Metaphysical Realism and the Argument from Cluelessness”; and Burch-Brown, “Clues for Consequentialists.”

understand how it could ever be “not consonant to utility to consult utility” – before explaining, with equal zest, that what his contemporary feared, quite rightly, were the reforms utilitarianism would prescribe for institutions that bestowed and perpetuated unmerited privilege.⁴²

But the implementation of act utilitarianism, critics claim, would have consequences more troubling than those that worried eighteenth-century elitists. First, if people were to set aside the common-sense morality that prevails today and were to adopt the practice of making decisions according to the act-utilitarian standard of maximizing overall well-being, decision-making itself would become cripplingly tedious and time-consuming.⁴³ Second, there would be much more selfish behavior, since predictions of consequences are often fraught with uncertainties and people have a well-known tendency to resolve such uncertainties in ways that agree with their own interests.⁴⁴ Third, coordination would break down, since people would expect one another not to stick to previously made plans, but to regard every choice point as a fresh opportunity for maximizing.⁴⁵ Fourth, there would also be breakdowns of socially beneficial virtues such as honesty, the keeping of promises, and the special ties constitutive of love and friendship, since such virtues require people to act on principles that are not fully captured by the act-utilitarian goal of maximizing overall well-being.⁴⁶ In sum, the result would be little more than slow and selfish decision-making conducted by uncoordinated moral deficients.

⁴² Bentham, *Comment and Fragment*, p. 447; see also p. 516.

⁴³ See, e.g., Railton, “Alienation, Consequentialism, and the Demands of Morality,” pp. 153–154.

⁴⁴ See Shaw, *Contemporary Ethics*, p. 146; and Hooker, *Ideal Code, Real World*, p. 143.

⁴⁵ See Hodgson, *Consequences of Utilitarianism*, chapter 2; and Shaw, *Contemporary Ethics*, pp. 146–147.

⁴⁶ See, e.g., Butler, *Works*, vol. II, pp. 190–192; Sidgwick, *The Methods of Ethics*, p. 136 and p. 405; Hodgson, *Consequences of Utilitarianism*, pp. 58–59 and p. 61; Stocker, “The Schizophrenia of Modern Ethical Theories,” pp. 458–461; Wolf, “Moral Saints,” pp. 427–430; and Kapur, “Why It Is Wrong to Be Always Guided by the Best,” pp. 489–494.

Such a prospect would, of course, reflect badly on any moral theory. But it is especially discrediting to act utilitarianism, critics argue, since that theory's core ideal is the maximization of overall well-being and the prospect just sketched is, among its many failings, an utter debacle on that score. On these grounds, act utilitarianism is often said to prohibit its own implementation, and to be "self-defeating."⁴⁷

Immoral implications

The most serious and influential objection to act utilitarianism concerns the moral judgments that act utilitarianism entails for particular cases of moral decision-making. According to this objection, there are countless cases – reflecting diverse aspects of morality – in which act utilitarianism entails judgments that are questionable or utterly unacceptable. Such cases, it is claimed, show that act utilitarianism misconstrues, or just runs roughshod over, many important aspects of morality.

One such aspect of morality includes the various special obligations that people are often thought to have. Promises, as well as figuring in the preceding section, make for an apt example here: people are often thought to have special obligations in virtue of promises they have made, but act utilitarianism is said to fail to give promises their proper moral weight. This claim is developed in a classic discussion from W. D. Ross:

Suppose ... that the fulfillment of a promise to *A* would produce 1,000 units of good for him, but that by doing some other act I could produce 1,001 units of good for *B*, to whom I have made no promise ... We should, I fancy, hold that only a much greater disparity of value between the total consequences would justify us in failing to discharge our *prima facie* duty to *A*. After all, a promise is a promise,

⁴⁷ See, e.g., Hodgson, *Consequences of Utilitarianism*, p. 3 and p. 60; and Parfit, *Reasons and Persons*, pp. 27–28 and pp. 40–41.

and is not to be treated so lightly as the theory we are examining would imply.⁴⁸

The special obligations that stem from promises are not the only ones that act utilitarianism is said to neglect. Others include the special obligations that people have to other people in virtue of what those other people have earned, or deserve; and the special obligations that people have to their family and friends. All of these cases, it is said, show that morality is not just a matter of maximizing overall well-being.

A second aspect of morality that act utilitarianism is said to violate has to do with treating individuals justly: act utilitarianism is said to be too ready to impose grave harms on some people in order to provide benefits to others. Consider this case provided by T. M. Scanlon:

Suppose Jones has suffered an accident in the transmitter room of a television station. Electrical equipment has fallen on his arm, and we cannot rescue him without turning off the transmitter for fifteen minutes. A World Cup match is in progress, watched by many people, and it will not be over for an hour. Jones's injury will not get any worse if we wait, but his hand has been mashed and he is receiving extremely painful electrical shocks.⁴⁹

Scanlon asserts that we should rescue Jones immediately, regardless of how many people are watching the match. But act utilitarianism implies that if the viewers are numerous enough, we should wait. In a similar vein, other theorists have argued that act utilitarianism condones the judicial punishment of innocent people, depending on the facts of the situation. Cases can be imagined in which the harm experienced by the innocent person is outweighed by other benefits, such as quieting social unrest or deterring other people from performing harmful acts.⁵⁰ Act

⁴⁸ W. D. Ross, *The Right and the Good*, pp. 34–35.

⁴⁹ Scanlon, *What We Owe to Each Other*, p. 235.

⁵⁰ See, e.g., McCloskey, "A Non-Utilitarian Approach to Punishment," pp. 255–257; and Boonin, *The Problem of Punishment*, pp. 41–52.

utilitarianism's readiness to regard harms to some people as outweighed by benefits to others is part of the basis of John Rawls's famous claim that act utilitarianism "does not take seriously the distinction between persons."⁵¹

A third aspect of morality where act utilitarianism is said to go astray involves the sacrifices that it requires individuals to make in order to provide benefits to other people. Cases illustrating this issue are structurally similar to those concerned with treating individuals justly, except that the person experiencing the harm is the agent himself or herself, rather than another person. Paradigm cases concern the extent of the obligations of affluent people to donate money to poverty-relief programs. As it happens, such cases are asserted by proponents of act utilitarianism, as well as by opponents of the theory. The former argue that because of the moral imperative of promoting overall well-being, affluent people are obligated to donate much larger sums of money than is generally thought to be obligatory;⁵² the latter argue that because act utilitarianism is so demanding, it must be wrong.⁵³ (The former's *modus ponens* is the latter's *modus tollens*.) Despite some proponents' candor about the demandingness of act utilitarianism, this aspect of the theory remains one of the main grounds on which critics claim that it has a distorted view of morality.

Indirect utilitarianism

Overview

Mindful of the foregoing objections, contemporary proponents of act utilitarianism tend to advance a particular form of the view that is often called "indirect utilitarianism" (though, as noted earlier, this term is also often used to refer to rivals of act utilitarianism such as rule utilitarianism). Although proponents of this view intend for it to overcome or mitigate all of the foregoing objections, the second objection provides an especially convenient point of entry into this view. In response to this objection – that the implementation of act utilitarianism would be

⁵¹ Rawls, *A Theory of Justice*, p. 27.

⁵² See, e.g., Singer, "Famine, Affluence, and Morality"; and Unger, *Living High and Letting Die*.

⁵³ B. Williams, "A Critique of Utilitarianism," pp. 108–118.

harmful, making act utilitarianism self-defeating – defenders of act utilitarianism point out that this objection presupposes the use of act utilitarianism as what might be called a *decision procedure*. That is, it presupposes that people use act utilitarianism as a procedure for deciding what to do in ordinary choice situations. And defenders of act utilitarianism then concede that act utilitarianism is not well suited to be used in that way, for precisely the reasons stated in the second objection. But they claim that act utilitarianism is a defensible moral theory nonetheless, because it offers the correct *criterion of rightness* – the correct account of what makes actions right and wrong. On this view, the fact that act utilitarianism is not well suited for use as a decision procedure reflects, at most, something unfortunate about the psychological and social costs of pursuing the aims of morality too directly, and not any failure on the part of act utilitarianism to provide a sound account of what ultimately determines the moral values of acts.

This, then, leaves act utilitarians with the question of what decision procedure to recommend. The principle underlying their answer is simple: for any given person, the ideal decision procedure is the one whose possession and employment by that person would maximize overall well-being. For most people, the ideal decision procedure is probably some variant of common-sense morality: a decision procedure giving considerable weight to values such as honesty, the keeping of promises, the special ties constitutive of love and friendship, and so on. Of course, the exact contours of the ideal decision procedure for any given person is a complicated empirical question involving all of the myriad considerations mentioned in the articulation of the objection about the harmfulness of implementing act utilitarianism. Whatever the exact contours of the ideal decision procedure turn out to be, indirect utilitarianism is characterized by (1) affirming act utilitarianism as the correct criterion of rightness and (2) regarding the ideal decision procedure to be the one that best advances the goal of maximizing overall well-being.⁵⁴

⁵⁴ Some of the many notable works in the development of indirect utilitarianism are Bales, “Act-Utilitarianism”; Hare, *Moral Thinking* (especially chapter 2); Railton, “Alienation, Consequentialism, and the Demands of Morality”; and Crisp, “Utilitarianism and the Life of Virtue.”

It is important to avoid the misperception that the ideal decision procedure proposed by indirect utilitarianism is essentially act utilitarianism augmented with a collection of rules and guidelines carefully tailored to enable an agent who is consciously searching for the act that will maximize overall well-being to identify it, in the way that a shopper intent on buying the most delicious tomato might apply guidelines that recommended choosing tomatoes with a particular color, weight, firmness, or aroma. Rather, the ideal decision procedure being sketched here is one in which the agent values the goods mentioned above – honesty, the keeping of promises, the special ties constitutive of love and friendship, and so on – for their own sakes, even though act utilitarianism entails that these things matter merely as means to the promotion of overall well-being. Psychologically, such valuing might take the form of explicitly regarding certain rules as morally binding – perhaps the rules endorsed by rule utilitarianism – or might take the form of an unarticulated (but nonetheless firm) motive or disposition to act in certain ways in certain situations. In any case, whereas an agent using act utilitarianism as her decision procedure would unhesitatingly set any of the aforementioned goods aside when convinced that doing so would lead to the maximization of overall well-being, an agent with the ideal decision procedure would feel pangs of guilt at the prospect of setting any of them aside – even when she is convinced that doing so would maximize overall well-being. A person’s decision procedure is, in effect, her conscience, with all of the moral emotions that concept suggests. So, the ideal decision procedure proposed by indirect utilitarianism is not just a well-informed act utilitarianism. It is, rather, act utilitarianism complemented by other moral rules, motives, and dispositions. Although the resulting decision procedure contains elements with non-act-utilitarian content, it is recommended by act utilitarianism because of its favorable impact on overall well-being.

So, indirect utilitarianism rests on divorcing the notion of a criterion of rightness from the notion of a decision procedure, and maintaining that the correct criterion of rightness will not necessarily be advisable, or even self-endorsing, as a decision procedure. As a result, indirect utilitarianism contrasts interestingly with rule utilitarianism – the view that an act is right if and only if it would be allowed by (what is here called) the ideal decision procedure. Like rule utilitarians, indirect

utilitarians regard act utilitarianism as self-defeating, in the sense described above. But whereas rule utilitarians ensure agreement between the correct criterion of rightness and the ideal decision procedure by, in effect, regarding the ideal decision procedure as constituting the correct criterion of rightness, indirect utilitarianism divorces the two notions in order to maintain act utilitarianism as the correct criterion of rightness.

Assessment

The merits of indirect utilitarianism are a subject of ongoing debate. In support of the view, one might attempt to rebut the objections surveyed earlier by claiming that indirect utilitarianism improves on direct act utilitarianism by being easier to implement, by being more beneficial when implemented, and by endorsing the having of moral commitments that closely match what are often regarded as important aspects of morality.

But indirect utilitarianism is vulnerable to various criticisms as well. First, one might dismiss, as irrelevant, what moral commitments a moral theory endorses *the having of*; what matters, one might claim, are the theory's *implications*, and on this score indirect utilitarianism can offer no improvement over direct act utilitarianism, since indirect utilitarianism's criterion of rightness is simply the principle of act utilitarianism.⁵⁵ Second, indirect utilitarianism seems to confirm rather than answer a longstanding additional objection to act utilitarianism – the objection that it is ineligible to serve as society's publicly affirmed morality.⁵⁶ This follows from the substantial overlap between the idea of society's publicly affirmed morality and the idea of a moral theory as a decision procedure. Given that indirect utilitarianism involves disavowing act utilitarianism as a decision procedure, it seems to thereby concede that it cannot well serve as society's publicly affirmed

⁵⁵ Elsewhere, I argue against this dismissal; see Eggleston, "Practical Equilibrium."

⁵⁶ See, e.g., Sidgwick, *The Methods of Ethics*, pp. 489–490; and Hodgson, *Consequences of Utilitarianism*, p. 46. For contemporary responses to this objection, see Lazari-Radek and Singer, "Secrecy in Consequentialism"; and Eggleston, "Rejecting the Publicity Condition."

morality – or, at least, cannot well serve as the entirety of society’s publicly affirmed morality.

Indirect utilitarianism is clearly more complicated than direct act utilitarianism, and it challenges several conventions of moral theory as traditionally practiced. It might, however, be the most promising theoretical framework in which to embed the principle that morality is, fundamentally, simply a matter of maximizing overall well-being.