

Explain the 'collapse objection' to rule utilitarianism. Is the 'collapse objection' fatal to the project of developing a convincing rule-based alternative to act utilitarianism?

1. Yes, the "collapse objection" is fatal to rule-based alternatives to act utilitarianism, the ethical theory that an action is right only if it leads to the greatest total welfare as compared to any other possible action. Rule-based alternatives attempt to overcome criticisms of act utilitarianism such as the epistemic and moral debasement objections by redefining rightness as being exactly when an action is in accordance with a set of rules which, when uniformly followed, lead to the greatest total welfare as compared to any other set of rules. However, this rule-based approach fails to counter such challenges convincingly. I will outline these two objections to act utilitarianism, explain rule utilitarianism's responses, and show that these replies are either incoherent with utilitarian principles or simply indistinguishable from act utilitarianism. My goal is to demonstrate that rule utilitarianism cannot come to the rescue of its act-based sibling in countering such arguments, and so it collapses into effectively the same theory.
2. The epistemic objection argues that act utilitarianism does not fulfil its deliberative goal of providing instruction on what actions to take, and so cannot be a correct ethical theory. We can formalise the objection as follows:

P1: For an ethical theory to be true, it is necessary that it can be used as a guide for moral decision-making.
P2: Act utilitarianism requires computation of actions' consequences stretching out for an indefinite period of time and involving a great number of unknown factors.
P3: Such calculation of welfare over an indefinite time period is not humanly practicable and cannot possibly be relied on as a basis for decision-making.
C: Therefore, act utilitarianism cannot be true.

As Crisp (1997, p101) notes, this objection does seem to pose a substantial problem for act utilitarianism. If (as we currently have it defined) the right action is the one which actually leads to the greatest welfare, then not only is arriving at a decision about which action to take computationally intractable, the right action is in some sense undefined until some end point in time which may not even exist. Williams (1993, p85) suggests that an attractive feature of act utilitarianism is that under it, "all moral obscurity becomes a matter of technical limitations". If these limitations are insurmountable by all, though, that is hardly a selling point for the theory.

3. A rule utilitarian such as Urmson (1953) might respond that their theory offers a solution: rather than requiring agents to calculate the consequences of every action they take and selecting the one which maximises welfare, they can instead follow a pre-determined and interpretable list of rules which specify what action is right given the circumstances. Whilst it is true that, were such a set of rules to exist, they would sidestep the epistemic objection, a rule utilitarian must explain how an agent (or indeed humanity) would arrive at this set of rules in the first place.
4. Unfortunately, rule utilitarians here find themselves faced with the same problem as act utilitarians. For instance, although it might seem that welfare is maximised when the general rule "Do not murder" is followed, no human can really be sure that this is the case, given that we cannot compute with certainty the long-term effects on welfare if the routine were routinely broken. Instead, we fall back on observations made up to the present point to conclude that the maxim "Do not murder" is part of the rule set which maximises welfare. Rule utilitarianism doesn't rely on us having certainty that a particular action recommended by the theory is the right one, merely that we have good reason to think it is.

5. However, lowering the standard of proof to this level means act utilitarianism is also able to clear it. If we are to do what we have good reason to believe the criterion says is right, then an act utilitarian can just turn to probabilism (Crisp 1997, p100) and perform the action which a rational agent would expect to maximise utility, letting the unknowable and effectively random long-term consequences of an action cancel out. Moreover, act utilitarianism does not exclude agents from using heuristics in place of computing welfare every time they take an action. Since computation is itself an action which consumes time and is either right or wrong (Williams 1993, p90), falling back on a rule of thumb may be reasonably considered the right action from an act utilitarian standpoint. As Mill (2.24) argues, there is no need for a utilitarian to constantly return to first principles when heuristics will do the job.¹ So, if one only needs a reasonable expectation that the decision criterion tracks the rightness criterion to apply the theory, then act utilitarianism also becomes robust against the epistemic objection, and thus rule utilitarianism provides no advantage here.
6. Another objection to act utilitarianism is that it leads to a debasement of moral standards, and endorses (indeed, requires) actions which appear to us to be wrong.² Williams (1993, p91) points to the conviction of an innocent man to prevent wider disorder as one such case.³ Rule utilitarianism would seem to provide a different, more welcome conclusion: since uniform adherence to the rule “Do not convict innocent people” leads to greater welfare than non-adherence, we should follow it even if doing so leads to lower welfare in a particular instance. Yet this is simply “pure irrationality” (Williams 1993, p94). As Smart (1973, p12) argues compellingly, if rule utilitarianism truly does view welfare as a fundamental constituent of the criterion of rightness, the rule would surely be modified to “Do not convict innocent people, except where doing so prevents mass riots”. There would of course be exceptions to this exception, and so on with infinite regress, until the rules given are specific to every single possible situation, and reducible to the simple act utilitarian maxim of “Do whichever action will lead to the greatest welfare”. Alternatively, rule utilitarians might insist on following their rule, but this would then prioritise rule-following over welfare, removing any of its utilitarian qualities (Williams 1993, p95).
7. In light of these arguments, I conclude that the “collapse objection” is fatal to rule-based alternatives to act utilitarianism, and that such alternatives provide no advantages in fulfilling the deliberative or explanatory goals of an ethical theory.

References

- Roger Crisp, “Routledge Philosophy Guidebook to Mill on Utilitarianism” (Routledge, 1997): 95-124.
- J.S. Mill, “Utilitarianism” in “The Blackwell Guide Mill's Utilitarianism” ed. Henry R. West (Blackwell, 2006).
- J.J.C. Smart, “An Outline of a System of Utilitarian Ethics” in Smart & Williams, “Utilitarianism: For and Against” (Cambridge, 1993).
- J. O. Urmson, “The Interpretation of the Moral Philosophy of J. S. Mill”, *The Philosophical Quarterly* vol. 3, no. 10 (1953): 33-39.
- Bernard Williams, “Morality: An Introduction to Ethics” (Cambridge, 1993): 82-98.

One might here object “Why should following particular heuristics necessarily lead to the right action?”, but for the purposes of this essay it is sufficient to reply that one can equally fairly object “Why should following particular rules necessarily lead to the right action?” against rule utilitarianism. If humans were superintelligent act utilitarians, then we would very likely make different choices to the ones selected by an ordinary act utilitarian. But note the important qualification in the act utilitarian’s criterion of rightness: “the welfare-maximising action of all *possible* actions”. In some sense we can argue that the theoretical welfare-maximising action for a more knowledgeable agent is not within the action space for ordinary humans, because we would have to spend too long trying to compute our way towards it, itself an action that would be wrong (since it would not lead to the maximum welfare). Put simply, given the constraint of human knowledge and rationality, our set of possible actions is such that the way to maximise the expected utility we produce is to follow established heuristics, except for occasions when these second principles conflict.

² Williams asserts that it “seems wrong” to convict an innocent man, but this is not a particularly powerful argument. Whilst we might expect moral axioms such as the Greatest Happiness Principle to line up with our intuitions, that does not necessarily mean that every conclusion derived from them also should (Smart 1997, p56). In logic, the principle of *ex falso quodlibet* is surprising and seems wrong on first glance, but step-by-step analysis from first principles can reassure us that it is indeed correct. Similarly in ethics, valid reasoning from sound premisses may lead us to correct conclusions which nonetheless appear undesirable or repugnant.

³ Note that it is not necessarily the case that act utilitarianism would actually advise this course of action. Some critics assume that utilitarians are not able to consider the foreseeable second-order effects of their actions, when clearly any rational person would know that convicting an innocent man has negative side-effects which must be weighed carefully against the benefits it may bring. Mill himself pre-empts this objection by acknowledging that “certain moral requirements [collectively named justice]... stand higher in the scale of social utility... than any others”, and so should not be violated without extremely good reason (5.37), leaving open the possibility for pure act utilitarianism to (almost always) respect principles of justice without a need to introduce rule utilitarianism.