Rule-consequentialism and Internal Consistency: A Reply to Card

BRAD HOOKER
University of Reading

Rule-consequentialism has been accused of either collapsing into act-consequentialism or being internally inconsistent. I have tried to develop a form of rule-consequentialism without these flaws. In this June's issue of *Utilitas*, Robert Card argued that I have failed. Here I assess his arguments.

Perhaps the most influential objection to rule-consequentialism has been that either this moral theory must collapse into extensional equivalence with the simpler theory act-consequentialism, or it escapes collapse only by rendering itself internally inconsistent. Since 1993, I have argued that this objection can be decisively refuted.[1] In a recent article, Robert Card contends that I have not adequately answered the objection.[2] His article forces me to make some clarifications and concessions. But I will argue here that these are not fatal to rule-consequentialism.

The best way of defining consequentialism must avoid suggesting that the only form is *impartial maximizing act*-consequentialism. As an account of wrong acts, consequentialism should be defined as holding that whether an act is wrong depends solely on how the act's consequences compare with the consequences of alternative acts, or on how the consequences of rules or practices or motives that allow the act compare with the consequences of rules or practices or motives that prohibit the act. This definition shows why different theories share 'consequentialism' as their family name.

I think the most plausible form of consequentialism has a foundational principle that is impartial, maximizing and focused on the consequences of rules. The rule-consequentialism I develop holds that

[1] My 'Rule-consequentialism, Incoherence, Fairness' was presented in 1993 at many universities and at the Pacific Division meeting of the American Philosophical Association. It appeared in print in the autumn of 1994 in *Proceedings of the Aristotelian Society* 95. My *Ideal Code, Real World: A Rule-consequentialist Theory of Morality* (Oxford, 2000) develops the relevant points.

[2] Card, 'Inconsistency and the Theoretical Commitments of Hooker's Rule-consequentialism', *Utilitas* 19.2 (2007). My references to this article will be placed in my text.

© 2007 Cambridge University Press
doi:10.1017/S0953820807002786
Utilitas Vol. 19, No. 4, December 2007
Printed in the United Kingdom

https://doi.org/10.1017/S0953820807002786 Published online by Cambridge University Press

an act is wrong if it is forbidden by a code of rules whose internalization by 90 per cent of each future generation maximizes expected value, impartially construed. Here I do not have space for explanation.[3]

One argument for rule-consequentialism starts from the idea that the overarching goal of morality is to produce the best consequences. Employing act-consequentialism is not likely to produce the best consequences because trying to make decisions on a moment-by-moment basis by always calculating the expected value of alternative possible actions and choosing what to do on the basis of that calculation would be a terribly counter-productive procedure. Such a procedure would take more information, time, energy, powers of reasoning and impartiality than people have, or could cost-effectively be brought to have. So act-consequentialism is not the principle whose employment would produce the best consequences.

Perhaps the rules whose internalization and employment would produce the best consequences, or at least the highest expected value, are:

- Rules forbidding imposing physical harm on those not threatening any innocent person
- Rules against taking or harming others' property
- Rules against promise-breaking and lying
- Rules requiring one to take special responsibility for those with whom one has certain special relationships, and
- A rule requiring one to sacrifice one's own lesser good for the sake of benefiting others more, at least until one has reached some significant level of aggregate self-sacrifice.

There are important issues about how to interpret these rules and the vagueness in them. Still, it is highly plausible that these rules are the ones whose internalization and employment would produce maximum expected value. And it is also plausible that the moral theory that fits best with these rules is rule-consequentialism. Call this the *consequentialist argument for rule-consequentialism*.

It is a bad argument. First, it conflates the question of which theory is correct with the question of which theory is such that belief in it will produce the most good. Second, it *starts* from a consequentialist premise, namely the premise that the overarching point of morality is to maximize the good. An argument for a consequentialist theory should *not* start from a consequentialist premise, or from any other controversial premise.

[3] I do explain in my *Ideal Code, Real World*; in my 'Reply to Arneson and McIntyre', *Philosophical Issues* 15 (2005); and in B. Hooker and G. Fletcher, 'Variable versus Fixed-rate Rule-utilitarianism', *Philosophical Quarterly* 58 (2008).

https://doi.org/10.1017/S0953820807002786 Published online by Cambridge University Press

My argument for rule-consequentialism is not consequentialist. To be justified to us, moral theories *must* economically systematize (or, if no system is available, at least endorse) the moral convictions we have after careful reflection. Furthermore, moral theories *should* identify a fundamental principle that both (a) provides a unified explanation of why our more specific considered moral convictions are correct and (b) justifies them from an impartial point of view. Perhaps rule-consequentialism can explain why our more specific considered moral convictions are correct and provide some impartial justification for them. If rule-consequentialism can do this, then we have a 'reflective-equilibrium' argument for rule-consequentialism. This argument is *not* a *consequentialist* one.[4]

Properly formulated, rule-consequentialism does not collapse into extensional equivalence with act-consequentialism. Though compliance with rules is of course very important, there are costs and benefits associated with the internalization of rules that are not fully reflected in acts of compliance with those rules. In particular, there are mainly cognitive costs associated with internalizing more complicated sets of rules. And there are various costs associated with internalizing rules that require greater self-sacrifice. Here I am pointing to the costs of internalizing the rules apart from the costs of complying with them. Because of all these costs, the rules whose internalization has the greatest expected value are not ones extensionally equivalent with act-consequentialism. Rather, complying with the rules whose internalization has the greatest expected value will sometimes require doing what act-consequentialism condemns.

Now the objection comes: if rule-*consequentialism* contains an overarching commitment to maximize the good, then surely there is some incoherence in the theory's ever requiring acts that would not maximize the good. I defended rule-consequentialism against this objection by pointing out that it need not contain an overarching commitment to maximize the good.

Rule-consequentialist agents could have, as their fundamental moral motivation, a commitment to act in ways that are impartially defensible. And these agents might believe both that complying with impartially justified rules is impartially defensible and that rule-consequentialism is the best account of impartially justified rules. Agents with that fundamental moral motivation and these beliefs would be rule-consequentialist agents but would not have an overarching commitment to maximize the good.

[4] Systematization for moral convictions and a unifying impartial justification for them are metaethical desiderata, *not* values at the same level as consequentialist ones.

https://doi.org/10.1017/S0953820807002786 Published online by Cambridge University Press

Not only rule-consequentialist agents but also the theory itself need not have an overarching commitment to maximize the good. The theory is composed of the proposition that rules are to be assessed by their consequences plus the proposition that acts are to be assessed by those rules. Neither of these propositions contains an overarching commitment to maximize the good.

Card suggests my rule-consequentialism is internally inconsistent because, despite my claims, rule-consequentialism *does* contain an overarching commitment to maximize the good. I accept that a rule-consequentialist agent should be concerned about the consequences of acts, not just the consequences of rules. One of the rules an ideal code would contain is a rule requiring agents to prevent disasters when they can. I argue that this rule should be overridingly strong, though limited in scope so as not to require more than a certain amount of aggregate self-sacrifice over time. As Card notes (p. 249), this rule requires sensitivity to the consequences of the different acts one could do.

As well as that overriding rule about preventing disasters, I accept that rule-consequentialism supports a more general rule about doing good for others. Again, this rule needs to be limited so as not to require more than a certain amount of aggregate self-sacrifice over time. Furthermore, this rule isn't overriding and perhaps isn't even often salient. For if it were often salient, then we would be regularly trying to calculate what would be best for others. That would get us into some of the difficulties about lack of information, time, etc., mentioned earlier.

So I concede that the rule-consequentialist agent should have a standing commitment to *do good for others* up to some limit of aggregate self-sacrifice. And I concede that this agent should also have an overriding commitment to *prevent disasters* up to some limit of aggregate self-sacrifice. But I will now argue that these concessions are not fatal.

First, the overriding commitment to *prevent disasters* is *not* the same as an overriding commitment to *maximize the good* (maximizing the good requires far more than preventing disasters). Second, a *standing commitment to do good for others* need not be an *overarching commitment to maximize the good*. What I meant by 'overarching' was 'overriding'. The standing commitment to do good for others need not, and should not, be overriding.[5] Hence, though I agree with Card that rule-consequentialist agents should have an overriding commitment to

[5] A standing commitment to do good *for others* and an overarching commitment to maximize *the good* also differ in aim: the good includes the agent's own good as well as the good of others.

https://doi.org/10.1017/S0953820807002786 Published online by Cambridge University Press

prevent disasters and a standing commitment to do good for others, neither of these amount to an overriding commitment to maximize the good. None of his arguments show that rule-consequentialist agents must have an overriding commitment to maximize the good.

His arguments target not so much the rule-consequentialist agent as the theory itself. It would be wrong to suggest that rule-consequentialism pays no attention to the actual consequences of people's following rules. Card thinks this point pushes my theory into a corner. Rule-consequentialism, he thinks, 'has some form of commitment to maximizing the good at both the level of rules and acts, and hence has an overarching commitment to maximizing the good' (p. 248).

However, to attend to a consideration does *not* entail being committed to thinking that consideration *always* decisive. The most it entails is the idea that the consideration is decisive in *some* circumstances. Rule-consequentialism attends to the consequences of acts of following a code, but the theory need hold neither that this is the only thing to attend to, nor that it is the thing of overriding importance. The theory can hold that while consequences of acts are important, so are the broader consequences of the internalization of one code rather than another, and often these broader consequences are more important.

Card notes that I hold that throwing life-preservers to the desperate people in the water is required even if others on board equally well placed to throw life-preservers refuse to do so. He thinks what drives me to this view is act-consequentialist thinking (p. 250). In reply, I admit that, if we keep our focus sharply on any particular case considered in isolation, there is tremendous intuitive appeal in the principle that one is required to keep making sacrifices for others as long as the benefit to them is much bigger than the cost to oneself. But this admission stops short of the act-consequentialist view that one is required to keep making sacrifices for others as long as the benefit to them is even a little bigger than the cost to oneself.

As Card notes, I argued that rule-consequentialist assessment of different possible rules about aid would favour one that assesses sacrifice aggregatively, not iteratively. Such a rule would not require (though would of course permit) self-sacrifice when one has already made, or will otherwise make, significant personal aggregative sacrifice.[6] This limit on the amount of aggregate self-sacrifice morality can require for the sake of helping strangers seems counter-intuitively mean when we look only at an isolated case and ignore its place in a pattern of cases. On careful reflection, however, rejecting such a limit

[6] Obviously, this rule is vague. But any less vague rule about aid seems unlikely to apply plausibly to a wide enough range of situations.

https://doi.org/10.1017/S0953820807002786 Published online by Cambridge University Press

is even more counter-intuitive. Thus, on balance, the rule about aid that I argued rule-consequentialism would support seems intuitively plausible.[7]

B.W.Hooker@Rdg.ac.uk

[7] This article was written while I held a Research Readership from the British Academy of Arts and Social Sciences. I thank the British Academy, and Gustaf Arrhenius, Nir Eyal, Peter Kail, Tim Mulgan and John Skorupski, for comments.

https://doi.org/10.1017/S0953820807002786 Published online by Cambridge University Press