

## **Sophisticated Rule Consequentialism: Some Simple Objections**

**Richard Arneson**

The popularity of rule-consequentialism among philosophers has waxed and waned. Waned, mostly; at least lately. The idea that the morality that ought to claim allegiance is the ideal code of rules whose acceptance by everybody would bring about best consequences became the object of careful analysis about half a century ago, in the writings of J. J. C. Smart, John Rawls, David Lyons, Richard Brandt, Richard Hare, and others.<sup>1</sup> They considered utilitarian versions of rule consequentialism but discovered flaws in the view that attach to the wider consequentialist doctrine. In the eyes of many, the flaws were decisive.

Brad Hooker has produced brilliant work that unsettles this complacent consensus.<sup>2</sup> Over a period of several years he has produced a sustained and powerful defense of a version of rule consequentialism that does not obviously succumb to the criticisms that have been thought to render this doctrine a nonstarter. He acknowledges intellectual debts to Richard Brandt. But Hooker avoid certain excrescences in Brandt's efforts to conceive of morality as an ideal code of rules. Most notably, Hooker eschews Brandt's misguided attempt to derive some version of rule utilitarianism from an underlying commitment to some form of contractualism. Moreover, Hooker has worked to articulate a version of rule consequentialism in sufficient detail that one can see how the different parts of the doctrine hang together and how the best version of the

doctrine must confront many discrete choices of formulation and make the best choice at each of these many decision nodes.

Hooker rightly reminds us that if rule consequentialism is a candidate moral theory, it should be assessed as a moral theory, according to the standards appropriate for moral theories. According to Hooker, these standards are: a moral theory must develop pretheoretical intuitions about morality, must be internally consistent, must cohere with our considered moral judgments in ideal reflective equilibrium, must "identify a fundamental moral principle that both (a) explains why our more specific considered moral convictions are correct and (b) justifies them from an impartial point of view" (p. 4), and must provide guidance for dealing with controversial and unsettled issues.

Call a version of rule consequentialism that is not immediately vulnerable to three standard objections "sophisticated rule consequentialism." The standard objections are (1) rule consequentialism is guilty of rule worship, (2) the doctrine is utopian in a bad sense, and (3) rule consequentialism either collapses into act consequentialism or (if interpreted to avoid collapse), is manifestly implausible.<sup>3</sup> The rule worship objection is that whereas rule consequentialism purports to tie morality to the production of best consequences, the doctrine at crucial junctures advocates obedience to rules when doing so produces suboptimal consequences. The utopianism charge points to what is claimed to be an egregious class of rule worship cases. Rule consequentialism holds that one ought to obey the code that would lead to best consequences if everybody accepted it (or followed it),

but such a code appears to instruct agents to comply with this ideal code even when others are not accepting (or following) it. The collapse objection asserts that for any construal of rule utilitarianism according to which it appears to dictate conduct different from what act consequentialism would dictate, there must be an alternative candidate rule utilitarian code that eliminates the putative conflict with act consequentialism and must be judged superior from the rule consequentialist standpoint.<sup>4</sup> Or if not, so much the worse for rule consequentialism.

The three objections as usually stated all presuppose that rule consequentialism must have an overarching commitment to maximizing good consequences or embrace the aim of maximizing good consequences. Hooker denies this presupposition. He observes that according to his understanding of rule consequentialism, rules are to be assessed by the consequences of their general acceptance, acts are to be assessed by their conformity to the rules that are ideal according to this exercise, and the theory is to be assessed by its overall fit with our considered judgments in reflective equilibrium. His development of this line of thought is ingenious and tricky.

Hooker elaborates a sophisticated rule consequentialism and argues it is superior, as a moral theory, to its main rivals. In a nutshell, the claims are these: Sophisticated rule consequentialism beats act consequentialism because the latter delivers verdicts about what we ought to do that conflict with our carefully considered moral judgments. Sophisticated rule consequentialism beats

Ross-style intuitionist deontology because it does no worse than intuitionism at avoiding conflict with our considered moral judgments and provides a principle that explains and justifies our considered particular moral judgments whereas intuitionism in the end presents morality as a motley heap of unconnected judgments recommended only by the fact that we are inclined to endorse each of them taken separately. In passing we note also that sophisticated rule consequentialism proves itself superior to absolutist deontologies that assert that there are some things we absolutely must not do, whatever the consequences. These rival views offend many of our most strongly held moral convictions. Sophisticated rule consequentialism thus reveals itself to be better than the main going rivals in moral theory and at least entitled to further exploration and respectful consideration.

Hooker defends this version of rule consequentialism: "An act is wrong if it is forbidden by the code of rules whose internalization by the overwhelming majority of everyone everywhere in each new generation has maximum expected value in terms of well-being (with some priority for the worst off). The calculation of the code's expected value includes all costs of getting the code internalized. If in terms of expected value two or more codes are better than the rest but equal to one another, the one closest to conventional morality determines which acts are wrong." (p. 32)<sup>5</sup>

DISASTER AVOIDANCE

The sophisticated rule consequentialist understands the utopianism worry. The worry is that rule consequentialism identifies the rules that one ought here and now to obey with the rules that would produce the best reasonably expected consequences if it were the case that they were taught to and accepted by almost everyone. This means rule consequentialism identifies what it is morally right here and now to do with conformity to rules that would work out for the best in counterfactual circumstances. In the actual circumstances the agent faces, it may be the case that no one or hardly anyone is following these ideal rules, and the results of the agent's here and now conforming to them might be anything at all—good, bad, or ugly. The sophisticated rule consequentialist has sophisticated replies to the worry. Unfortunately, the replies do not succeed.

One suggested gambit is to insist that the ideal code of rules must contain a consequentialist escape clause that says: "Above all, avoid disaster." This says that if following the set of ideal rules (apart from the disaster avoidance rule itself) would reasonably be expected to lead to an avoidably disastrous outcome, one ought to choose a course of action that avoids the looming disaster. "Above all" indicates that this rule trumps other rules with which it conflicts, the ones that in the agent's circumstances are heading to disaster.

To see the inadequacy of this response to the utopianism worry, consider that it might turn out that in a great many decision problems faced by agents, following the ideal code of rules would result in near disaster or bad consequences in the neighborhood of a near disaster. This problem arises so far

as I can see, pretty much independently of how the rule consequentialist understands the vague idea of a "disaster."<sup>6</sup> Wherever one draws the line of disaster, the question will arise, what should be done when following the ideal code would give rise to bad consequences below the threshold of disaster. Sophisticated rule consequentialism with the disaster avoidance rule set in place must hold that the moral agent ought to soldier on and follow the ideal code and bring about expectably bad consequences. The act consequentialist will say that sophisticated rule consequentialism here reveals itself guilty of rule reverence, a paler version of the superstitious rule worship that critics claimed was the underlying normative motivation of simple rule consequentialisms. If following the ideal code of rules even when doing so leads to disaster is irrational and morally wrong, why shouldn't we agree that following the sophisticated ideal code of rules even when doing so leads to near disaster is also irrational and morally wrong?

The sophisticated rule consequentialist has further replies. These are supposed to block the conclusion of the argument of the previous paragraph.

One reply is that standing fast by the ideal code of rules even when doing so does not produce the best consequences in the circumstances accords with common sense moral judgment. Indeed, critics of act consequentialism have urged that consequentialism allows and even requires acting against important moral rules just on the bare ground that doing so would produce better consequences in the agent's actual circumstances. Common sense morality

takes a contrary position. According to common sense morality, the moral obligation to tell the truth, keep one's promises, and in general to conform to significant moral rules continues to hold and should constrain the conduct of the morally conscientious agent even when lying or breaking one's promise or the like would bring about somewhat more good than standing fast by the moral rules. If sophisticated rule consequentialism accords with common sense morality on this point, this is a point that supports rule consequentialism, not a stinging criticism of it.

This reply fails. The problem is that the sophisticated rule consequentialist position does not imply recommendations for conduct that coincide with the recommendations of common sense morality, so even if we have reason to accept the latter, that still leaves us with good and sufficient grounds to reject the former. Here I am not endorsing the position of common sense morality, which I shall suppose to be roughly equivalent to the intuitionist ethics of W. D. Ross. I am making the point that agreement with common sense morality on the point at issue does not generate reason to support rule consequentialism. This is so because sophisticated rule consequentialism with its disaster avoidance component tells us to obey the ideal code of rules in scenarios in which common sense morality would rebel from this conclusion.

Consider situations in which the ideal code of rules, or at least the portion of it that is in question on this occasion, is not in fact accepted by most people and not followed by most people. Consider a rule that would produce ideal

consequences if everybody or nearly everybody conformed their behavior to it, but would produce no good consequences otherwise. Here is a simple example: In war, soldiers fighting for a just cause ought to stand by their post when attacked, unless outnumbered by attacking enemy so that even stout defense would be futile. Suppose this rule, followed by nearly everybody, would produce ideal results. But the rule in fact is not internalized by the military forces fighting for a just cause in a particular war. The enemy have attacked and most of your fellow troops have run away. You can stand and fight, in conformity with the ideal rule, or you can run and live to fight another day. The consequences of conformity to the rule would not be disastrous, but would be decidedly negative. You will die and gain very little if anything for your side.<sup>7</sup> What should you do? Common sense morality, which holds that the obligation to obey hypothetically useful rules is sensitive to the actual degree to which others are complying here and now, surely says one should run and live to fight another day. Act consequentialism to its credit says the same. Rule consequentialism, even sophisticated rule consequentialism with the disaster avoidance proviso added, would have to hold that one ought to stand and fight and die. So much the worse for sophisticated rule consequentialism. Here it is revealed to be Quixotic in a bad sense.

#### RULE CONSEQUENTIALISM QUALIFIED TO DEATH?

Consider again the implications of rule consequentialism of the form we have been discussing. We are imagining its application to a scenario in which

the ideal code of rules is not actually accepted by most people in society, so the good consequences that would accrue if everyone accepted the ideal code are not going to be forthcoming in any event, whatever the agent does on the particular occasion of choice being examined. The agent we suppose could act in conformity with the ideal code, but in these nonideal conditions doing so would do no good. We further stipulate that if the agent were to follow the ideal code on this occasion, the results would be bad for her or bad for other people who would be affected, but not so bad as to cross the threshold of disaster that would trigger the disaster avoidance rule (for now we are agreeing that this rule is included in the ideal code). In the example as so far described, there is nothing to be said for the action the agent is contemplating except that it conforms to the ideal code of rules, the internalization of which by almost everyone everywhere would have consequences at least as good as the internalization of any other possible alternative code. There are alternative actions, not endorsed by the ideal code, that the agent could perform instead, any of which would do some good for the agent or for others, compared with results following the ideal code will produce.

Notice that sophisticated rule consequentialism might be able to handle scenarios in which, although nearly everyone accepts the code deemed ideal, some significant number of people, or even nearly everybody, actually violates a rule of the code on some (types of) occasions. An agent's acceptance of a code is compatible with failure actually to comply with its requirements sometimes.

Hence the ideal code can include rules that deal with this noncompliance. Also, consistent with the supposition that nearly everyone accepts the candidate code, it could still be the case that some people do not accept it, and again, the code could include rules designed to guide people's conduct in response to such nonacceptance.

The problematic cases are those in which many or most people do not in fact accept the ideal code, the nearly universal acceptance of which would produce best consequences, so the agent's conformity with the code in these circumstances will lead to suboptimal, perhaps significantly suboptimal, consequences.

Suppose the sophisticated rule consequentialist responds to the difficulty by adding higher-order rules. The revised formulation runs so: An act is wrong if it is forbidden by the code of rules whose general acceptance would produce best consequences, except that if this code of rules is not generally accepted, one should follow the code of rules the general acceptance of which would produce best consequences in that set of circumstances (i.e., the scenario in which the first-order code is not generally accepted). In principle one could add further iterations, yet higher-order rules to deal with nonacceptance at the previous level.

The idea of dealing with the problem of the implausibility of ideal code recommendations in situations of widespread nonacceptance by adding to the ideal code a second-order rule tailor-made for that case is immediately exposed

to difficulty. The difficulty is that formulating a rule for situations of nonacceptance that would have good consequences if it were accepted by nearly everyone everywhere is not what is wanted. We are worried about the bad consequences of following the ideal code in particular circumstances. What sort of code would be ideal if the code included a provision for this sort of situation and that revised code were accepted by nearly everyone everywhere is not to the point. To see this, notice that it might well be the case that the ideal rule for dealing with certain sorts of nonacceptance, battlefield desertion for example, would produce fine consequences if it were the case that nearly everyone everywhere accepted that ideal rule for this contingency, but not otherwise, and the agent's actual situation is otherwise. A simple example illustrates the point: It could be the case that the rule that specifies whether one should desert one's battlefield post when everyone else is deserting and no good consequences would result by one's lone steadfastness—the rule for this situation whose acceptance by nearly everyone everywhere would have best consequences, would have those consequences only in virtue of the hypothesized general acceptance. Suppose that when desertion is rife, if universally the last man who might desert steadfastly and bravely and hopelessly did stand fast by his post, that shining widespread example would so shame potential deserters and marginally loyal soldiers as to have large-scale positive consequences for troop morale and battlefield efficacy. So the ideal code rule for situations of mass desertion is: the last man left at the post must stand fast by it. Now we imagine

applying this rule in a battlefield situation, and I am the last man in my unit who might desert, all my mates having already done so. My standing fast, not actually being part of a widespread practice of last men everywhere holding up the flag and dying gloriously, will accomplish nothing except bring about my immediate death at the hands of the enemy. Nonetheless, looking to sophisticated rule consequentialism with rules for dealing with nonacceptance and noncompliance included in the ideal package, I would be instructed to stand fast and would be required morally to do that.

My conclusion is that the strategy of responding to the utopianism objection by presenting a nested series of ideal codes, with each level past the first postulating nonacceptance of the code at the previous level by most persons (and considering what rules would be ideal for that situation, if everybody accepted them), cannot succeed. Counterexamples persist. The question that the rule consequentialist frames is simply not the question that the person trying to decide what is the morally best course of action needs to decide. What one ought to do in a given situation depends on the consequences and qualities of the alternative acts one might perform. What an ideal code of rules would prescribe for this situation, a code that is ideal in the sense that best consequences would result if nearly everybody were to accept it, is just not the relevant question.

This point does not presume that any version of consequentialism is correct. Perhaps deontology is correct. Perhaps people have natural rights, and

it is worse to do what violates rights than to allow rights violations, and worse to bring about a rights violation as one's goal or the means to one's goal than to bring about a rights violation in a way that is not intended in that sense.

Perhaps the list of basic natural rights is fixed by intuition after ideal reflective scrutiny, reflective equilibrium at the ideal limit of deliberation. That may be.

What cannot be, what does not really make sense when one tries to work out its implications, is the idea that what one ought to do here and now is fixed by the answer to the question, asked about any candidate course of action, what if everybody did the same, or what if everybody were to internalize a code of rules that included a rule specifying that candidate course of action for this situation.

However, sophisticated rule consequentialism as elaborated by Hooker is not the same as the iterated rule consequentialism that we have been considering and rejecting. Hooker proposes two additional rules for inclusion in the ideal code that are designed to deal with the utopianism problem, the implausibility of the implications for conduct of simple rule consequentialism when other people are not accepting, or conforming to, the ideal code. One rule is "Avoid disaster!" and the second is roughly, "If acting on the otherwise ideal code would be unfair to oneself or others in situations of general nonacceptance of the ideal code, do what is fair."<sup>8</sup> I shall refer to these rules as disaster avoidance and fairness.

These rules so far as I can see jump the tracks and are not really rule consequentialist.<sup>9</sup> I allow as rule consequentialist the view I called "iterated rule

consequentialism.” This version adds further levels to the doctrine—rules (that would produce best consequences if nearly everybody everywhere were to accept them) for dealing with situations in which there is general nonacceptance of the ideal rules specified at the first level, similar rules for dealing with situations in which there is general nonacceptance of the ideal rules specified at the second level, and so on. I submit that the iteration strategy is not successful, because versions of the counterexamples that demonstrated the odd implausibility of the simple one-level doctrine can be reproduced at each higher level. The what-if character of rule consequentialism brings it about that the questions the doctrine is posing, at whatever level it is pitched, are not germane to the question, what action is singled out by moral reasons all things considered as what the agent should do in her actual circumstances.

Hooker’s version of rule consequentialism gives up the idea that in situations of nonacceptance of the first-best ideal code, the thing to do is to ask a restricted version of the question, what rules if accepted by nearly everybody everywhere would produce best consequences. Instead in situations of nonacceptance, one should behave fairly. Moreover, in any situation, whatever the actual level of acceptance, if following the otherwise ideal rules would lead to a disaster, one should adopt some other course of action that avoids disaster. This version of rule consequentialism is a hybrid or compromise. Consider just the fairness component. To reiterate, it says that when there is general

nonacceptance of the ideal code, and following it would be unfair to oneself or others, one should act fairly.

This hybrid is unstable. To see this, notice that however one elaborates the independent and free-standing norm of fairness that is doing the work here, the question arises, why is not the value of treating people including oneself fairly a value that should play a role in determining what is morally right and wrong, permissible and impermissible, across the board. If fairness is a nonconsequentialist value that matters morally, I don't see how its writ can be restricted to conditions of general nonacceptance. We are on the road to affirming some version of pluralist intuitionism not any version of rule consequentialism.

There are some indications in Hooker's text that he would reject the version of his proposal that I have been discussing, the one that takes his doctrine outside the ambit of rule consequentialism entirely. Instead he would say that the disaster avoidance and playing fair rules that he introduces are indeed singled out by asking the basic rule consequentialist question at the second level: What code of rules if accepted by nearly everybody everywhere would produce best consequences, given that we are asking for second-level rules that would produce best consequences for situations in which there is general non-compliance with the first-level ideal code. But this construal moves his position back to what I have called "iterated rule consequentialism" and renders it vulnerable to the objection that is decisive against that doctrine.

## SOPHISTICATED RULE CONSEQUENTIALISM ON HOME GROUND

The sophisticated rule consequentialist acknowledges that problems emerge for her preferred form of rule consequentialism in situations of partial or general nonacceptance. Perhaps with respect to these loose ends rule consequentialism should be considered a research project, a work in progress. But she is fully confident that rule consequentialism of some form provides the right answers to questions about how to live when what we are envisaging is the agent's choice of conduct in a world in which the ideal code of rules is accepted by almost everyone everywhere. Is this confidence justified?

The ideal code of rules that is to determine what is right and wrong in every situation is supposed to be a set of rules acceptance of which by nearly everyone everywhere would produce best consequences. This is the rule consequentialist proposal. The determination of what rules are ideal must take into account people's limited knowledge of relevant facts when they must decide what to do, their limited cognitive ability to use the information that is available and to discern the reasons for and against any given course of action they might consider and their strength, and their deeply entrenched tendency to partial motivation, to prefer self over others friends and kin over others, and more generally and those near and dear over strangers. Hence the rules must be reasonably simple and must accommodate people's partial motives. According to Hooker the ideal rules represent an optimal compromise between the consequences that would ensue, given that everyone accepts the rules, and the

costs of training people actually to accept the rules. He suggests that to calculate the latter cost, the relevant question for any candidate set of rules is what costs would accrue from training each new generation of people so that nearly everyone everywhere accepts these rules, starting from scratch in each generation.

Now we imagine that such an ideal code of rules is in place and society is humming along on this basis. To my mind familiar act consequentialist considerations undermine the idea that the ideal code of rules, established in a society, is the theoretical determiner of right and wrong. A rule, to be maximally expedient, must be reasonably simple and not overburdened with exception clauses. This means that circumstances can arise in which the agent will predictably forego accessible good consequences if she conforms to the ideal rules. She has made a promise, but as it turns out, an impartial calculation of best consequences obtainable by this agent in her actual circumstances would demonstrate that breaking the promise would produce better consequences overall than keeping it. Given that keeping the promise would not lead to disaster and hence trigger the disaster avoidance rule, rule consequentialism must dictate that the ideal rule should be followed despite the cost in good consequences foregone.

Hooker professes to be untroubled by such examples. He celebrates them, in fact. His thought is that common sense morality (think for concreteness of Ross-style intuitionism) holds that the duty to keep a promise overrides

modest increments of good consequences that could be gained by breaking it. The same goes for other common-sense moral rules. That sophisticated rule-consequentialism yields judgments about such cases that agree with common-sense judgments is to the credit, not the discredit, of this doctrine.

This response is misplaced. If one disagrees with the judgments yielded by act consequentialism, one is holding that the deontological distinctions in the ways that an individual's agency results in good or bad themselves directly affect the moral judgment concerning what act is morally right and ought to be done. But if the do/allow and intended/foreseen distinctions matter morally in and of themselves, then no version of consequentialism is correct. Acceptance of the claim that these deontological distinctions matter morally and must be incorporated in the correct formulations of fundamental moral principles strikes at the heart of rule consequentialist program.

The sophisticated rule consequentialist aims to cater to the deontological and common-sense moral judgments about cases without giving these deontological and common-sense elements any place in the statement of fundamental principles. The idea, openly espoused by Hooker, that we should tweak the detailed characterization of the rule consequentialist formula so that, in actual and likely circumstances, it tends to produce verdicts about what should be done in particular circumstances that cater to nonconsequentialist common-sense judgments strikes me as an unstable hybrid strategy. Either the deontological and other common-sense judgments that we are trying to

accommodate should be accepted, in which case they should be reflected directly into the formulation of fundamental principle, or they should be resisted not accepted, explained away as an understandable mistake to which we are prone. But in that latter case the deontological intuitions pose no bar to acceptance of straight old-fashioned act consequentialism.

But again, if rule consequentialism by the rigidity of its rules tends to recommend courses of action that tend to coincide with common-sense morality rather than with act consequentialism when acting against ordinary moral rules would be expedient, the coincidence seems superficial—just based on contingent facts. But if we are jiggling and prodding and tweaking the exact formulation of rule consequentialism to bring it about that it does agree with deontological judgments about cases in many situations, why not just openly admit the deontological and common-sense intuitionist elements directly into our moral principles?

I do not insist that everything in a moral theory must be packed into its first-level principles regulating conduct. A theory might contain an additional level of principles specifying norms of blameworthiness and praiseworthiness, and perhaps a foundational level containing basic principles that explain and justify the principles at other levels. A moral theory might have several tiers. My point is that the tiers must cohere. Considerations taken to be morally significant in themselves at one level should not be undercut or vanish entirely at other levels. Hooker's levels do not cohere. His first-order principles register

deontological convictions, and his formulation of the rule-consequentialist criterion is explicitly shaped by these considerations, but they do not register at all at the fundamental level, at which rules are just assessed by the outcomes that would ensue if they were generally accepted.

An example of a parameter whose value the sophisticated rule consequentialist can pick with an eye to increasing the degree to which rule consequentialism accords with non-consequentialist deontological intuitions about moral rightness and wrongness is the percentage of people who accept the ideal code. Rule consequentialism might hold that rightness and wrongness of acts is fixed by the ideal code of rules acceptance of which by everybody everywhere would produce best consequences. The requirement that everybody internalizes the ideal rules is too stringent. First, this specification would seem to abstract away from almost all nonconformity to moral rules and preclude having moral rules to deal with determined antisocial violators of moral rules. Second, the costs of training the tail of the distribution of sociability would become huge. If unavoidably by genetic endowment and unfortunate early socialization some small percentage of people in any society will grow up deeply averse to moral rules, the costs of programs of training and socialization to bring it about that literally everyone internalizes the ideal code will become prohibitive. Worse, these costs will exert pressure to trim the rules from the ideal code that the most antisocial types find most egregiously objectionable. So if one is trying to massage the formulation of rule consequentialism to render its verdicts about

what agents should do congruent in likely circumstances with Ross-style intuitionism about common-sense moral rules, one will want to adjust the percentage-of-people-internalizing-the-code downward, to prevent prohibitions on lying or promise keeping from dropping out of the code. At the other end, as one moves the percentage downward, one will want to avoid lowering the percentage to the point that including strong duties of altruism approaching the stringent act-consequentialist requirements will become cost-effective to train people to accept. Whether the balance of these two pressures can really produce a version of rule consequentialism that allows options not to do the act that would produce best consequences on each occasion of acting and requires constraints against consequentialist maximizing that violates cherished moral rules is an open question, ultimately an empirical question, I would suppose. But I question the rationale for this rigmarole. If one believes that consequentialism is wrong because it fails to accord with a morality of moral constraints and moral options, why not straightforwardly incorporate constraints and options directly into the formulation of fundamental moral principles?

The argument I have been making is hypothetical in form. If one believes that an adequate morality includes constraints and options, then one should accept a morality of constraints and options, rather than jiggle an indirect consequentialism so it ends up mimicking the morality of constraints and options.<sup>10</sup>

PUBLICITY

Act consequentialism countenances the possibility that the maximization of good consequences would be produced by promoting consequentialism for an elite and some simpler morality of religious commands or primitive rules for the masses. Since act consequentialism states a criterion of right and wrong action and does not purport to state a method or decision procedure that should actually guide each agent as she considers what course of action to adopt, act consequentialism cannot rule out the possibility that by its lights the criterion of correct moral action should be kept secret, reserved for the few (or even for none, since what would promote best consequences over the long run just might possibly be suppressing all knowledge of this criterion among humans). Hooker rejects the idea that the true morality should be esoteric, known only by a few. "Such paternalistic duplicity would be morally wrong, even if it would maximize the aggregate good" (p. 85), he writes. But how can a consequentialist of any stripe be sure about this? The rule consequentialist should be open-minded as to whether common-sense moral aversion to duplicity or paternalism or anything else is really justified at the end of the day.

What Hooker means is that by its very definition sophisticated rule consequentialism must be a public code of rules. The code of rules that would produce expected best consequences if it were accepted by nearly everyone everywhere must be a code that nearly everyone is trained to accept. The content of such a code cannot be a secret known only to a few. The content of the code must be publicly accessible, available to anyone who seeks to discover

it. As it were the value of publicity comes for free if you accept sophisticated rule consequentialism.

The commitment to a rule-consequentialist code carries with it a commitment to publicity. Also, if one works out what rules such a code would contain, manifestly rules against duplicity and against interfering in other people's private affairs by restricting their liberty for their own good would appear on the list. Paternalism and duplicity are ruled out at two levels of moral thought. So argues Hooker.

In response, it must be noted that these broad claims about rule consequentialism and publicity are incorrect. Rule consequentialism states a criterion of morally right and wrong action. The criterion asserts that what an agent morally ought to do, here and now, is set theoretically by a specified counterfactual scenario: What would be the expected consequences if one or another proposed moral code were accepted by nearly everyone everywhere. The code whose consequences would be best in the counterfactually specified circumstances determines what, one morally ought, here and now, to do. This is true quite independently of the further issue, whether one morally ought to try to make this ideal code a publicly recognized code in one's actual circumstances.<sup>11</sup> Suppose we find ourselves in a situation in which most people are not internalizing the rule-consequentialist ideal code of rules. Nothing in the notion of sophisticated rule consequentialism guarantees that in this situation the ideal code of rule specifies that we should now strive to make this code the

established dominant moral code, into which as many people as possible are socialized. Hooker's confidence that the sophisticated rule consequentialism he espouses is allergic to duplicity and paternalism and wedded to publicity is so far as I can see misplaced.

More important, publicity itself should not be endorsed as intrinsically morally valuable. That the correct moral code along with its intuitive rationale should be accessible to all people is usually a good means to promoting moral value. But in coherently describable possible situations, this will not be so. In those situations, we do not obtain the best accessible consequences if we insist on publicity. Let us follow Hooker and identify best consequences with maximal well-being weighted to give some priority to gains for the worse off. Consider then a simple example in which act consequentialism and sophisticated rule consequentialism as interpreted by Hooker will diverge in their recommendations regarding publicity. Social planners can bring about either of two scenarios. In one, consequentialism is broadcast as the public morality, and the badly off members of society are thereby rendered worse off in lifetime well being than they need be. Perhaps the establishment of consequentialism as the public morality in our circumstances brings it about that poor people suffer more violent crime than they otherwise would. The alternative scenario that can be achieved is a form of "government house consequentialism," with an esoteric consequentialist morality internalized by a few and a religion-based ethics internalized by the many. In this latter scenario the ignorance of the rational

basis of morality that the many suffer is a blight on their lives, but is more than offset by greater immunity from crime. Faced with this example, the act consequentialist unhesitatingly chooses to bring about the second scenario and gives no weight in and of itself to the fact that society fails to satisfy the putative publicity requirement. I do not see that the act consequentialist's resolutely instrumentalist view of publicity constitutes any sort of embarrassment for her position.

#### MORAL REASONS

Another way to approach what seems paradoxical and wrong-headed about sophisticated rule consequentialism is to note a basic feature of the reflective equilibrium or coherence methods in ethics that Hooker seems to endorse. In evaluating a candidate moral principle, it is not enough that the principle would recommend plausible courses of action, ones that chime in with our considered judgments, if it is applied to actual decision problems that we and others are likely to face, the world being as it is. An acceptable moral principle must pick out the course of action we would deem right after ideal critical scrutiny and pick it out for reasons that agree with the reasons that after critical scrutiny we deem to support or underpin the choice. For example, John Rawls argues that even if a clever philosopher could show that the act utilitarian principle would recommend policies concerning the upholding of freedom of speech and the suppression of human slavery in the actual and likely situations human societies are likely to face, there is still something profoundly wrong

about utilitarianism that emerges when we consider its application to hypothetical cases that involve our firmest considered moral judgments such as those condemning slavery and the suppression of free speech. Utilitarianism according to Rawls even if it delivers verdicts that accord with our considered convictions does not deliver the right verdicts for the right reasons, the reasons we would embrace after ideal scrutiny.<sup>12</sup>

By this same ideal reflective equilibrium or ideal coherence method sophisticated rule consequentialism can be shown to merit rejection. Consider possible future scenarios in which the world changes dramatically and in such a way that the content of the ideal code of rules is revealed to be quite different than we might have supposed by projecting familiar circumstances of earth as we know it into the indefinite future. Consider a simple example. A thousand years from now, biological science achieves a breakthrough. From then on, technology is available that allows education and socialization by direct and precise stimulation of the brain. This new technology drastically alters the costs of bringing it about that nearly everyone everywhere comes to accept various proposed candidate moral codes. We can in this altered world successfully inculcate in people far more sophisticated and complex and motivationally demanding codes than hitherto.

These possible distant future facts would have an enormous impact on the content of the rules that sophisticated rule consequentialism selects as ideal and thereby as theoretical determiners of what individuals morally ought to do in

particular situations here and now. If the future of earth were as imagined, the ideal rules would represent a compromise between the rules that work best for people now and the rules that would work best for people after the world-changing technological advance.

Our response to this imaginary scenario should be that what happens a thousand years from now (provided what happens is not affected by actions we take now) and what moral rules it would be useful to train people to accept a thousand years from now has no impact whatsoever on the answer to the question, what makes sense for someone to do, what is the morally right or permissible thing for someone to do, here and now. Sophisticated rule consequentialism cannot deliver this result, so is disqualified. Sophisticated rule consequentialism does not pick out as right the actions that we deem right after critical scrutiny for the reasons we find them right. Rule consequentialism makes the determination of what is right here and now hostage to contingencies concerning what rules would produce good consequences if internalized by future people in whatever circumstances those future people happen to face. Intuitively those contingencies do not seem to be determiners of right and wrong. What is revealed to be transparently problematic about rule consequentialism in the counterfactual scenario just considered is also, though less manifestly, problematic about rule consequentialism as it would operate in future scenarios that resemble our familiar world for the most part.

## CONCLUSION

Hooker is to be commended for developing a version of rule consequentialism that is artfully responsive to criticisms that have been made against prior rule consequentialisms. His sophisticated rule consequentialism states the idea in what may well be its most defensible form. In this essay I have argued that Hooker's sophisticated rule consequentialism is vulnerable to objections that are variants of the standard criticisms. If these arguments are sound, then perhaps Hooker's ultimate achievement will have been to show that the rule consequentialist research program should be abandoned. Act consequentialism remains standing and merits further development and comparison with alternative nonconsequentialist moralities that resonate strongly with some features of common sense moral opinions.

#### Notes

---

<sup>1</sup> . See Rawls (1955), Smart (1956), Brandt (1963, 1967, 1979), Lyons (1965), and Hare (1981).

<sup>2</sup> . (Hooker 2000), and the citations there to earlier work by Hooker. Further references to this work are given by page numbers enclosed by parentheses in the text.

<sup>3</sup> . Act consequentialism is the doctrine that one ought always to do an act that brings about an outcome no worse than the outcome that any other act one might have done instead would have brought about.

<sup>4</sup> . Following Brandt, Hooker formulates rule consequentialism in terms of rules the general acceptance of which would have best consequences, not rules the general conformity with which would have best consequences. In the utility of general

---

acceptance he counts the utility and disutility of training people in each new generation to accept the rules. The rules must then be simple and learnable. This move renders rule consequentialism not at all likely to yield implications for conduct that are the same as those of act consequentialism, so the collapse worry disappears. [On the collapse objection, see (Lyons, 1965).] In passing I note that the general conformity test would also pretty clearly emerge as a distinct alternative to act consequentialism if in the utility of general conformity one counted the utility and disutility of training people in each new generation to conform to the proposed rules.

<sup>5</sup>. Some of the details of this canonical formulation play no role in my discussion. In particular, I shall make nothing of the difference between saying an act is wrong if it would be forbidden by a code of rules of which the **expected** consequences would be best and saying that an act would be wrong if forbidden by a code of rules of which the **actual** consequences would be best. Sometimes in this discussion I write as though Hooker had opted for the latter formulation. So far as I can see this difference does not matter for any argument I press.

<sup>6</sup>. The lower the threshold of disaster is set, the less the divergence in practice between Hooker's sophisticated rule consequentialism and act consequentialism. At the limit, one might deem any act that produces worse consequences than another act that might have been done instead an act that leads to disaster. Hooker evidently does not want to take this path.

<sup>7</sup>. My description of this example invites the reply that death for the individual would qualify as a disaster and should trigger the disaster avoidance rule. If that reply is tempting, imagine that the balance of expected gains and losses from standing fast by the

---

generally useful rule in these circumstances would be bad, but just short of qualifying as disastrous. Suppose that standing fast by one's post will result in serious bodily injury for the resolute soldier and very tiny expected gain for any impacted military objective.

<sup>8</sup> . On disaster avoidance, see (Hooker, pp. 98-99). On fairness, see (Hooker, pp. 121-25).

<sup>9</sup> . It bears emphasis that the issue whether Hooker's position is properly labelled "rule consequentialism" or not is terminological and not substantive. If some development or modification of rule utilitarianism as traditionally conceived were reasonably deemed acceptable all things considered, the question whether we should call this doctrine "rule consequentialism" or "rule-consequentialism-with-an-asterisk" or something else would be unimportant.

<sup>10</sup> . Being an act consequentialist, I myself do not believe that an adequate morality does include constraints and options.

<sup>11</sup> . The issue turns on this question: Does the ideal code include a rule to the effect that one should strive to promulgate this and the other rules that constitute the code and secure its establishment as a publicly recognized morality? The rule on this point cannot be too demanding of agents, on pain of excessively raising the costs of training people into the code. It's not at all clear to me that the promulgation rule might not take this form: in some circumstances, passively accept whatever code is dominant in society; in other circumstances, strive to change it toward the ideal (when this can be done at modest cost); in still other circumstances, promulgate a vulgar morality and render the ideal code itself esoteric. The answer depends on empirical facts that are hard to discern.

---

<sup>12</sup> . To clarify: I do not accept Rawls's view that the considerations he mentions amount to a decisive objection against act utilitarianism (Arneson, 2000). I do accept the point of method he urges: A moral theory must not only recommend courses of action in actual or likely circumstances that will strike us as intuitively acceptable after ideal deliberation. A theory must recommend courses of conduct that strike us as intuitively right and recommend them for reasons that we accept after ideal reflection as the right reasons. Hence an acceptable moral theory must be counterfactually stable: in nonactual but possible scenarios the theory must yield implications for choice of conduct and policy that accord with considered judgments after ideal deliberation.

## References

Arneson, Richard. (2000) "Rawls versus Utilitarianism in the Light of Political Liberalism," in The Idea of a Political Liberalism, Victoria Davion and Clark Wolf, ed. (Lanham, MD: Rowman and Littlefield).

Brandt, Richard B. (1963) "Toward a Credible Form of Utilitarianism," in Morality and the Language of Conduct, H. N. Castaneda and G. Nahnnikian, ed. (Detroit: Wayne State University Press), pp. 107-43.

\_\_\_\_\_. (1976) "Some Merits of One Form of Rule-Utilitarianism," in University of Colorado Studies in Philosophy, pp. 111-36.

---

\_\_\_\_\_. (1979) A Theory of the Good and the Right (Oxford: Oxford University Press).

Hare, R.M. (1981) Moral Thinking: Its Levels, Method, and Point. (Oxford: Oxford University Press).

Lyons, David (1965). Forms and Limits of Utilitarianism. (Oxford: Oxford University Press).

Rawls, John. (1955) "Two Concepts of Rules," Philosophical Review, 64, 3-32.

Smart, J. J. C. (1956) "Extreme and Restricted utilitarianism," Philosophical Quarterly, 6, 344-54.