Rule Consequentialism

I. Act Consequentialism and Rule Consequentialism

Act consequentialism holds that the rightness of an act depends entirely on whether or not it maximizes the good.

Rule consequentialism holds that the rightness of an act depends not on the goodness of its consequences, but on whether or not it is in accordance with a certain code of rules, which has been selected for its good consequences.

II. They Differ with Respect to Their Criteria of Rightness, Not Necessarily with Respect to Their Decision Procedures

It is important to distinguish a *criterion of rightness* (a criterion that specifies what the fundamental right-making and wrong-making features of acts are) from a *decision procedure* (a procedure for deciding what to do).

The act consequentialist should reject the following decision procedure:

> **DP1** On every occasion, an agent should decide what to do by attempting to ascertain which of all the acts available to her would maximize the good.

There are good reasons to think that such a decision procedure would be counter-productive:

1. The costs of obtaining as much information as possible about the likely consequences of all of the available act alternatives would often exceed the benefits of doing so.
2. There is often insufficient information available to agents, such that even if it weren't costly to obtain as much information as possible about the likely consequences of all of the available act alternatives, it would still be better to follow some simple tried and tested rules.
3. Collectively, the expectation effects of everyone following some simple tried and tested rules is better than the expectation effects of everyone following DP1, for if everyone were to follow DP1, people would worry that they can't trust other—that they can't, for instance, trust people to keep their promises.

For these reasons, act consequentialists typically endorse both following and internalizing the following decision procedure:

DP2 On every occasion, agents should decide what to do by following certain tried and tested rules, such as "Don't harm others," "Don't steal," "Keep one's promises," "Tell the truth," etc. Included amongst these rules would be a rule saying that, when it is quite clear that following one of these other rules would have disastrously bad consequences, you should violate the given rule. These are the rules whose general acceptance would produce the best consequences.

Rule consequentialists also endorse DP2 as the correct decision procedure to follow. The act consequentialist differs from the rule consequentialist in thinking that every time an agent follows DP2 and yet fails to maximize the good, she acts wrongly. The rule consequentialist, by contrast, believes that it is often impermissible to break such rules even when doing so will maximize the good.

III. Formulating of Rule Consequentialism:

A. Compliance versus Internalization

COMPLIANCE: One complies with a code of rules if and only if one acts in accordance with that code of rules.

INTERNALIZATION: One has internalized a code of rules if and only if one both believes that the right thing to do is what is specified by the code and has the psychological dispositions and character traits that dispose one to do what is specified by the code, to feel guilty when one violates the code, and to resent it when others violate the code.

So we can formulate rule consequentialism in either of two ways:

THE COMPLIANCE FORMULATION: An act is morally permissible if and only if it is allowed by the code of rules, which if generally complied with, would produce the most good.

THE INTERNALIZATION FORMULATION: An act is morally permissible if and only it is allowed by the code of rules whose internalization by the vast majority would produce the most good.

B. Why Hooker Favors the Internalization Formulation

There are at least two reasons:

1. Internalization can have benefits that mere compliance doesn't have. To illustrate, suppose that you have internalized the rule that says "Retaliate

against an attacker" and that you are totally transparent in the sense that people know your dispositions just by looking at you. Your having internalized the rule will, then, have the benefit of deterring potential attackers without your ever having to retaliate against any attackers. This is a benefit that merely complying with the rule doesn't have.

2. It might be more costly to get one code of rules internalized than another, and we should not overlook this cost since, absent obtrusive governmental policing and sanctions, internalization is necessary to ensure general compliance. So one possible objection to a code might be that it is so complicated or so demanding that its internalization costs outweigh the benefits that would come from having people internalize it as opposed to some simpler or less demanding code. To illustrate, suppose that code X and code Y are identical except that whereas code X has a rule requiring people to give 20% of their incomes to charity, code Y has a rule requiring people to give 10% of their incomes to charity. Now let's suppose that the costs and benefits are as follows:

Code:	Internalization benefits:	Internalization costs:	Total:
X	10,000	-6,000	4,000
Y	6,000	-1,000	5,000

C. The Collapse Objection

Another reason for not formulating rule consequentialism in terms of compliance is the collapse objection.

Which of the following two codes of rules would produce the most good if generally complied with?

- CODE 1: "Don't kill," "Don't steal," "Keep your promises," etc.
- CODE 2: "Don't kill except when killing will maximize the good," "Don't steal except when stealing will maximize the good," "Keep your promises except when breaking them will maximize the good," etc.

The answer is clearly Code 2. Thus, we see that, if formulated in terms of compliance, rule consequentialism is extensionally equivalent to act consequentialism. This would be bad news for rule consequentialism. If both theories yield all the same moral verdicts, why opt for the more complicated of the two? Rule consequentialism would have all of act consequentialism's counterintuitive implications and be more complicated to boot.

Which is the code whose internalization by the vast majority would produce the most good?

To answer this question, we must consider the wider costs and benefits of having the code in question internalized. The good would not in fact be maximized by the internalization of Code 2, for three reasons. First, as we have already seen, act consequentialism (i.e., DP1) does not make for a very good decision procedure; internalization does not insure compliance. Second, if everyone internalized the rules stated in Code 2, awareness of this fact would eventually be widely recognized and, as a result, there would be a disastrous breakdown of trust. Third, the costs of getting such rules internalized would be extremely high, for people have a special concern for their own loved ones, and to get them to acquire dispositions to lie to, steal from, and kill their loved ones when doing so would maximize the good is going to be very costly.

D. Rule Consequentialism and the Issue of Non-Compliance: An Ideal Code for the Real World

Since internalization of the ideal code by 100 percent of the population is, practically speaking, impossible and doesn't guarantee 100 percent compliance, rule consequentialism must be formulated so that it selects as the ideal the code the one whose internalization by less than 100 percent of the population would produce the most good. In this way, the rule consequentialist can ensure that there will be rules pertaining to non-compliance. These will include: (1) rules about how to deter those who haven't internalized the ideal code or who might be tempted to violate the ideal code, (2) rules about punishing or rehabilitating violators of the ideal code, and (3) rules about what to do when others around you are, according to the ideal code, supposed to be doing their fair share in some collective enterprise but aren't.

If we were instead to formulate rule consequentialism so that it selects as the ideal the code the one which if internalized by 100 percent of the population would produce the most good, then the ideal code wouldn't necessarily include rules for dealing with noncompliance, for the addition of such rules would have some added internalization costs but little to no added benefit, since rules for dealing with noncompliance would have little utility in a world in which there is 100 percent internalization.

E. Hooker's Formulation

RULE CONSEQUENTIALISM: "An act is wrong if and only 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 a code's expected value includes all costs of getting the code internalized [in each new generation]. 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 what acts are wrong." (Hooker 2000, 32)

Included in Hooker's formulation is the following important footnote:

* "Assume that new generations are not changed genetically. If genetic engineering alters human genetic makeup, the codes that are best will probably be different" (Hooker 2000, 32).

We should also add some sort of determination clause—that is, a clause stating that what makes morally wrong acts morally wrong is that they are forbidden by the ideal code.

F. The Ideal Code will include a But-Above-All-Prevent-Disasters Rule

Since it is reasonable to suppose that things will go better if people are more concerned with preventing disasters than with following other rules, we should suppose that rule consequentialism will include a rule that says "Break one of the other rules if doing so is necessary to prevent disastrously bad consequences."

Just how bad the consequences need to be to count as disastrously bad will vary depending on what the rule in question is. For instance, what will count as sufficiently bad to warrant breaking the rule that dictates keeping one's promise to meet someone for lunch won't necessarily be sufficiently disastrous to warrant breaking the rule that dictates refraining from the commission of murder.

Hooker notes that "disaster prevention is an *obligation* when the disaster is to *others*, and a *permission* when the disaster is to oneself" (2005, 273). This is curious, though, for it would seem that a code of rules that included a rule that made disaster prevention an obligation whether the disaster was to oneself or to others would have greater expected utility than one that made it permissible not to prevent disasters to oneself. In making a distinction between preventing a disaster to oneself and to others, Hooker seems to be trying to accommodate the self-other asymmetry that we find in commonsense morality. But it seems that rule consequentialism is incompatible with this aspect of commonsense morality.

III. Why Accept Rule-Consequentialism Over Its Rivals?

A. Hooker's Criteria for Theory Selection

According to Hooker we should assess moral theories according to the following five criteria:

(1) Moral theories must start from attractive general beliefs about morality.

(2) Moral theories must be internally consistent.

(3) Moral theories must cohere with (i.e., economically systemize, or, if no system is available, at least endorse) the moral convictions that we have after careful reflection.

(4) Moral theories should identify a fundamental principle that both (a) explains why our more specific considered moral convictions are correct and (b) justifies them from an impartial point of view.

(5) Moral theories should help us deal with moral questions about which we are not confident, or do not agree. (Hooker 2000, 4)

Hooker thinks that, as does rule consequentialism, all of its leading rivals start from attractive general beliefs about morality. To illustrate, consider the following theories:

o Rule-consequentialism: Morality derives from the worry about what if everyone felt free to do that. Morality is a collective enterprise undertaken for the sake of collective benefit.

o Act-consequentialism: All that ultimately matters from the moral point of view is whether individuals are benefited or harmed. It can never be wrong to do what produces the most good.

o Contractualism: Morality consists of rules to which everyone would consent to in some original position.

Thus, Hooker thinks that rule consequentialism and its leading rivals are all on par with respect to (1). Hooker argues, though, that rule-consequentialism outshines it rivals with respect to one or more of the other four criteria. For instance, Hooker believes that rule consequentialism's implications about what is right or wrong in particular circumstances match our confident moral convictions better than act consequentialism's implications do. Thus, Hooker claims that rule consequentialism is superior to act consequentialism on criterion (3). And Hooker claims even if rule consequentialism doesn't fare any better than Rossian Pluralism on criterion (3), rule consequentialism outshines Rossian Pluralism on criterion (4).

B. Some Worries about Hooker's Criteria

First, Hooker's criteria leave out the following important desideratum: Moral theories should cohere with (i.e., economically systemize, or, if no system is available, at least endorse) the *non*-moral convictions that we have after careful reflection. For instance, our moral theories should cohere with our convictions about metaphysics and practical reasoning and rationality. To illustrate, consider (1) the divine command theory and the considered *non*-moral conviction that the quantity and quality of evil we find makes the existence of God improbably and (2) rule consequentialism and the considered *non*-moral convictions about practical reasons that led us to find act consequentialism compelling. Hooker admits that moral theories should be consistent with the best theories of science, metaphysics, rationality, etc., but he believes that "all the main moral theories *are* consistent with the most plausible systems of belief about other things" (Hooker 2000, 238). But this is questionable. Isn't the divine command theory one of the "main moral theories," and is it clear that it is consistent with the best metaphysics? And not all the main moral theories are compatible with all plausible systems of beliefs about practical reasoning. The teleological conception of practical reasons is a plausible theory about practical reasons, and it is not, on certain plausible assumptions about the relationship between morality and practical reasons, compatible with all the main moral theories. Hooker might admit as much, but he would respond as follows:

> The problem here is that when views about rationality are put forward as premises in arguments supporting one of these normative moral theories over its rivals, the views about rationality are controversial and even questionable. So they do not ground a *compelling* argument in favour of any one normative moral theory. (2000, 230)

Now I don't think that such a response is adequate. All philosophical theories are controversial, and so it's not surprising that all theories of rationality will be controversial. But this is compatible with the thought that some of these theories will be better than others, and some of these will cohere better with some moral theories than others will. Hooker should not just assume that all the main moral theories are consistent with the best theories of science, rationality, and metaphysics, for we have yet to work out what the best theories are. For instance, there is still a lot of interesting work going on in the areas action theory, personal identity, and practical reasoning, and this work is potentially relevant to the determination of what the best moral theory is. So we should include among our criteria the thought that, other things being equal, a moral theory is better if it coheres better with our beliefs about things such as science, metaphysics, and rationality. Hooker's criteria omit this important desideratum.

Second, Hooker's first criterion doesn't allow us to say that one theory is, other things being equal, superior to another if it starts from more attractive general beliefs about morality. Also, I'm not sure that a theory's starting from attractive general beliefs about morality should be a necessary condition as opposed to a desideratum.

Third, I don't understand why (3) is formulated as a necessary condition as opposed to a desideratum. I think that some theories will do a better job of cohering with our convictions (where this is conceived of as involving mutual support and not mere consistency) than other theories will. And I think that some theories will do a better job of systematizing our moral convictions than others will. So given that coherence comes in degrees, I don't understand the "must" in 3. Surely, it doesn't mean that moral theories must cohere with our other convictions to the maximum degree to which it is possible for a set of propositions to cohere.

Fourth, these criteria can point us in different directions, and yet Hooker's criteria for theory selection don't include some criterion for resolving conflicts between these various criteria. For instance, one theory may do a better job of systematizing our moral convictions and another theory may do a better job of justifying them from an impartial perspective. And one theory may do a better job of justifying our moral convictions from an impartial perspective and another theory may do a better job of helping us deal with moral questions about which we are not confident, or do not agree. Hooker's criteria for theory selection don't tell us how to select among theories with different strengths.

Fifth, I've always been unclear as to what it would be to justify some act, judgment, or principle from an impartial point of view, and why, if impartiality is distinct from universalizability, impartiality should be a requirement (or even desideratum) for a plausible moral theory. Why, for instance, is it not simply enough to be able to justify a given act, judgment, or principle? Suppose one could fully justify that act, judgment, or principle. Would this justification be lacking for the fact that the justification wasn't from an impartial point of view, whatever that entails? Perhaps, a requirement for any plausible moral theory is that it identifies some set of fundamental moral principles that allow us to justify our conduct to others, but why does that justification have stem from an impartial point of view?

For these reasons, I think that we do better by applying the following general method of inquiry to the task of theory selection in moral philosophy:

WIDE REFLECTIVE EQUILIBRIUM

1. Reflect on the interconnections among our beliefs.
2. Leave nothing out of our reflections. Reflect not only on our own initial beliefs and their interconnections, but also on ourselves and our competency to judge, and reflect on alternative beliefs, theories, and arguments.
3. Settle any conflicts that emerge, deciding what to believe on the basis of what seems to be the most intuitively plausible upon reflection. (DePaul 2006, 616)

IV. Rule Consequentialism and the Too Demanding Objection

The too demanding objection to traditional act consequentialism: (1) traditional act consequentialism requires huge sacrifices from us even in the most ordinary of circumstances and (2) traditional act consequentialism requires us to make sacrifices, big and small, even when the aggregate good would thereby be only slightly increased.

Now, at first, it may seem that rule-consequentialism can avoid this objection simply by pointing out that if everyone gave a little bit of time and money (say, only 1 hour and $100 per year), then it would be unnecessary, and in fact bad, to give more.

But consider the following case: Suppose that there are two people running to catch a plane and two kids drowning in a lake just off the runway, and assume that neither would have to miss her flight (or sacrifice her non-refundable ticket) if each pitched in. Nevertheless, suppose that one refuses to pitch in, and so the only way to save both kids is if the other person saves them both, missing her flight.

Recall that rule consequentialism is best formulated in terms of less than 100% internalization.

What should the rule be then? Perhaps, the rule might be "when you happen to be surrounded by others who are not helping, then prevent the disaster even if this involves doing more than you would have to if others would help." But now this seems to be very demanding, because we are confronted with such situations on an hourly basis.

HOOKER'S SUGGESTION: "[P]eople should help others in great need when they can do so at modest cost to themselves, cost being assessed aggregatively, not iteratively." More precisely, Hooker thinks that rule consequentialism would favor a rule that would require that the sum of the altruistic self-sacrifices that one makes

over the course of one's lifetime meet some significant but fairly modest threshold.[1] Because the cost to the agent is assessed aggregatively rather than iteratively, the rule does not require one to help another in great need whenever the cost of helping on that particular occasion is modest. Having to help others whenever doing so on that occasion involves modest cost could easily be very costly in the aggregate. Thus having a rule which required doing so would be very difficult and costly to internalize. Thus a code with rules that assess costs aggregatively as opposed to iteratively would be the ideal.

AN OBJECTION TO THIS RULE: Suppose that, having already met the threshold of self-sacrifice required by the above rule, you notice yet another person drowning in a lake at you rush to catch a flight and whom you can save only if you miss your flight and thereby forfeit the cost of your non-refundable ticket. Doesn't it seem counterintuitive to suppose that you can just let this person drown now that you've already filled your quota of altruism? "How can it ever be morally permissible to 'shut the gates of mercy on mankind'?" (Hooker 2000, 168).[2] Here's Hooker response to this objection, which he calls the "objection about meanness":

> [A]ny principle which allows you to say that you've done your part to help others and now want to work on other interests of yours can sound counterintuitively mean…. On the other hand, any principle which tells you that you must take *every* opportunity to come to the rescue will look counterintuitively demanding, once we realize the virtual infinity of opportunities you have to come to the rescue. I cannot see how to keep morality from being outrageously demanding without drawing some kind of line limiting your duty to help strangers. But any such line will be opposed by the objection about meanness. A compromise must be struck. Given the pressure to compromise, the rule about aiding others I think rule-consequentialism would endorse seems intuitively plausible. (2000, 168)

Given that the ideal code will have rules that assess costs aggregatively as opposed to iteratively, we can be confident that rule consequentialism will not be too demanding in a world in which rules requiring more than modest sacrifices in the aggregate have very high internalization costs.

But even if rule consequentialism isn't very demanding in a world where demanding rules have high internalization costs, it is very demanding in a world where the internalization costs associated with demanding rules is relatively small given future advances in bioengineering or other fields. What's worse, the actual world may be one in which the benefits of a code that includes some very demanding rules exceeds its internalization costs (the costs of having it internalized by each new generation) given future advances in bioengineering or other fields.

---

[1] He thinks that the ideal code will also include the following rule: "When necessary to save the world, or even just some significant proportion of humanity, one may be required to make an extreme – even the ultimate – self-sacrifice" (Hooker 1998, 32).

[2] Richard Muirhead independently raised this objection in class.

As Arneson points out, "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" (2005, 249).

In response to this objection, which Hooker acknowledges is a powerful one, he suggests the following revised version of his theory:

HOOKER'S REVISED VERSION OF RULE CONSEQUENTIALISM: "Moral wrongness is determined by the code of rules whose internalization by the overwhelming majority of everyone everywhere in each new generation (not including generations after any new development that significantly reduces the costs of internalizing more complex and demanding codes) has maximum expected value in terms of well-being with some priority for the worst off" (2005, 268-269).

I take it that, on this view, every time there is this kind of new development what it is right and wrong to do changes such that what it was right and wrong for the generation before the development to do differs from what it is right and wrong for the generation after the development to do. In correspondence, Hooker has confirmed (tentatively) that this is correct.

I'm not sure what to make of this theory—see V.B below.

V. Other Objections to Rule Consequentialism (RC)

A. Rule Consequentialism Is Incoherent

Some argue that that RC is guilty of rule worship and that this makes RC incoherent. The question is: "If the goodness of consequences is ultimately what matters, then why should we follow the code even in those instances where we know that violating the code would produce more good?" So some argue that RC has an overarching commitment to maximize the good (RC's ultimate goal is to maximize the good), but yet RC is also committed to the view that agents ought not to maximize the good when this involves violating the ideal code. Thus some claim that RC is incoherent in that it both embraces and eschews an overarching commitment to maximize the good.

Hooker's response is simple and, I believe, correct: RC does not have an overarching commitment to maximize the good. RC is as it is defined in II.E.; RC is not the claim in II.E. plus the claim that the ultimate goal is to maximize the good.

B. The Counterintuitive Contingencies Objection

On RC, the extent to which I'm, say, obligated to make sacrifices for sake of helping others depends on such contingencies as what human nature is like, how many people are in need of help, what sort of technological advances there have been in recent years, and what's going on in far off lands or even distant planets (assuming that humans will at some point colonize other planets). This is so because RC requires us to weigh the costs and benefits of having various alternative codes of rules internalized by the overwhelming majority of *everyone everywhere* in each new generation (*not including generations after any new development that significantly reduces the costs of internalizing more complex and demanding codes*).

As many of us know, millions of people on this planet are suffering for lack of potable water, basic healthcare, and adequate nutrition. And, as many of us also know, we (the well-to-do) could alleviate and/or prevent some of this suffering by making certain self-sacrifices, e.g., by donating some of our income to organizations such as Oxfam and UNICEF. Now suppose that we (the well-to-do) are wondering to what extent each of us is morally obligated to sacrifice his or her own welfare for the sake of doing more to alleviate such suffering. Could the answer to this question depend on the existence of beings on some distant planet, say, Zargon, over which we have not had, and cannot have, any influence? Suppose that there is nothing we can do to influence the lives of Zargonians in any way. We can neither harm nor benefit them; we cannot even have the slightest affect on their thoughts or experiences, for their planet is billions of light years away from ours and, consequently, beyond the reach of causal powers. We know about them only through the supernatural abilities of an oracle, who we know always tells the truth and who tells us everything about them. But although we know about them, they do not know about us, for we have no way to communicate with them, let alone affect their welfares. Given the lack of any influence that we have over their lives, how could their existence affect how much one of us is required to sacrifice for the sake of alleviating some of the suffering here on Earth? It seems absurd to suppose that it could. Yet this is precisely what rule consequentialism, as recently developed and defended by Brad Hooker, implies.

On Hooker version of rule consequentialism, an "act is wrong if and only 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)" (Hooker 2000, 32). In other words, we are morally required to act in accordance with the code of rules that has maximal total expected value; this is what's called the ideal code. The total expected value ('TEV' for short) of a code of rules is a function of two things: (1) the expected value it would have once it

has been internalized by the overwhelming majority of everyone everywhere in each new generation (call this the 'internalization benefits' or 'IB' for short) and (2) the expected costs of having that code internalized by the overwhelming majority of everyone everywhere in each new generation (call this the 'internalization costs' or 'IC' for short).

To see how, on rule consequentialism, the extent to which we are required to make self-sacrifices for the sake of aiding those fellow Earthlings who are in great need can depend on whether Zargonians exist, compare the following two possible worlds: W_1 and W_2. In W_1, only Earthlings exist. In W_2, both Earthlings and Zargonians exist. In each of these worlds, let us assume that rule consequentialism will favor one of following two codes. Code U is a rather undemanding code, requiring that the sum of the altruistic self-sacrifices that one makes over the course of one's lifetime meet some significant but fairly modest threshold. Code D, by contrast, is considerably more demanding, requiring that the sum of the altruistic self-sacrifices that one makes over the course of one's lifetime meet a threshold that is ten times greater than that on Code U. Assume that Zargon is ten times more populous than Earth and that the internalization costs of getting Zargonians to internalize some rule is ten times less than what is to get Earthlings to internalizing some rule; this is due to advances in genetic engineering on Zargon, which have allowed Zargonians to genetically hardwire certain codes into each new generation. Here is the breakdown of the internalization benefits and costs for each world:

W_1

Code	Internalization benefits (IB)	Internalization costs (IC)	Total expected value (TEV)
U	+2,000	−500	+1,500
D	+19,000	−18,000	+1,000

W_2

Code	IB (Earth)	IC (Earth)	IB (Zargon)	IC (Zargon)	TEV (Earth + Zargon)
U	+2,000	−500	+20,000	−500	+21,000
D	+19,000	−18,000	+190,000	−18,000	+173,000

So, in W_1, the ideal code is Code U, but, in W_2, the ideal code is Code D. Now what we well-to-do Earthlings are morally obligated to do for sake our less fortunate brethren depends on what the ideal code is. So, in W_2, Earthlings are required to sacrifice ten times more than what they would be required to sacrifice in W1, and the only difference between the two worlds is the fact that in W_2, but not W_1, the Zargonians exist. But why would what Earthlings are

morally required to do depend on whether Zargonians exist, when even if they did exist nothing we could do would have any affect on them?

See my "[Rule-Consequentialism and Irrelevant Others](#)" for more on this.

C. The One-Code-Doesn't-Fit-All Objection

Suppose that there is a small group of people who are much smarter and much more prone to performing self-sacrificing, altruistic acts than the overwhelming majority of people are. This group of people could, then, without undue cost be made to internalize a code of rules that would, due to its greater complexity and demandingness, produce more good than the ideal code of rules (as defined by Hooker) would. Nevertheless, RC implies that it is the ideal code that determines what it is right and wrong for them to do. But why should someone capable of following a complex code be forced to comply with a simple code when this would entail doing less good?

D. The Contrary-to-Reason Objection

Let a_1 be the act of sacrificing one's own life or the life of one's loved one (which, we'll assume, are both well worth living) for the sake of saving some stranger's life, where doing so would maximize the good, and assume that W_2 is the possible world in which the ideal code requires S to perform a_1.

1. If RC is true, then, in W_2, S is morally required to perform a_1. (From the definition of 'RC'.)
2. In W_2, S has decisive reason to refrain from performing a_1. (From intuition.)
3. In all possible worlds, if S has decisive reason to refrain from performing a_1, then S does not have sufficient reason to perform a_1. (From the definitions of 'decisive reason' and 'sufficient reason'.)
4. In all possible worlds, if S is morally required to perform x, then S has sufficient reason to perform x. (From moderate moral rationalism.)
5. Therefore, RC is false. (From 1-4.)