CrossMark

Utilitas (2021), **33**, 204–220
doi:10.1017/S0953820820000369

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

From Compliance, to Acceptance, to Teaching: On Relocating Rule Consequentialism's Stipulations

Timothy D. Miller

Lee University
Corresponding author. Email. tmiller@leeuniversity.edu

Abstract

Several recent formulations of Rule Consequentialism (RC) have broken with the consensus that RC should be formulated in terms of code *acceptance*, claiming instead that RC should focus on the consequences of codes' being *taught*. I begin this article with an examination of the standard case for acceptance formulations. In addition to depending on the mistaken assumption that *compliance* and *acceptance* formulations are the only options, the standard case claims advantages for acceptance formulations that, upon closer examination, favor teaching formulations. In the remainder of the article, I defend this new teaching-centered approach against some recent criticisms. I argue that preoccupation with the somewhat technical problem of identifying the best criterion for making choices under conditions of uncertainty has distracted rule consequentialists from paying more careful attention to the advantages and disadvantages that result from decisions concerning where they locate RC's stipulated assumptions within the theory.

Several recent formulations of Rule Consequentialism (RC) have broken with the widespread consensus that RC should be formulated in terms of the consequences of a code's *acceptance*, claiming instead that RC should focus on the consequences of a code's being *taught*. Proponents of these new theories have made some bold claims about the advantages they hold over acceptance formulations; for example, Dale Miller describes his Uniform-Moral-Education RC as "a sea change in how we formulate RC,"[1] and I have claimed that my own Calculated Rates RC provides a solution to the partial acceptance problem that has plagued acceptance formulations.[2] In contrast to the optimism of its proponents, Kevin Tobia has judged this new approach a "failure," albeit one that offers instructive lessons concerning the role of assumptions in RC's attempts to identify an Ideal Moral Code.[3] Of course, failure and success in moral theorizing are often matters of degree rather than all-or-nothing affairs. I contend that these new teaching formulations of RC hold far more promise than Tobia's

[1]D. Miller, Moral Education and Rule Consequentialism, *Philosophical Quarterly* (forthcoming).
[2]T. Miller, Solving Rule Consequentialism's Acceptance Rate Problem, *Utilitas*, 28 (2016), 41–53.
[3]K. Tobia, Rule-Consequentialism's Assumptions, *Utilitas* 30 (2018), 458–71 (p. 458).

© The Author(s), 2020. Published by Cambridge University Press

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

unqualified judgement of failure suggests. Furthermore, I will argue that there are more important lessons to be learned from reflection on the role of assumptions in formulations of RC than those Tobia emphasizes. Preoccupation with the somewhat technical problem of identifying the best criterion for making choices under conditions of uncertainty has distracted rule consequentialists from paying more careful attention to the advantages and disadvantages that result from decisions concerning where to locate RC's stipulated assumptions within the theory.

1. The standard case for acceptance formulations reconsidered

RC is typically understood as seeking to identify the Ideal Moral Code – that is, the code that would have better consequences than all of its rivals.[4] Of course, the very idea presupposes the possibility of comparing the consequences of various codes, and many of the debates among rule consequentialists have concerned the criterion by which such comparisons should be made. The fundamental source of difficulty is the fact that moral codes do not produce consequences, at least not in any direct, straightforward sense. Their effects are always indirect; for example, their consequences result from their being taught, internalized, and acted upon. Hence, when we ask what consequences we could expect[5] from a code, that question is really just an abbreviated gloss of some more complex question: what consequences could be expected to result given various assumptions concerning the context in which it is embedded? Following Kagan, I will refer to the set of assumptions used in comparing the values of codes as their *embedding conditions*.[6] Amidst the many disagreements over how to formulate RC, one point of general consensus has been that the embedding conditions should compare codes on the basis of the expected value of their being *accepted* or *internalized*. The explanation of this consensus has much to do, I suspect, with the way the set of available options has been framed. In *Ideal Code, Real World*, Hooker presents it as a choice between just two alternatives, compliance or acceptance, and offers what I will refer to as the *standard case* for acceptance.[7]

One element of that case is the observation that codes have consequences beyond those that result from the actions (and inactions) they require; hence, compliance formulations seem to miss some of a code's significant consequences. Hooker illustrates

[4]In order to account for the possibility that two or more codes might tie for best, formulations of RC often identify *an* ideal code as one that would have consequences *at least as good as* any of its rivals. I am glossing such issues for ease of exposition.

[5]The theory can be formulated in terms of either *actual* or *expected* consequences. Most formulations follow Hooker's lead and state the theory in terms of expected consequences. See B. Hooker, *Ideal Code, Real World* (New York: Oxford University Press, 2000), pp. 72–75.

[6]For a helpful introduction to the role of embedding conditions in formulations of RC, see S. Kagan, *Normative Ethics* (Boulder, CO: Westview Press, 1998), pp. 227–39.

[7]*Ideal Code, Real World*, pp. 75–80. Many of the points Hooker makes are also made in various places by Richard Brandt, who often refers to a moral code's "prevalence" or "currency" in a society. See, e.g., Brandt, Toward a Credible Form of Utilitarianism, in *Morality and the Language of Conduct*, ed. by H.-N. Castañeda and G. Nakhnikian (Detroit: Wayne State University Press, 1963), pp. 107–43; Some Merits of One Form of Rule-Utilitarianism, in *Morality, Utilitarianism, and Rights* (New York: Cambridge University Press, 1992), pp. 111–36; Fairness to Indirect Optimific Theories in Ethics, also in *Morality, Utilitarianism, and Rights*, pp. 137–57; Problems of Contemporary Utilitarianism: Real and Alleged, in *Ethical Theory in the Last Quarter of the Twentieth Century*, ed. by N. Bowie (Indianapolis: Hackett, 1983), pp. 81–105; and *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979).

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

this point by noting examples in which the fact that others are aware that one has internalized a rule helps ensure that occasions to comply with it are unlikely to arise in the first place.[8] For example, if it were widely known among students that I accept a rule of failing anyone who is caught submitting plagiarized papers in my courses (and that I am pretty adept at sniffing these things out), then I may have few occasions to act on this rule. Most of its benefits would result from the fact that my students' expectations concerning my response would deter them from plagiarizing in the first place. These benefits would result from my acceptance of the rule (and the fact that others were aware that I accepted it), not from my actually complying with it. Another way to see the significance of acceptance is to focus on the fact that when a code has been internalized, violations of its demands are typically accompanied by feelings of guilt. Such feelings can be rather costly, but the desire to avoid them is also one of the psychological mechanisms that help to ensure that acceptance of rules translates into compliance, and hence, if compliance with a rule is important enough, what we might refer to as a rule's "guilty conscience costs" can be outweighed by its benefits.[9] However, some moral codes contain pointless prohibitions. In some cases, acts of compliance or non-compliance with pointless prohibitions may have little effect on the amount of well-being in the world aside from the effects they have on the consciences of those who have internalized the rules. In fact, those who have internalized pointless prohibitions can feel painful guilt and anxiety from merely experiencing a desire to violate them, even if they never follow through with such actions.[10] Compliance formulations have no obvious means of accounting for such costs.

A second element of the standard case is the observation that acceptance formulations, unlike compliance formulations, help to ensure that RC avoids selecting excessively complicated or demanding rules. Highly complex rules might be expected to produce very good consequences *given the assumption that they are complied with*, but they if they are too complex for humans to learn and apply, they would fail to produce those benefits in the real world. Similarly, rules that demand high levels of individual sacrifice for the sake of some greater communal benefit may lead to wonderful outcomes *given the assumption that they are complied with*, but if actual persons would be unable or unwilling to internalize and comply with such demands, the hypothetical good outcomes would, once again, fail to translate into real-world benefits.

The standard case for acceptance seems compelling, and it has not, to my knowledge, been directly challenged in the literature. However, I believe it is flawed, in part because it starts from the mistaken assumption that acceptance and compliance formulations are the only alternatives. I noted above that codes can produce consequences indirectly as a result of their being *taught*, *internalized*, and *acted upon*. Compliance formulations focus on the benefits of codes being *acted upon* and acceptance formulations focus on the benefits of their being *internalized*; however, until recently, the third option of evaluating codes on the basis of the consequences of their being *taught* has been largely overlooked. Once this option comes into focus, some of the advantages that have been claimed for acceptance formulations can be seen to actually favor teaching formulations.

[8]Hooker makes this point using the example of accepting a rule that endorses retaliation if one is attacked. See *Ideal Code, Real World*, p. 77.

[9]Brandt, Some Merits of One Form of Rule Utilitarianism, p. 126; and *A Theory of the Good and the Right*, pp. 166–67.

[10]I am grateful to Dale Miller for calling this sort of example to my attention.

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

Consider again the second element in the standard case for acceptance formulations. The reason compliance formulations are thought to risk selecting overly complicated or demanding rules is that they evaluate codes in a way that simply builds the assumption of compliance into the embedding conditions without any concern for how compliance is achieved. That is, such formulations ask what would result if people *did successfully follow* the rules, but that way of posing the question prevents us from considering the prior question of whether people *could* or *would* successfully follow them in the first place. It seems to have gone unnoticed that if that is the source of the difficulty for compliance formulations, then acceptance formulations should be no better off. If instead of asking what would happen *if people complied with a code*, we ask what would happen *if people internalized it*, then we have merely shifted to evaluating codes in a way that builds the assumption that the codes are internalized into the embedding conditions, again without any concern for how internalization is achieved. That is, if we simply ask what would result if people *did successfully internalize* the rules, that way of posing the question prevents us from considering the prior questions of whether people *could* or *would* internalize those rules in the first place.[11] Hence, whether we assume that they are successfully complied with or that they are successfully internalized, RC faces the worry that it might select overly complex or demanding rules.

Acceptance formulations typically try to avoid this result by appealing to codes' internalization costs. For example, Hooker formulates RC as follows (emphasis added):

> RC: 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 . . . *The calculation of a code's expected value includes all costs of getting the code internalized.*[12]

However, it is easy to overlook the fact that the internalization costs mentioned in the italicized sentence are not automatically captured by acceptance formulations of RC. It can seem natural to read that sentence as merely an explication of a detail already implicit in Hooker's use of "internalization" in the first sentence, but that reading is mistaken. Acceptance formulations evaluate codes on the basis of an assumption that they *are internalized* and ask what good or bad consequences can be expected to result.[13] It might be helpful to think of these consequences as being situated

[11]One might contend that acceptance formulations would limit RC to the consideration of rules that it is *possible* for humans to internalize, and thus would place an upper limit on the complexity of the rules. However, even if that is right, it would not give acceptance formulations what they need. If RC is to avoid selecting rules that are too complex, it cannot merely ask whether a rule's internalization is a psychological *possibility*; rather, it must ask whether the effort needed to achieve its internalization would be *cost-effective*. I am grateful to Dale Miller for pressing me to clarify this point.

[12]*Ideal Code, Real World*, p. 32.

[13]Mulgan claims that "Hooker's focus on the cost of inculcation clearly suggests that we are evaluating *our attempts* to inculcate potential codes, rather than assuming that inculcation will succeed." See T. Mulgan, *Future People* (New York: Oxford University Press, 2006), p. 146, fn. 26. However, Hooker explicitly separates the calculation of a code's value on the assumption that it has been successfully internalized and the calculation of the costs that would be involved in getting it internalized as two distinct, and apparently independent, parts of the process of calculating a code's overall value (*Ideal Code, Real World*, p. 79). Moreover, Mulgan's teaching-centered interpretation of Hooker's theory cannot explain why Hooker goes on to ask what rate of acceptance we should assume to have been achieved in our evaluation of a code's value, since, as Mulgan himself observes, that question does not arise on teaching-centered formulations

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

downstream from acceptance. By contrast, the appeal to internalization costs is an upstream consideration from the vantage point of the assumption that a code is accepted; it enquires into the costs that would have been involved in getting to the point at which the code has been internalized. The italicized sentence does not merely clarify what it is involved in the assumption that a code is internalized; rather, it introduces an additional consideration to factor into the evaluation of codes.

Two important implications of this point are worth noting. First, if it is appropriate for acceptance formulations to look upstream from acceptance in order to capture the costs that help them avoid overly complicated and demanding rules, then the same move ought to be available to compliance formulations as well. They could not only consider the downstream consequences of compliance, but also the upstream costs that would be involved in achieving compliance. Since internalization is the most important way in which compliance with codes is achieved, internalization costs would be captured as a subset of this broader category of *compliance costs*.[14] Just as acceptance theorists argue that highly complex or demanding rules would be avoided because the internalization costs would be prohibitively high, a compliance theorist could argue that such rules would be avoided because their compliance costs would be prohibitively high. In other words, contrary to the standard case, it is not really the shift from compliance to acceptance formulations that enables RC to avoid selecting overly complicated and demanding rules, but rather the contrivance of including upstream costs. Since that contrivance is equally available to both acceptance and compliance formulations, the second element of the standard case does not obviously favor acceptance formulations after all.

The second point to notice is that by appealing to internalization costs, acceptance formulations essentially attempt to evaluate codes both on the assumption that they *are accepted* and on the assumption that we are *trying to get them accepted*, or put more simply, that they are *taught*. If some of the advantages that have been claimed for acceptance formulations are in fact achieved by the contrivance of looking upstream from acceptance and evaluating codes from the vantage point of the assumption that they are taught, then it seems worthwhile to consider whether acceptance really is the best focal point for evaluating codes. Why not simply shift the vantage point further upstream and evaluate codes on the assumption that they are taught? Since the goal of teaching a code is to get it internalized and acted upon, teaching-centered accounts of RC's embedding conditions would offer a broader framework that could incorporate the strengths of compliance and acceptance formulations, while more naturally resolving some of their difficulties, including the concern that they might select overly complicated or demanding rules.

(*Future People*, p. 147). Mulgan's misinterpretation of Hooker's theory on this point leads him to underestimate the extent of his own departure from Hooker; although he claims to "follow Hooker in formulating Rule Consequentialism in terms of acceptance (or internalization) rather than compliance" (*Future People*, p. 141), he in fact formulates his theory in terms of a teaching-centered approach to the embedding conditions.

[14]Sometimes enforcement methods can lead people to comply with rules in spite of the fact that they do not accept them, or in spite of the fact that they accept them but are tempted to violate them anyway. Compliance costs would include both the teaching costs needed to achieve compliance through acceptance and the enforcement costs that may be needed to motivate compliance among the non-acceptant and weak-willed. See T. Miller, Solving Rule Consequentialism's Acceptance Rate Problem, p. 51.

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

2. Recent teaching formulations of RC

Several recent contributions to the literature on RC have begun to develop the idea of recasting RC in terms of teaching-centered embedding conditions.[15] Tim Mulgan presents his Ideal Moral Outlook Utilitarianism (IMOU) as a version of RC, although he rejects the usual terminology of rules and codes in favor of the more general idea of a moral outlook because it "leaves open whether the utilitarian ideal is a code of rules, a set of dispositions, a package of virtues, a set of priorities, a general moral outlook or (as seems most likely) some combination of these."[16] Terminology aside, what matters for my present purpose is the criterion by which Mulgan evaluates moral outlooks. He writes, "IMOU ranks competing moral outlooks by asking what would happen if we (the present generation) attempted to teach that outlook to the next generation. The ideal outlook is the best one for us to teach to them."[17] Two other recent formulations of RC have adopted a similar approach with a more explicit emphasis on the ways our approaches to teaching and enforcing a code would inform our evaluations. I articulate the idea in my Calculated Rates RC (CRRC) by introducing the concept of an *expanded code*, which is the result of conjoining a moral code with an account of the teaching and enforcement methods that would be used to achieve internalization of and compliance with the code. CRRC's characterization of the ideal code is:

> (CRRC): An ideal code is a code included in an expanded code whose value is at least as high as the value of any alternative expanded code.[18]

Similarly, Dale Miller's Uniform Moral Education RC (UMERC) begins with the notion of a *system of moral education*, understood as "as a combination of teaching techniques used by one generation to instil moral rules into another."[19] Such systems can be uniform if they attempt to instill a single code into all members of the new generation, or non-uniform if they attempt to instill different codes within different sub-groups or individuals.[20] UMERC characterizes the ideal code as follows:

[15]Whether Mill's utilitarianism is best understood as a version of act utilitarianism, rule utilitarianism, or some other type has been the subject of much debate; I am grateful to an anonymous reviewer for calling my attention to the fact that some interpretations of Mill bear noticeable similarities to the teaching formulations of RC discussed below. For example, although he denies that Mill's view can be neatly characterized as either act or rule utilitarian, Gray's interpretation is certainly closer to the latter, and its emphasis on "the inculcation of sentiments and attitudes and the instilling of dispositions and inclinations" invites comparison with the teaching-centered approaches discussed below; J. Gray, *Mill on Liberty: A Defence* (New York: Routledge, 1996), pp. 31–32. See also, D. Miller, *John Stuart Mill: Moral, Social, and Political Thought* (Malden, MA: Polity Press, 2010), pp. 88–101.

[16]T. Mulgan, How Should Utilitarians Think About the Future?, *Canadian Journal of Philosophy*, 47 (2017), 290–312 (p. 291).

[17]Mulgan, How Should Utilitarians Think About the Future?, p. 291; IMOU pulls together various ideas Mulgan has defended previously, e.g., in *Future People* and in *The Demands of Consequentialism* (Oxford: Clarendon Press, 2001).

[18]T. Miller, Solving Rule Consequentialism's Acceptance Rate Problem, p. 50.

[19]D. Miller, Moral Education and Rule Consequentialism.

[20]The uniformity in question is only to be understood as uniformity within whatever group the code is intended to be taught to. Later in his essay, Miller discusses the question of a code's scope and distinguishes between those theories that aim to identify a single universal code that applies to everyone everywhere and those that aim to identify an optimal code for a particular social group; he indicates his preference for the latter sort of theory.

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

(UMERC): the ideal code is the code that would be the subject of the optimal uniform system of moral education.[21]

Whatever subtle differences there may be between CRRC and UMERC, what is noteworthy for present purposes is that through their notions of *expanded codes* and *systems of moral education* they both characterize the embedding conditions by focusing on how codes would be taught and reinforced.

3. A response to Tobia's criticisms of CRRC

I and the other proponents of these teaching-centered approaches have claimed that their advantages include the promise of providing solutions to problems that have plagued the acceptance formulations that preceded them, including the problem of partial acceptance. However, Tobia has criticized the new teaching-centered approach as I articulate it in CRRC, claiming that it "merely relocates"[22] and "merely masks"[23] the partial acceptance problem, and that instead of offering an alternative to the previous theories it aims to replace, it ultimately ends up falling back upon one or another of those theories. If these criticisms are all correct, then the advantages claimed for CRRC would seem to be overhyped, and the teaching-centered approach to RC might be little more than a disappointing dead end. However, I will argue that Tobia's judgements concerning the failure of CRRC rely on too narrow an understanding of the issues at stake among competing formulations of RC. His discussion emphasizes the role of criteria for comparisons when outcomes are uncertain; however, I will argue that the more fundamental issue at stake between teaching-focused and acceptance-focused formulations of RC concerns the structure of the theories. Tobia's criticism that CRRC "merely relocates" the problem of partial acceptance misses the crucial point that where this and other problems are located in relation to the theory's assumptions can result in important theoretical advantages or disadvantages. I will argue that the disadvantages of locating these problems where acceptance formulations place them are severe. Teaching-centered formulations like CRRC may not make the problems disappear, but I hope to show that they do at least locate them within the theory's structure in a way that renders them more manageable.

3.1 The problem of partial acceptance

In my discussion of compliance and acceptance formulations above, I ignored an internal problem that arises within each of those approaches. Due to the consensus in favor of acceptance formulations, the problem is typically referred to as the "problem of partial acceptance", although a parallel "problem of partial compliance" arises for compliance formulations. In short, the basic problem is that any calculation of the expected value of a code based upon the assumption that the code is *accepted* (or *complied* with) must make further assumptions concerning what rate or level of acceptance (compliance) the code has. Since the ideal goal would be universal acceptance (compliance), it might be tempting to think that we should evaluate codes based upon that ideal. However, the assumption of universal acceptance (compliance) runs into well-

[21]D. Miller, Moral Education and Rule Consequentialism.
[22]Tobia, Rule-Consequentialism's Assumptions, p. 458.
[23]Tobia, Rule-Consequentialism's Assumptions, p. 463.

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

known difficulties, the most important of which is that it would select a code that offered no guidance for how to deal with less than ideal circumstances.[24] As Hooker notes, "An adequate ethic must provide for situations created by people who are malevolent, dishonest, unfair, or simply misguided. In short, for use in the real world, a moral code needs provisions for dealing with non-compliance."[25]

Most rule consequentialists have followed Hooker's suggestion that we should evaluate codes under assumptions of less than universal acceptance, although they have not been able to reach any sort of agreement concerning how that is to be done. A growing list of alternative formulations of RC are all differentiated by the ways they factor acceptance rates into their criterion for selecting an optimal code. Fixed Rate RC evaluates codes by stipulating a specific acceptance rate; for example, Hooker proposes a rate of 90%, judging it to be both high enough to capture the idea that the code is meant for everyone, and low enough to ensure that the code would account for the difficulties posed by non-compliance.[26] However, Fixed Rate RC has been commonly criticized for the arbitrariness of its stipulated rate; several alternative formulations have attempted to avoid this arbitrariness by evaluating codes at every possible rate of acceptance, although they disagree over their criterion for comparing codes. Optimum Rate RC[27] identifies the optimum code as the code with the highest potential value, comparing competing codes to each other at whatever possible acceptance rates maximize each code's potential. Variable Rate RC[28] identifies the optimal code as the one that has the highest average value across all acceptance rates, and Maximizing Expectation Rate RC[29] identifies it as the code with the highest *weighted* average value across all acceptance rates, with each possible acceptance rate weighted according to its likelihood of obtaining. Each of these formulations faces important difficulties, although I will not attempt to summarize all of the criticisms here.[30] I trust that enough has been said to help clarify the nature of the problem that CRRC claims to solve.

The key to CRRC's "solution" lies in the fact that it shifts the focus of RC's account of the embedding conditions from the assumption that codes are accepted to the

[24]Teemu Toppinen has recently attempted to show that universal acceptance formulations have the resources to overcome this difficulty if they include "robustly demanding goods" in their value theory. T. Toppinen, Rule Consequentialism (and Kantian Contractualism) at Top Rates, *Philosophical Quarterly*, 66 (2016), 122–35.

[25]Hooker, *Ideal Code, Real World*, p. 80.

[26]Hooker, *Ideal Code, Real World*, pp. 80–85. Brandt also sometimes mentions 90% as a targeted acceptance rate; see, e.g., Brandt, Some Merits of One Form of Rule Utilitarianism, p. 120; and *A Theory of the Good and the Right*, pp. 192–93, although there Brandt explicitly cautions that "[i]t would be a mistake to press too hard for an exact answer to the question just what a rational person is aiming at when he supports the 'currency' of a certain moral code."

[27]H. Smith, Measuring the Consequences of Rules, *Utilitas*, 22 (2010), 413–33 (p. 418). This and several of the other theories discussed in this paragraph are formulated as versions of rule utilitarianism rather than rule consequentialism. Since the differences between rule utilitarianism and rule consequentialism are not relevant to the issues discussed here, I will refer to them all as versions of rule consequentialism.

[28]M. Ridge, Introducing Variable-Rate Rule Utilitarianism, *Philosophical Quarterly*, 56 (2006), 242–53 (pp. 244–6).

[29]K. Tobia, Rule Consequentialism and the Problem of Partial Acceptance, *Ethical Theory and Moral Practice*, 16 (2013), 643–52 (p. 645).

[30]For such criticisms, see the sources from the prior three notes, as well as B. Hooker and G. Fletcher, Variable *Versus* Fixed Rate Rule Utilitarianism, *Philosophical Quarterly*, 58 (2008), 344–52; T. Miller, Solving Rule Consequentialism's Acceptance Rate Problem, pp. 42–48; and D. Miller, Moral Education and Rule Consequentialism.

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

assumption that they are taught. Any account of RC has to stipulate inputs for some of its assumptions in order make predictions about (i.e. "solve for") others. Those which evaluate codes on the assumption that they are accepted have to make stipulations about the levels of acceptance they will receive in order to solve for other variables such as the internalization costs that would be required to achieve those levels of acceptance and the benefits that could be expected to result downstream from the acceptance. Although it might be tempting to think that only Fixed Rate RC is committed to any such stipulation, that would be a mistake. Fixed Rate RC somewhat arbitrarily chooses to make the stipulation once (typically at a rate of 90%); by contrast, the other acceptance formulations make the stipulation repeatedly for every possible rate of acceptance. That is, like the fixed rate theory, they must stipulate a 90% acceptance rate to solve for the costs and the benefits that could be expected to obtain at that rate; they must also stipulate an 80% acceptance rate in order to solve for the costs and benefits that could be expected at that rate, and so on for all other possible acceptances rates. Rather than avoiding making such stipulations, the Optimum, Variable, and Maximizing Expectation Rate formulations greatly multiply them.

By reframing the embedding conditions, teaching-centered formulations like CRRC change which variables in RC's calculations are stipulated and which are "solved for" on the basis of those stipulations. Stipulating the methods of teaching and enforcement that would be used to get codes internalized has the effect of shifting acceptance rates from the set of variables stipulated as inputs to the calculations to the set of variables that are calculated as outputs. However, Tobia has objected that this "merely relocates" the partial acceptance problem, and that instead of solving the problem it "merely masks" it. He rightly points out that stipulating methods of teaching and reinforcement will not determine a unique acceptance rate; instead, we will have to deal with a range of possible acceptance rates that might result from those methods, and that, he argues, will leave us with the same old acceptance rate problem we began with. According to Tobia,

> this reveals that CRCC [sic] is not a true competitor to these other theories. It is ultimately dependent on one of these other views, arguing that we should choose the Code x Method combination that offers the best value given an acceptance rate calculation – value ultimately calculated in terms of Variable or Optimum or Maximizing Expectation or some other theory.[31]

However, note that in order assert that CRRC is "not a true competitor" for the other theories, Tobia must assume that what is most essential to the various theories are their criteria for comparison (whether they aim for an optimized expected utility, a highest average expected utility, a highest weighted average expected utility, etc.); his criticism presupposes that the structure of the theories – which values are stipulated as inputs to the calculations and which are calculated as outputs – is a matter of no real importance. However, I will argue that those structural differences are of primary importance; the significant advantages offered by teaching formulations like CRRC are revealed when we take notice of the way they alter the order of RC's calculations.

This brings me to Tobia's second criticism of CRRC, which attempts to "problematize" its theoretical structure. He asks:

> Is it really plausible that we can stipulate what methods of training and enforcement will obtain? In the real world, this is a question influenced by various

[31]Rule-Consequentialism's Assumptions, p. 464.

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

> political processes and complex social norm interactions. The far more plausible assumption is that we can estimate the likelihood of certain methods obtaining, given our intention for them to obtain. Similar debates to the partial acceptance debates arise here: the various positions include endorsing the code that does best on average across various methods, the code that does the best at one possible method, or the code that does the best as a weighted average.[32]

A first line of response to this line of questioning might be that it is not at all obvious that CRRC does claim that we can stipulate what method of training and enforcement *will obtain*. CRRC says that we should evaluate codes by asking what the consequences would be *given the assumption that they were taught and enforced in certain ways*. It need not make any implausible claims to the effect that our stipulations for the theoretical purpose of code comparison are guaranteed to obtain in the real world.

Still, it might be objected that our comparisons of codes' values are of limited practical use if they ignore the likelihoods of those codes and teaching methods actually obtaining. I am sympathetic to this concern, and it brings me to a second line of response. Questions concerning the likelihoods of various codes and teaching methods obtaining in the real world are bound up with the broader question of how RC should approach internalization costs. This is an important issue that has not received the amount of attention it deserves; however, since it is too large an issue to fully address here, I will limit myself to two general points. First, the issues Tobia presents as problematizing the structure of CRRC are not at all unique to CRRC; they are issues that all formulations of RC will have to address in some way or other. In order to count them against CRRC, Tobia would need to show that acceptance formulations offer a better framework for addressing them than teaching formulations like CRRC can offer. That brings me to my second point: because of the structural differences between acceptance and teaching formulations, they do locate this and other problems in different stages of their calculations of a code's value. However, a closer examination of these structural differences will reveal that it is the teaching formulations that are better situated to address many of RC's problems.

3.2 The structural disadvantages of acceptance formulations

I will begin by considering the ways these issues arise within the structure of acceptance formulations. As we noted earlier, acceptance formulations are forced to appeal to the contrivance of looking both downstream and upstream from their stipulated acceptance rates in order to take into consideration all of the costs and benefits associated with a given code. On the face of it, the idea seems straightforward enough. Suppose we wish to know what the value of a code would be if, let's say, 80% of a society accepted it. First, we look downstream from that point. As a rough gloss, we will need to add up the benefits that would result from the good behavior of the 80%, and subtract from that the costs that would result from the bad behavior of the other 20%.[33] Finally, we turn our attention upstream to consider the costs that would have been involved in achieving that 80% rate. Sum it all up, and we've calculated the expected value of the code at 80%.

In actuality, each step in this process is thoroughly problematic. Begin with what appears to be the simplest of the steps. Although it may seem like a stipulated 80%

[32]Rule-Consequentialism's Assumptions, p. 464.

[33]This is a merely a rough gloss for various reasons, including the fact that it ignores the added complications of accounting for the benefits of acceptance that go beyond those of compliance.

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

acceptance rate would give us a fairly solid handle on the good results that we can expect from the 80%, things are not quite as straightforward as they initially appear. Acceptance of a moral rule is no sure guarantee of compliance with it, since people fairly regularly do things that they acknowledge to be wrong. If we wish to know what good results we can expect from the 80%, we would need to know the extent to which their acceptance of the code translates into compliance. Of course, acceptance theorists could always grant this fact and respond by adding an additional layer of their favorite method of conducting comparisons in conditions of uncertainty. For example, as a proponent of the maximizing expectation criterion, Tobia might wish to calculated the weighted average value of all of the ways an 80% acceptance rate could play out. The calculations we are imagining would be wildly complex, but since complexity bedevils all versions of consequentialism, let us leave that concern to the side. The more fundamental difficulty with attempting that move at this juncture is that without knowing something about the upstream considerations there would not even be any determinate likelihoods to attach to the various downstream possibilities. After all, those likelihoods would depend on factors such as the depth or intensity to which the rules were internalized,[34] as well as the extent to which agents have been prepared to overcome temptations to violate rules they accept. But such factors depend crucially on upstream considerations of how the code is taught and what sorts of enforcement mechanisms are in place to support it.

The situation is even worse when it comes to the downstream bad consequences we can expect from the 20% who do not accept the code. As Holly Smith has pointed out, "[t]he possible ways of rejecting a given code are legion, and they can lead to very different actions, even among agents facing the same situation."[35] If a code is rejected in favor of another that is quite similar to it, then the results of non-acceptance may be barely distinguishable from those of acceptance. Alternatively, a code might be rejected in favor of one vastly worse (a repugnantly twisted code), or one that is vastly more demanding (a code that only moral saints could live out), or anything in between. Some might even reject all moral codes, living as amoralists. The details of how and

[34]Although the most recent literature has devoted much attention to the issue of how widely a code is internalized within a social group, it has paid comparatively little attention to the depth or intensity of internalization. Brandt raised the issue fairly consistently; see, e.g., Brandt, Toward a Credible Form of Utilitarianism, pp. 177–78; Some Merits of One Form of Rule-Utilitarianism, p. 126; Fairness to Indirect Optimific Theories in Ethics, pp. 142–47; Problems of Contemporary Utilitarianism: Real and Alleged, p. 98; *A Theory of the Good and the Right*, pp. 291–92; and, A Utilitarian Theory of Excuses, in *Morality, Utilitarianism, and Rights* (New York: Cambridge University Press, 1992), pp. 215–34 (pp. 228–32). Hooker, following Brandt, appeals to the differing levels of motivation and aversion that the consciences of "morally well-trained people" would attach to different rules when he addresses the problem of how RC should handle conflicts among the rules (see *Ideal Code, Real World*, pp. 88–92, 131–34). However, Eggleston has argued that Hooker's theory does not allow for this appeal to conscience to add anything that isn't already included within the ideal code itself, since his theory equates conscience formation with code internalization; see B. Eggleston, Conflicts of Rules in Hooker's Rule-Consequentialism, *Canadian Journal of Philosophy* 37 (2007), 329–49. Hooker's and Brandt's repeated references to proper moral training and education suggest that a teaching formulation of RC would offer a more natural framework for incorporating their approach to conflicts of rules. Since teaching formulations focus primarily on the best approaches to inculcating codes, and they identify the ideal code as whatever code that ideal approach to moral education would attempt to inculcate, they allow for the possibility that an ideal approach to moral education would involve more nuanced conscience formation than is captured within the content of the rules themselves.

[35]Smith, Measuring the Consequences of Rules, p. 427.

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

why a code is rejected clearly matter a great deal. A further complicating factor is that, depending on how the code we are evaluating is enforced, some of those who reject the code might nevertheless comply with its rules in spite of the fact that they do not accept them. The simple fact is that the downstream costs of a 20% non-acceptance rate are thoroughly indeterminate. As the example of the moral saint demonstrates, there is not even a guarantee that non-acceptance will always lead to costs instead of added value; much depends upon what would be accepted in the code's place, and much depends as well on the various ways in which compliance and acceptance can come apart from each other. Again, an acceptance theorist, could attempt to consider all of the possible forms and distributions of non-acceptance within the 20% and add another layer of weighted average value of calculation to the theory.[36] Ignore the complexity of the imagined calculation. The more fundamental difficulty, once again, is that in the absence of the upstream information (e.g. how the moral code in question is taught and enforced) there could not even be any determinate likelihoods to plug into the weighting of these downstream possibilities. Tobia would seem to be left with two possible options: either join CRRC in stipulating the upstream considerations concerning teaching and enforcement methods, or somehow try to "solve for" those upstream factors based on the stipulated acceptance rate.

Let us consider, then, the upstream portion of acceptance formulation calculations. The goal is to estimate the internalization costs that would be involved in achieving our stipulated 80% acceptance rate. The first thing to note is that the concerns Tobia raises against CRRC's proposal to stipulate methods of teaching and enforcement arise at this point within acceptance formulations as well. Recall that he objected to the idea of stipulating teaching and enforcement methods based on the difficulty that the methods that would obtain "in the real world" are "influenced by various political processes and complex social norm interactions." However, acceptance theories cannot avoid such issues; they merely locate them upstream from their stipulated assumptions rather than downstream from them as in CRRC. How, after all, will acceptance theories go about calculating the upstream costs of achieving an 80% rate of acceptance? Those costs would depend upon a variety of factors certain to include the methods of teaching and enforcement utilized in teaching the code; moreover, the likelihoods that various teaching methods would obtain, as well as the extent of their effectiveness if they were to obtain, would be influenced by the very sorts of contingencies of political processes and social norm interactions that Tobia regards as problematic for CRRC.

Acceptance theorists have had little to say about such real-world complications. In fact, the approach to internalization costs that Hooker endorses is purposefully designed to ignore them.[37] His formulation of RC accounts for the internalization costs that result from the relatively fixed facts about human nature (e.g. our cognitive limitations and our self-centered inclinations),[38] since their inclusion helps acceptance

[36]This response is recommended by Shang Long Yeo, Measuring the Consequences of Rules: A Response to Smith, *Utilitas*, 29 (2017), 125–31 (p. 130).

[37]However, its success in avoiding them has been questioned; see T. Mulgan, Ruling out Rule Consequentialism, in *Morality, Rules, and Consequences*, ed. by B. Hooker, E. Mason, and D. Miller (New York: Rowman & Littlefield, 2000), pp. 212–21.

[38]These costs are only "relatively" fixed because under certain future contingencies, such as those involving genetic engineering, they might turn out to be more flexible. See Hooker, *Ideal Code, Real World*, p. 32, fn. 1; cf. R. Arneson, Sophisticated Rule Consequentialism: Some Simple Objections, *Philosophical Issues*, 15 (2005), 235–51 (pp. 248–49); and B. Hooker, Reply to Arneson and McIntyre, *Philosophical Issues*, 15 (2005), 264–81 (pp. 268–9).

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

formulations avoid codes with overly complex or demanding rules. However, his theory gives no weight to what we might think of as *transitional* internalization costs – the costs of getting societies to give up various unjustified elements of their current codes in order to replace them with an ideal code. Hooker brackets off the consideration of transition costs by counting only the costs of getting codes internalized by future generations who have not yet internalized any sub-optimal codes and ignoring the complications involved in getting a code accepted by current generations (who have already internalized a sub-optimal conventional morality).[39] Bracketing off transition costs in this way helps to ensure that RC's selection of its ideal code is not tainted by unjustified elements of conventional morality, and it conveniently insulates his formulation from dealing with the sort of messy contingencies Tobia cites as problematic for CRRC. However, it purchases that purity at the cost of practical applicability, since such bracketing amounts to magically assuming into existence an ideal teaching generation without making any attempt to account for the costs and feasibility of bringing it about.[40] The strategy does not solve the real-world difficulties introduced by transition costs so much as magically wish them away, and this fact puts Tobia's criticism of CRRC in an awkward position. If acceptance theories are entitled to magically wish such difficulties away, then teaching formulations should be entitled to the same. On the other hand, if Tobia wants to hold CRRC accountable to addressing those difficulties, he should hold acceptance formulations accountable to the same standard. Either way, CRRC seems to be at no disadvantage to acceptance formulations with regard to such "real world" complications.

The second thing to note about the upstream portion of acceptance formulation calculations is that they are problematic for reasons that mirror the issues affecting their downstream calculations. Just as it is impossible to determine what the downstream costs and benefits of an 80% acceptance rate would be independent of the upstream considerations of teaching and enforcement methods, so it would be impossible to calculate those upstream costs on the basis of only an acceptance rate without further information about the downstream details. We have already noted that an 80% acceptance rate can play out in many very different ways downstream (different levels of compliance among the accepting, different patterns of rejection among the 20%, etc.). As a result, there are bound to be at least as many different ways that an 80% rate could be achieved upstream, and hence, many possibilities concerning how costly achieving that rate would be. Once again, an acceptance theorist might concede the fact and insist on performing yet another weighted average value calculation over all of the possible ways of achieving that acceptance rate. Ignore the complexity of the imagined calculation. The deeper problem is that without specifying which of the various downstream possibilities we are considering, there would be no fact of the matter concerning the probability weights to assign to the various upstream possibilities. Hence, reverse engineering the teaching and enforcement methods needed to achieve a stipulated level of acceptance in order to calculate the upstream internalization costs is not a viable strategy; but if such variables cannot be "solved for" within the acceptance formulation

[39]Hooker, *Ideal Code, Real World*, pp. 79–80. Mulgan endorses Hooker's strategy and incorporates it in his teaching-centered formulation of RC; see Mulgan, How Should Utilitarians Think about the Future?, p. 291; and Mulgan, *Future People*, pp. 140–41, 146–47.

[40]Cf. Hooker's charge that if we assume complete acceptance of a code "we have simply imagined out of existence unmitigated amoralists and those who are conscientious but misguided" (*Ideal Code, Real World*, pp. 82, 84).

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

structure, the only alternative would seem to be that they must be stipulated as inputs to the calculations.

The general lesson to be drawn is that many of the challenges facing contemporary RC are exacerbated by two disadvantages of the decision to evaluate codes based on stipulated acceptance rates. I observed earlier that in order for codes to have consequences, they have to be *taught, internalized, and acted upon*; by choosing to stipulate values for the variable that is located in the middle of this process, acceptance formulations are forced to attempt to work both backwards and forwards from that stipulation. Secondly, the variable they stipulate is a quite indeterminate one; a single acceptance rate is compatible with many different combinations of the other factors they are connected to (e.g. compliance rates, patterns of non-acceptance, teaching and enforcement methods, etc.). To put the point in terms borrowed from the lexicon of philosophy of mind, since a given acceptance rate is multiply realizable by a wide variety of sets of values for the other variables that inform rule consequentialist calculations, stipulating a value for that variable gives us very little of the information needed to calculate either the upstream or downstream aspects of a code's expected value.

3.3 The structural advantages of teaching formulations

Teaching formulations of RC avoid both of those difficulties. They locate the initially stipulated variables as far upstream as possible, so that all subsequent calculation can be carried out downstream; moreover, the variables they stipulate are capable of being made more or less determinate, to the extent that a more detailed account of the teaching and enforcement methods promises to provide better insight into codes' expected values. Moreover, the stipulated assumptions are grounded in the most immediate questions we face with regard to moral codes: namely, *which code should we teach* and *how should we teach it*. Since these are the most immediate objects of choice, it seems only natural that they should be the variables we stipulate for the sake of comparing expected outcomes. By contrast, rates of acceptance and compliance cannot be aimed at directly; such goals can only be achieved indirectly through the choices we make concerning the methods we use to inculcate a moral code.

Teaching-centered approaches to embedding conditions have several important advantages that their proponents have noted. For one thing, by avoiding the false choice between compliance and internalization, they are able to account for the various ways in which teaching and enforcement of a code can influence both of those variables.[41] Moreover, they offer a more intuitive approach to the problems of partial acceptance and compliance; since we have much experience to draw upon in discerning the realistic limits of the success we can expect from our efforts at moral education, teaching-centered formulations will be sure to select codes that provide guidance concerning how to deal with non-compliance. Similarly, since we can anticipate that our attempts to teach highly complex or demanding rules would result in confusion, frustration, or outright rejection, teaching formulations will easily identify such rules as sub-optimal.

Teaching-centered formulations also avoid what Mulgan describes as the "artificiality of assuming that the internalization of a code is largely an all or nothing thing."[42] Whereas acceptance formulations are forced to divide people crudely into two categories, those who have accepted a code and those who have not, teaching formulations can

[41]For more on this point, see T. Miller, Solving Rule Consequentialism's Acceptance Rate Problem, p. 51.
[42]Mulgan, *Future People*, p. 147.

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

naturally account for the reality that the internalization of a code is almost always a matter of degree, both in terms of which of its rules are internalized and the depth of their internalization. As a result, teaching formulations can account for the fact that various rules within a code can be expected to achieve varying levels of internalization and compliance.[43] Instead of an unnuanced stipulation of an overall acceptance rate for an entire code, teaching formulations allow the varying levels of acceptance that different rules would achieve to be located among the variables that are "solved for" when we think about the expected results of our attempts to teach a code. This is a crucial advantage, since securing high levels of acceptance and compliance is more important for some rules than it is for others. Moreover, as Dale Miller has noted, teaching formulations also provide a helpful perspective for thinking about the consequences of non-acceptance:

> Experienced classroom instructors have some empirical basis for predictions not only about what percentage of students are likely to completely master the material if it is taught via a certain method but also about what percentages of students are likely to fall short of complete mastery in different ways. Similarly, our collective experience with moral education gives us some basis for predictions not only about what percentage of the population is likely to internalize the rules they are being taught if a particular system of moral education is employed but also what percentages of people are likely failing to internalize them and how, e.g. by internalizing rules that are subtly different in various ways or by failing to internalize any rules whatsoever.[44]

Although making predictions about the negative effects of non-acceptance is bound to remain one of the most difficult aspects of RC's calculations, teaching-centered formulations seem to render even this difficult problem at least somewhat more manageable.

Of course, Tobia is correct to note that teaching formulations do not eliminate the question of how RC should compare rules and codes in the face of uncertainty about the outcomes of those choices. CRRC, for example, still does have to "solve for the acceptance rates" and such calculations will inevitably be made under conditions of uncertainty.[45] However, it is a mistake to suppose that the criterion for comparison under conditions of uncertainty is the most critical and defining question for formulations of RC to address. In fact, I am skeptical of Tobia's apparent presupposition that there is a single "best" criterion to be identified,[46] in part because the uncertainties we

[43]See Mulgan, *Future People*, p. 147; and T. Miller, Solving Rule Consequentialism's Acceptance Rate Problem, p. 50. Cf. Sidgwick's observation that "the difference between the moralities of any two societies is often more strikingly exhibited in the different emphasis attached to various portions of the moral code in each, than in disagreement as to the rules which the code should include" (*The Methods of Ethics*, p. 484).

[44]D. Miller, Moral Education and Rule Consequentialism.

[45]Proponents of teaching formulations have not suggested otherwise. See, e.g., Mulgan, How Should Utilitarians Think About the Future?, p. 292; and T. Miller, Solving Rule Consequentialism's Acceptance Rate Problem, p. 52.

[46]Tobia does entertain the possibility that different criteria might be appropriate to different questions RC faces; e.g. he suggests that RC might employ one method of comparison with respect to the partial compliance problem, and a different one when considering possibilities concerning how similar the future will be to the present; see Tobia, Rule-Consequentialism's Assumptions, p. 468. However, his discussion of these issues seems to assume that for each particular question RC faces, there will be a single "best" approach to making such comparisons, and that is the assumption I am questioning here. Even if we limit our focus to a

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

face often run deeper than Tobia's treatment of the issue suggests. His preferred maximizing expectation criterion would have us approach conditions of uncertainty by considering all of the possible outcomes of our various options, multiplying the value of each outcome by its likelihood of obtaining in order to identify the option(s) that maximize expected value. Anticipating the objection that this approach "demands overwhelming and unrealistic amounts of knowledge and computation," Tobia responds that it "perfectly accommodates a lack of knowledge; one of the main reasons for using expected value to measure the value of some code is to deal with uncertainty and lack of knowledge."[47] However, this response ignores the difficulty that the limitations of our knowledge often extend farther than simply not knowing with certainty which outcome would occur. We are often unable to even anticipate in advance the full range of possible outcomes that might occur, and even when we can do that, the likelihoods of their obtaining may be inscrutable to us. The maximizing expectation criterion may offer a useful guide for choices faced under highly idealized cases of uncertainty, but Tobia's claim that it is "plausible across the board"[48] seems unduly optimistic. After all, how many of our deliberations concerning the contents of the ideal code will be so idealized? Under less idealized situations of uncertainty, other criteria may be more appropriate.[49] These are worthy issues for rule consequentialists to explore, but the assumption that any one criterion will be definitive of the "best" formulation of RC seems doubtful.

4. Conclusion

The unchallenged influence of the standard case for acceptance formulations has locked contemporary RC into a theoretical structure that unhelpfully stipulates an indeterminate variable located in the middle of its calculations. As long as RC remains within the confines of that structure, the prospects for an adequate formulation of the theory look grim; no amount of technical ingenuity seems likely to overcome the structural disadvantages of making stipulated acceptance rates the starting point for RC's comparisons of codes' values. However, recent work has begun exploring the more promising alternative of reorienting RC's structure in terms of teaching-focused accounts of embedding conditions. I have argued that the criticisms Tobia has directed at CRRC's teaching formulation miss their mark, in part, because they fail to recognize the importance of where RC locates its stipulated variables in relation to those it solves for. Teaching formulations are still in the early stages of their development, and they raise a number of questions of formulation that space prohibits me from addressing here. Most obviously, they will have to address the question of how they interpret the scope of both the teaching and learning generations, and addressing those questions will require decisions concerning whether the theory should aim to identify a universal code that applies to everyone, both everywhere and everywhen, or whether it should be developed in

single type of question, such as how we should compare rules given uncertainty concerning the levels of acceptance and compliance they would achieve, it is not obvious that a single criterion will apply to all of the various choices that fall under the question.

[47]Tobia, Rule Consequentialism and the Problem of Partial Acceptance, p. 650.

[48]Tobia, Rule-Consequentialism's Assumptions, p. 468; cf. Yeo, Measuring the Consequences of Rules: A Response to Smith, p. 130.

[49]Cf. Yeo, Measuring the Consequences of Rules: A Response to Smith, p. 128.

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press

ways that are relativized to more specific cultural and/or historical contexts.[50] However, in light of the argument I have offered in this article, it seems unlikely that the task of answering these questions will reveal any difficulties that could provide reasons to prefer acceptance or compliance formulations over teaching formulations. Whatever challenges these questions may pose for teaching formulations, the same challenges seem bound to arise in even more problematic forms within the upstream portion of the calculations attempted by acceptance and compliance formulations.[51] The structural advantages of situating RC's stipulated assumptions as far upstream as they can be located provide reason to be hopeful that teaching formulations might provide RC with a path of escape from the theoretical dead-end in which acceptance formulations seem to be stuck.[52]

[50]Other proponents of teaching-focused RC have already devoted some attention to these questions. Mulgan's IMOU endorses a temporally relativized approach; see Mulgan, How Should Utilitarians Think About the Future?, pp. 292–97. Likewise, Dale Miller endorses an understanding of UMERC that is relativized on both temporal and cultural dimensions; see D. Miller, Moral Education and Rule Consequentialism.

[51]In fact, they already have. Hooker makes allowance within his acceptance formulation for a limited degree of temporal relativism conditioned on changes that would fundamentally alter the fixed internalization costs of codes (see fn. 38 above). Moreover, in response to epistemic concerns regarding our ability to identify a truly ideal code, he has embraced an incrementalist approach to improving moral codes, introducing a further degree of temporal relativism into his account; see Hooker, Reply to Arneson and Macintyre, p. 278. Finally, Leonard Kahn has defended a culturally relativized version of RC within the context of Hooker's acceptance formulation; see L. Kahn, Rule Consequentialism and Scope, *Ethical Theory and Moral Practice*, 15 (2012), 631–46.

[52]I am grateful to the Appalachian College Association for its generous funding of a faculty fellowship and to Lee University for a sabbatical leave that has afforded me the time to work on this article. I am also indebted to Dale Miller and two anonymous referees for helpful comments and suggestions.

Cite this article: Miller TD (2021). From Compliance, to Acceptance, to Teaching: On Relocating Rule Consequentialism's Stipulations. *Utilitas* **33**, 204–220. https://doi.org/10.1017/S0953820820000369

https://doi.org/10.1017/S0953820820000369 Published online by Cambridge University Press