Rule Consequentialism and the Problem of Partial Acceptance

1. Introduction

Often the consequences of the acceptance of a rule depend on how many people accept that rule. The question of how to deal with differing levels of acceptance is one that a wide range of ethical theories must answer. Partial compliance and acceptance worries have stirred recent debate in Rule Utilitarianism (e.g. Ridge 2006, Hooker 2007), but also raise problems for Kantianism (e.g. Korsgaard 1986) and Contractualism (e.g. Watson, 1998). The question seems particularly relevant to Rule Consequentialism, since the number of people following a code of rules typically has a direct influence on the consequences of those rules.

In an effort to determine the best version of Rule Utilitarianism[1], I will analyze three established types of Rule Utilitarianism before presenting my own suggestion.[2] I will first discuss a Fixed Rate suggestion (Hooker, 2000) and Michael Ridge's Variable-Rate Rule Utilitarianism (Ridge, 2006). I will then turn to Holly Smith's Optimum Rate Utilitarianism (Smith, 2010) before introducing my own proposal, Maximizing Expectation Rate Rule Utilitarianism. After analyzing these variants of Rule Utilitarianism, I will argue that Optimum Rate and Maximizing Expectation are the two most plausible versions of Rule Utilitarianism, though Maximizing Expectation captures our intuitions about solving partial acceptance problems without the costs of the other three theories discussed.

---

[1] Because much of the debate on this topic has been framed in the context of Utilitarianism rather than Consequentialism, I will discuss Rule Utilitarian formulations. Most arguments here can be easily extended to Rule Consequentialist theories.

[2] There is one preliminary distinction to draw about Rule Utilitarianism or Rule Consequentialism's formulation: the difference between compliance and acceptance forms. The difference between these two is well described by Tim Mulgan (2006, 138):
Compliance Rule Consequentialism: To find the optimal rule set we ask what would happen if a given set of rules were always complied with by everyone.
Acceptance Rule Consequentialism: To find the optimal rule set we ask what would happen if a given set of rules were accepted by everyone.

2. Hooker's Fixed Rate Rule Utilitarianism

Brad Hooker has developed a version of Rule Consequentialism, Fixed Rate Rule Consequentialism, which claims we should follow the rules whose acceptance by some fixed percentage of the population would lead to the best consequences. Hooker's formula is Consequentialist, rather than Utilitarian, though the distinction between Consequentialist and Utilitarian formulations is not relevant to this paper, so for comparison with the competing theories I will discuss a Utilitarian Fixed Rate theory.[3] Hooker formally defines his Fixed Rate Rule Utilitarian (Fixed Rate RU) as follows:

> 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 … The calculation of a code's expected value includes all costs of getting the code internalized. If in terms of expected value two or more codes are better than the rest but equal to one another, the one closest to conventional morality determines what acts are wrong (Hooker, 2000, 32).

Hooker later specifies that the "overwhelming majority of everyone everywhere" occurs at 90% acceptance (Hooker, 2000, 83). Other Rule theorists have advocated for a 90% rate as well (e.g. Brandt, 1992).

One of the most common objections to Fixed Rate theories like Hooker's is that it seems arbitrary; why not choose 99% or 60% acceptance Fixed Rate Rule Utilitarianism? The answer to this question may be easy enough; Hooker says we should seek a percentage close to absolute acceptance, but also one that acknowledges that codes might face partial acceptance problems (Hooker, 2000, 84). Thus, Hooker might argue, a 60% acceptance Fixed Rate code fails to come close enough to absolute acceptance and a 99% acceptance Fixed Rate code does not adequately account for reasonable partial acceptance issues.

---

[3] Note also that Hooker discusses a Utilitarian counterpart to his Consequentialist theory (e.g. Hooker and Fletcher, 2007).

It is certainly possible and also very likely, an objection to Fixed Rate theories begins, that there is a Fixed Rate Rule Utilitarianism with a rate other than 90% that satisfies these two conditions. An 89% Fixed Rate Rule Utilitarianism, for example, seems to recognize partial acceptance problems while remaining close enough to absolute acceptance. Since this is the case, it seems that an 89%, 90%, or maybe even a 95% internalization of Fixed Rate Rule Utilitarianism could work just as well given Hooker's considerations. There is no *prima facie* reason to choose specifically a 90% acceptance Fixed Rate Rule Utilitarianism and, furthermore, there is no *prima facie* reason to choose any specific fixed rate.

Other problems for Fixed Rate theories come from their focus on just one acceptance level. For Hooker's theory an act is wrong if and only if it is forbidden by the code of rules whose acceptance by roughly 90% of everyone everywhere in each new generation has maximum expected value in terms of well-being, but it may well be the case that another code leads to even better consequences at 95% or 83% acceptance. It is possible that using a fixed acceptance rate might lead to accepting a code whose expected well-being is less than that of another code accepted at a different acceptance level.

While choosing 90% as the level of acceptance of the ideal code might work surprisingly well in bringing about the best consequences, it still seems a relevant issue that Fixed Rate Rule Utilitarianism focuses on just one level of acceptance. It is therefore worth pursuing alternative theories to Fixed Rate Rule Utilitarianism.

3. Ridge's Variable Rate Rule Utilitarianism

A second type of Rule Utilitarianism is Ridge's Variable Rate Rule Utilitarianism. This version attempts to improve upon Fixed Rate versions by creating a less arbitrary theory that also

provides guidance at multiple acceptance levels. Ridge defines Variable Rate Rule Utilitarianism (Variable Rate RU) as follows:

> An action is right if and only if it would be required by rules which have the following property: when you take the expected utility of every level of social acceptance between (and including) 0% and 100% for those rules, and compute the average expected utility for all of those different levels of acceptance, the average for these rules is at least as high as the corresponding average for any alternative set of rules (Ridge, 2006, 248).

Ridge defends his theory as less arbitrary than Fixed-Rate theories since it takes into consideration acceptance at every possible level rather than focusing on one arbitrary level of social acceptance. While his theory may be less arbitrary in that it does not fix a rate of acceptance, it is not clear that the theory is less arbitrary on the whole. One reason it seems arbitrary is that there are a variety of plausible ways to combine the expected utility for codes at different acceptance levels. An argument could be made to choose the code with maximized median, or for that matter mode, expected utility across possible levels of acceptance. Security-based reasoning might lead us to select the code with the greatest possible minimum, or the code with the greatest minimum on average across acceptance levels. There is certainly room to debate which of these variable rate methods would form the best criterion of rightness, but it is arbitrary to choose one without considering the others. Ultimately Ridge's criticism of Fixed Rate RU as arbitrary applies just as strongly (or weakly) to his own Variable Rate RU.

Furthermore, it has been shown that Ridge's averaging of all rates can lead to unappealing results if they are skewed by an anomaly (Fletcher and Hooker, 2007, 6). The argument is that if code A's expected value is 10 at all levels of acceptance, 0% to 100%, while code B has an expected value of 0 at every level, but an expected value of 102 only at 100% acceptance, then Variable Rate RU implies we should follow code B. The problem is that if we suppose 100% acceptance is highly unlikely, we would probably prefer code A. These considerations weaken

the claim that Variable Rate RU is non-arbitrary and call into question whether averaging expected utilities of codes across all possible acceptance levels is the best method for dealing with partial acceptance problems.

A related problem for Variable Rate RU is that it does not take into account what type of probability distribution over possible levels of acceptance we are to assume. By averaging expected utilities across all possible acceptance levels, Variable Rate RU assumes each possible acceptance level is equally likely to obtain. Under this uniform distribution, the probability that a given acceptance level will obtain is equal for all possible levels. It does not take much imagination to realize that codes leading to the best consequences at a 4% acceptance rate and codes leading to the best consequences at a 98% acceptance rate will be dramatically different. This would certainly present a challenge when choosing what principles to follow. A further problem is that the probability distribution over all possible acceptance levels is almost certainly *not* uniform; for a given code, certain acceptance levels are more likely than others to obtain.

Variable Rate RU enforces ignorance by intentionally removing beneficial information – it tells us to choose a code that does best on an average *across all acceptance levels* rather than to choose the code that does best taking into account the *probability each acceptance level will obtain.* Averaging across all acceptance levels is not a good tool to use against partial acceptance problems. Averaging expected values equally across all acceptance levels removes information crucial to the decision making process, reducing the expected utility of a chosen ideal code. The main problem for Ridge's Variable Rate RU is that it requires people to follow the rules that on average make things best across all acceptance levels, with no regard for how likely it is for certain acceptance levels to obtain.

4. Smith's Optimum Rate Rule Utilitarianism

Holly Smith has suggested a third alternative Rule Utilitarian solution to partial acceptance problems. Smith seeks to avoid both the arbitrariness of Hooker's Fixed Rate RU and misguided calculation of Ridge's Variable Rate RU. Her theory, Optimum Rate Rule Utilitarianism, posits that we should first look at how all codes will do across all acceptance levels. We should then accept the code that has the highest possible level of social welfare at *some* possible acceptance level. She formulates her Optimum Rate Rule Utilitarian theory (Optimum Rate RU) as:

> An ideal code is the code whose optimum acceptance level is no lower than that of any alternative code (Smith, 2010, 418).

For Smith's theory, it does not matter at which level of acceptance this optimum point occurs. We simply follow the code that *could* do the best, regardless of at what acceptance level this occurs. The ideal moral code could reach its optimum rate at 100% acceptance, 90% acceptance, or even 42% acceptance, though it does seem implausible the code with the potential to reach the highest expected social welfare would do so at an extremely high or low percentage acceptance. Smith graphically plots alternative codes at different acceptance profiles, or levels of acceptance, against the expected welfare at the given level (Smith, 417); see figure 1.1 below.

In this example, Code A is the ideal code that we should strive to follow and teach since it has the highest optimum expected social welfare (occurring at around 85% acceptance). Although Code B does better at 100% acceptance than code A, the loss in social welfare required to reach 100% acceptance makes Code A a better code because its optimum point (at around 85% level acceptance) is higher than the optimum point of code B (at around 75% level acceptance). Smith argues that choosing the code with the highest optimum is less arbitrary than

Fixed Rate RU and allows those who follow the ideal code to concentrate on the appropriate level of acceptance, as opposed to Variable Rate RU, which advocates averaging expected utilities equally across all possible levels of acceptance.

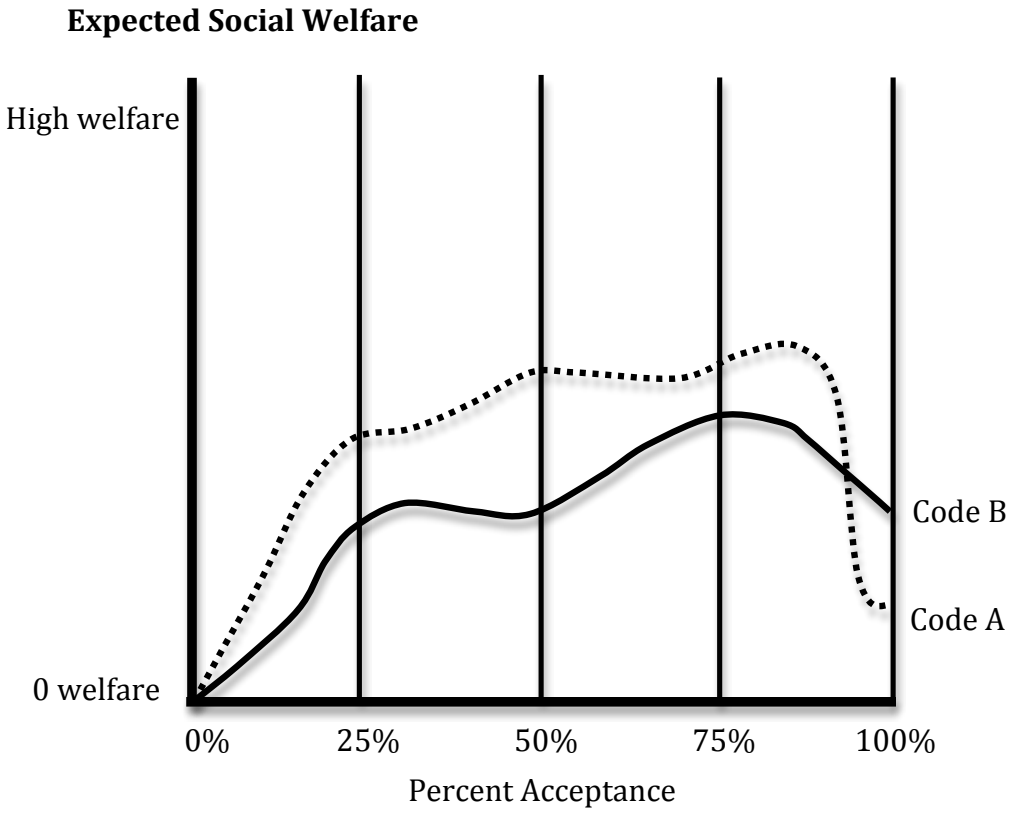

Figure 1.1

Optimum Rate Rule Utilitarianism certainly seems a step in the right direction, as it acknowledges important considerations missed by Fixed and Variable Rate RU. Namely, the code focuses on attaining the highest level of expected welfare and certainly seems less arbitrary than either of the previously discussed formulations of Rule Utilitarianism. There are, however, new challenges associated with an Optimum Rate proposal.

Optimum Rate RU advocates that we pick the code with the highest optimum point, which, though an improvement on Fixed Rate RU and Variable Rate RU, fails to acknowledge

that all codes with high optimum levels may perform poorly at other levels compared to more stable codes. For instance consider the following codes A, B and C in figure 1.2 below.

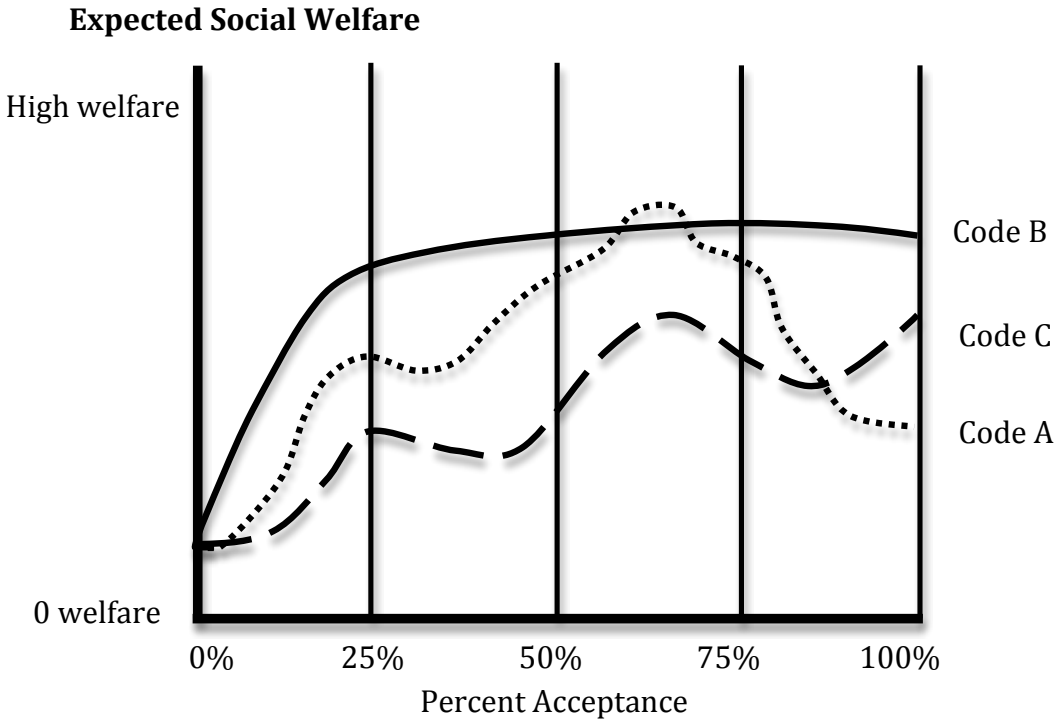

Figure 1.2

Here, Code A has the highest optimum point, though Code B appears better at a wide range of other acceptance levels; Code B consistently provides greater expected social welfare at nearly every level outside the narrow range from roughly 55 to 65 percent acceptance, it also seems like a safer choice than other codes since it provides high welfare across what we could expect to be the percent acceptance. Notably, even in the 55% to 65% range, codes A and B both provide a relatively similar expected social welfare.

The defense of Code B here rests on intuition, and it is possible that some may still find appeal in Code A. But that Code B almost always outperforms Code A indicates the tension between Smith's Optimum Rate RU and Ridge's Variable Rate RU; the code with the greatest optimum and the code that will do best on average are likely to be two different codes. This

objection to Optimum Rate RU is certainly in some tension with my earlier criticism of Variable Rate RU and average forms of Utilitarianism. In this case, code B appears to also do better on average and it is likely to be the code picked out by Variable Rate RU. I will argue later for another principle, not an appeal to which code does best across an average of all possible acceptance levels, that leads us to pick B here. My remarks here are meant as a critique of Optimum Rate RU, not an endorsement of Variable Rate RU.

It seems in some cases we want to select the code with the greatest optimum point, but in others we want to appeal to a sense of average expected utility at more than one acceptance level. Importantly, both Optimum Rate RU and Variable Rate RU (and Fixed Rate RU) fail to fully take into account how likely it is for a given level of acceptance to obtain for a given code. This helps motivate a new suggestion, Maximizing Expectation Rate Rule Utilitarianism, which I turn to in the next section.

5. Maximizing Expectation Rate Rule Utilitarianism

When attempting to solve partial acceptance problems, there seem to be at least four principles worthy of consideration:

1. Potential maximization of utility: The code we follow should have the potential to reach the highest level of utility possible among possible codes; the ideal code, if followed at x percentage of acceptance will have the highest utility of any code at any level of acceptance.
2. Highest average utility: The code we follow should, on average across possible acceptance levels, provide the highest level of utility of possible codes.
3. Greatest chance of acceptance at levels that would result in high levels of utility: The code

we follow should have the greatest chance of being accepted at levels at which it would provide a very high utility level.

4. Least chance of acceptance at levels that would result in low levels of utility. The code we follow should have the least chance of being accepted at levels at which it would provide a very low utility level.

Quite obviously, these principles conflict and compete with one another. Each of the previously discussed versions of Rule Utilitarianism prioritizes at least one of these principles. Smith seems to value principle 1 above all others; for Optimum Rate RU the best code *is* the code that satisfies principle 1. Hooker's Fixed Rate RU seems to be motivated by a desire to accommodate principle 1 as well. Hooker advocates for the code whose acceptance by the overwhelming majority of everybody everywhere will lead to the highest expected value in terms of well-being. Implicit in Hooker's argument might be a belief that this ideal code has the potential to reach the highest level of utility possible; whatever the ideal code is, its optimum rate will occur when it is accepted by an overwhelming proportion of people, at roughly a 90% acceptance. Fixed Rate RU may, arguably, satisfy principles 3 and 4 as well. One might argue that the code leading to the best consequences when an overwhelming percentage of people accept it is also the one with the greatest chance of acceptance at high utility levels of acceptance and least chance of being accepted at low utility levels of acceptance. Ridge gives priority to principle 2 here; his code is the one that satisfies this condition. By averaging the expected utility of codes across possible acceptance levels, Ridge rejects principles 1,3, and 4.

When discussing Ridge's Variable Rate RU, Smith asks, "why test its acceptance-utility at every possible level, when in reality it must be accepted at some level or the other?" (Smith, 2010, 416). This is an important objection, as a code will be accepted at some level (at a given

time). A further refinement is needed, though. Why value a code's acceptance-utility at only one possible level, when in reality the acceptance rate is uncertain? This type of critique seemingly pulls us back to Variable Rate RU, but there is a problem there too; we don't want to simply average all the codes. In order to evaluate a given code, we want to take an expected utility value at each acceptance level and *an expected value of the probability that each acceptance level will obtain.* We can use these two expectations together to determine the code that we can expect will lead to the best consequences. This calculation is at the center of my new Rule Utilitarian proposal, Maximizing Expectation Rate Rule Utilitarianism (Maximizing Expectation RU):

> Maximizing Expectation Rate RU: An ideal code is the code whose rules provide an expected utility at least as high as the corresponding expected utility for any alternative set of rules. The weighted expected utility of a code at a given acceptance level is the product of the expected utility of the code at that acceptance level and the percent likelihood that society will accept the code at this level. The expected utility of a code is the sum of the code's weighted expected utilities at every level of social acceptance between 0% and 100%, inclusive.

The force of this form of Rule Utilitarianism is that given uncertainty of facts, allowance for change, and people's differing willingness to accept rules at different levels of acceptance, we should follow the set of rules that we expect to have the best consequences. The code's mathematical formula gives the expected value of a code as the sum from 0% to 100% acceptance of the product of (1) the probability that the acceptance level will be l_n if the code is implemented and (2) the expected utility at the level, $E[u_n]$.The formula for a given code's expected value is then: $\sum p(l_n)(E[u_n])$ for acceptance levels n ranging from 0 to 100.

This new suggestion, Maximizing Expectation Rate RU, has several benefits. First, it captures two important intuitions about what our ideal code should be; any code strictly higher in expected utility at all levels than other codes is selected by Maximizing Expectation Rate RU

over these other codes. Also any code whose probability weighted average expected utility is higher than any other code's is the one we should follow.

I have argued earlier that both Fixed Rate RU and Variable Rate RU are arbitrary, but Maximizing Expectation Rate RU is non-arbitrary; it does not rely on an arbitrary fixed rate of social acceptance or an arbitrary method of weighing codes against each other at multiple levels of acceptance. By weighing codes against each other in terms of what outcomes we could expect, Maximizing Expectation Rate RU does not need to fix a rate of social acceptance.

There are three main foreseeable objections to Maximizing Expectation Rate RU, but I believe none critically damages the theory. The first possible objection is that the theory demands overwhelming and unrealistic amounts of knowledge and computation. The theory, though, 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. However, applications of Maximizing Expectation Rate RU using more extensive computations might lead to greater utility. In response, I argue that there is no strong reason to believe that the ideal code should be one at which everyone is meant to individually arrive. In fact, any requirement of total individual arrival would ensure a very simplistic code, almost certainly inferior to the actual ideal code. The need for computation is not an undermining fault for Maximizing Expectation Rate RU, but rather it provides reasons for division of tasks, communication and, fittingly, cooperation. This does not make the code any less practicable, but rather makes it one that accommodates new information and progress; new information can be used to update utility expectations and progress towards the ideal code will help simply Maximizing Expectation Rate RU calculations.

Secondly, it may be objected that under Maximizing Expectation Rate RU, the code with the highest optimum point may not be the ideal code. If there is some code whose acceptance could result in better consequences than ever possible for our code chosen by Maximizing Expectation Rate RU, how can we call the code chosen by Maximizing Expectation Rate RU the "ideal code?" One reply could be that while Maximizing Expectation Rate RU and Optimum Rate RU seem to be the two most plausible options for Rule Utilitarianism, the former chooses codes based on probable effects of acceptance while the latter chooses codes based on the best *possible*, though not necessarily probable, effects. Brad Hooker's book serving as the foundation for these debates is fittingly entitled *Ideal Code, Real World*. We are not choosing an ideal code for an ideal world, but rather an ideal code for a *real* world. The ideal code should be one that is practical and leads to high levels of well-being, not one that leads to unsatisfactory well-being with the promise of potential but unlikely maximal well-being.

A third objection is that Maximizing Expectation Rate RU faces a similar criticism to the one posed earlier against Variable Rate RU.[4] Suppose a code, A, is selected by Maximizing Expectation Rate RU, but also has a low probability of reaching an extremely high utility at one acceptance level and a medium expected utility at all other levels. A second code, B, has a high (but not extremely high) expected utility at every acceptance level. Maximizing Expectation Rate RU might lead us to choose A over B, even though we will probably be at an acceptance level with an expected utility lower than that of B. Consider Figure 3, below.

[4] I am thankful to an anonymous referee for this suggestion.

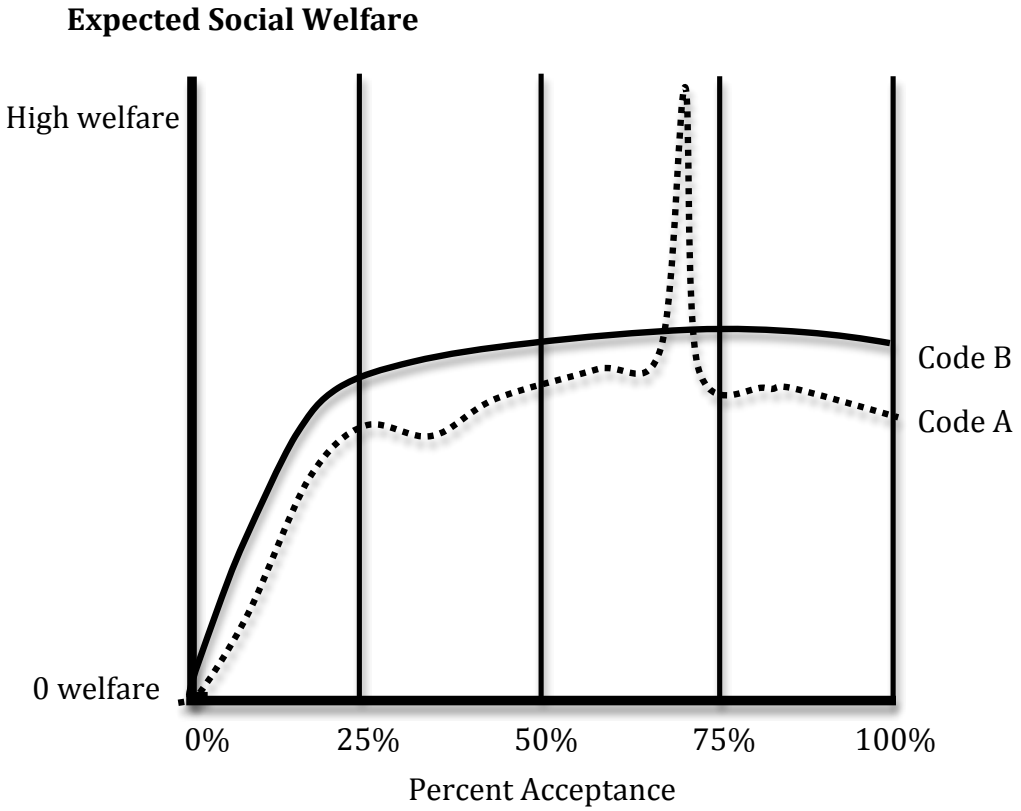

Figure 1.3

While similar in style to the objection leveled against Variable Rate RU earlier, this objection against Maximizing Expectation Rate RU does not have the same force. A premise of the objection here is that choosing code A over code B is wrong, unacceptable or unintuitive. However, the chance of attaining the extremely high expected utility (around the 70% acceptance level) should serve as a reason to pick code A, *proportional to the magnitude of the probability of reaching that expected utility.* In other words, if the probability of attaining the 70% acceptance level for code A is high enough to mathematically outweigh the weighted differences in expected utility at all other acceptance levels, then code A may be the correct choice over code B. Spikes like that of code A are problematic for Variable Rate RU, since Variable Rate RU

chooses codes by averaging across acceptance levels, and in doing so overweights the relevant impact of an unlikely but extremely high expected utility acceptance level.

Because of these reasons, I contend that Maximizing Expectation Rate RU is a better choice than Optimum Rate RU. Examples like the one described in Figure 1.2 can be constructed to show Optimum Rate RU's unappealing results. It would be very difficult for a proponent of Optimum Rate RU to construct a counterexample to Maximizing Expectation Rate RU that is equally unintuitive. Furthermore, if we revise Optimum Rate RU to a better formulation, it becomes a special case of Maximizing Expectation Rate RU. Optimum Rate RU selects as the ideal code the one with an optimum that is no lower than that of any other code, but as I have argued previously this fails to take into account the probability that a given acceptance level will obtain. In the case that the probability of reaching the acceptance level with the highest optimum is 1, then the code selected by Optimum Rate RU would also be picked out by Maximizing Expectation Rate RU as the ideal code. We should then look to Maximizing Expectation Rate RU as a more general formulation, encompassing an improved Optimum Rate RU.

6. Conclusion

In this paper I have sought to explain a new way in which Rule Utilitarian theories can handle problems of partial acceptance. I contend that current forms of Rule Utilitarianism, namely Fixed Rate, Variable Rate, and Optimum Rate Rule Utilitarianism, can be improved upon by taking into consideration the likelihood that given acceptance levels will actually obtain.

Fixed Rate RU and Optimum Rate RU both look at just one level of social acceptance; Fixed Rate RU selects the code that performs best at a *set level* of social acceptance, while Optimum Rate RU selects the code that performs best at *some level* of social acceptance.

Variable Rate RU selects the code that does best on average. Both Fixed Rate RU and Optimum Rate RU fail to take into consideration variability in acceptance levels, and while Variable Rate RU does acknowledge different acceptance levels, it does so without properly weighting them. I hope that by including the expected probability distribution of social acceptance for possible codes, Maximizing Expectation Rate RU offers a new way for Rule Utilitarian theories to respond to issues of partial acceptance.

References

Brandt, Richard. 1992: *Morality, Utilitarianism, and Rights*. Cambridge: Cambridge University Press.

Hooker, Brad. 2000: *Ideal Code, Real World.* Oxford: Oxford University Press.

Hooker, Brad and Fletcher, Guy. 2007: Variable vs. Fixed Rate Rule Utilitarianism*. The Philosophical Quarterly*, Vol 58, Issue 231, 344-352.

Korsgaard, Christine, 1986: The Right to Lie: Kant on Dealing with Evil. *Philosophy and Public Affairs*, Volume 15, Number 4, 325-349.

Mulgan, Tim. 2006: *Future People: A Moderate Consequentialist Account of Our Obligations to Future Generations.* Oxford University Press.

Ridge, Michael. 2006: Introducing Variable Rate Rule Utilitarianism*. Philosophical Quarterly*, Volume 56, 242-253.

Smith, Holly. 2010: Measuring the Consequences of Rules. *Utilitas*, 22, 413-433.

Watson, Gary. 1998: "Some Considerations in Favor of Contractualism" in *Rational Commitment and Social Justice, Essays for Gregory Kavka*, eds. Christopher Morris and Jules Coleman, Cambridge University Press, 168-187.