[Final version of this paper can be found here: https://doi.org/10.1017/S0953820816000145]

Measuring the Consequences of Rules: A Reply to Smith

SHANG LONG YEO

National University of Singapore

Abstract: In 'Measuring the Consequences of Rules', Holly Smith presents two problems involving the indeterminacy of compliance, which she takes to be fatal for all forms of rule-utilitarianism. In this reply, I attempt to dispel both problems.

In 'Measuring the Consequences of Rules', Holly Smith presents two problems involving the indeterminacy of compliance, which she takes to be fatal for all forms of rule-utilitarianism.[1] I beg to differ – and in this reply, I will attempt to dispel both problems.

Smith's first problem involves the supposedly indeterminate consequences of accepting moral codes with conditionalized rules. Such rules were proposed by rule-utilitarians to deal with cases where securing a public good requires most, but not all, agents to contribute. One example of this is the river pollution case discussed by Smith, Hooker and Lyons:[2]

> [t]hree industrialists own factories bordering the river, you among them. If two or more of the industrialists discharge waste into the river, it will be severely polluted. But if one or none discharge waste, and the others burn their waste [and hence do not discharge], the river will remain healthy, and no other environmental damage will be done[3]

[1] Holly M. Smith, 'Measuring the Consequences of Rules', *Utilitas* 22.04 (2010), pp. 413–33.
[2] Smith, 'Measuring the Consequences', p. 421, Brad Hooker, *Ideal Code, Real World* (Oxford, 2000), pp. 124–5, David Lyons, *Forms and Limits of Utilitarianism*, (Oxford, 1965), pp. 128–31.
[3] Smith, 'Measuring the Consequences', p. 421.

To deal with such cases, rule-utilitarians proposed using conditionalized moral codes like the following:[4]

R1: If two other industralists are discharging waste, you ought to discharge.

R2: If one other industrialist is discharging waste, while the other is burning waste, you ought to burn your waste.

R3: If no other industrialist is discharging waste, you ought to discharge your waste.[5]

Smith finds such codes problematic because they '*may have no determinate expected value at all*.'[6] Before examining her claim, it would be useful to recall the definition of an expected value (or expected utility). Suppose an act A has several different outcomes. To find the expected value of A, we multiply the value (or utility) of each outcome with its corresponding probability of occurring given A, and then add up all the products of these multiplications. In this formula, the probabilities used can be understood as the probability estimates justified by the available evidence.[7] In the rule-utilitarian context, expected values are most often used to evaluate the acceptance of a moral code. For instance, the expected value of universally accepting code C1 is obtained by multiplying the value of each outcome that might result from universally accepting C1 with the probability of that outcome occurring, and then adding up all the products of such multiplications.[8]

---

[4] Michael Ridge, 'Introducing Variable-Rate Rule-Utilitarianism', *The Philosophical Quarterly* 56.223 (2006), pp. 249–50.
[5] Smith, 'Measuring the Consequences', p. 421, 423.
[6] Smith, 'Measuring the Consequences', p. 421.
[7] Brad Hooker, 'Rule Consequentialism', *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/spr2011/entries/consequentialism-rule/> (2011), Rachael Briggs, 'Normative Theories of Rational Choice: Expected Utility', *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/rationality-normative-utility/> (2014).
[8] Many utilitarians have suggested using expected utility in one way or another. For a recent rule-utilitarian who uses expected utility, see Hooker, *Ideal Code, Real World*, pp. 1–2.

Returning to the river pollution case, notice that two patterns of action might emerge when there is universal acceptance of the proposed conditionalized code:[9]

Case 1: All three industrialists discharge waste.

Case 2: Two industrialists burn waste (and thus do not discharge), and one discharges waste.[10]

In Case 1, all three industralists follow rule R1 of the code. In Case 2, the two industrialists who are burning waste follow R2, while the one who discharges follows R3. Cases 1 and 2, however, have very different expected values – Case 1 results in a polluted river and a low expected value, while Case 2 results in a clean river and one industrialist saving on costs by discharging waste, thus having a high expected value.[11] Smith states that '[s]ince either of these patterns of action (note Case 2 has several distinct realizations) involves universal acceptance of the code, but has very different consequences, it is indeterminate what the consequences would be of universal acceptance of the code.'[12] I will focus on the epistemic version of this indeterminacy problem, which charges that 'we are unable to ascertain'[13] whether Case 1 or 2 will obtain, and hence cannot evaluate a moral code with conditionalized rules.[14]

My proposed solution is simple – the rule-utilitarian just needs to perform another iteration of expected value calculations in the face of such indeterminacy. From the usual

---

[9] For the sake of simplicity, I will assume – as Smith does – that universal acceptance leads to universal compliance. Nothing in my arguments will turn on this, however.
[10] Smith, 'Measuring the Consequences', p. 423.
[11] Smith, 'Measuring the Consequences', pp. 423–4.
[12] Smith, 'Measuring the Consequences', p. 424.
[13] Smith, 'Measuring the Consequences', n. 30.
[14] I believe this indeterminacy is actually an instance of what game theorists call *multiple equilibria*, and the river pollution case an instance of what is known as an *assurance game* in game theory. In order to map the assurance game onto the river pollution case, one simply needs to reinterpret the game theoretic payoffs as 'moral payoffs' that measure the moral value of each outcome, as judged by the accepted moral code. This analogy is good for the rule-utilitarian, because assurance games have been observed to be resolved quickly with communication. An introduction to the assurance game can be found in Avinash K. Dixit and Susan Skeath, *Games of Strategy*, 2nd edn. (New York, 2004), pp. 105–8, 394–6.

calculations, we already have the expected values of Case 1 and of Case 2. The expected value of the conditionalized moral code can then be obtained by multiplying the expected value of each case with the probability of that case occurring, and then adding up the products of these multiplications.[15]

Smith might object that this solution still does not eliminate the uncertainty – we still do not know if Case 1 or 2 will actually obtain. But to single out this uncertainty and demand that we resolve it without using expected utility seems quite *ad hoc*. This is especially so given that Smith already acknowledges the appropriateness of using expected utility in the face of other uncertainties. For instance, the Optimum-Rate Rule-Utilitarianism that she proposes in the same paper uses expected utility to deal with uncertainty about how the world (understood as everything other than the moral agents themselves) will turn out.[16] Hence it is unclear how she could object to my using expected utility for yet another source of uncertainty – namely, the one involving Cases 1 and 2.

Smith might also object that there is no way to obtain reasonable values for the probabilities of Cases 1 and 2 occurring. Without any information, this is indeed a problem – because we could take Cases 1 and 2 to be equally possible, and assign 0.5 probability to each, or we could take Case 1 and the three different realizations of Case 2 to be equally possible, hence assigning 0.25 probability to each. Both methods seem equally appropriate, yet they recommend very different sets of probabilities. Fortunately, we can avoid this by gathering empirical evidence for the respective probabilities. For instance, we could conduct an experiment where participants are given the river pollution case (or a suitably general one), and are told to follow the conditionalized moral code. Some proportion of Cases 1 and 2 (including 2's different

[15] This iteration of expected value calculations is exactly how expected utility theory deals with cases of nested uncertainty (the paradigm case being a lottery whose outcomes are further lotteries rather than payoffs). Iterated expected values are also used in Kevin Tobia, 'Rule Consequentialism and the Problem of Partial Acceptance', *Ethical Theory and Moral Practice* 16.3 (2012), pp. 643–52. However Tobia uses them to solve a different problem of how to evaluate a moral code whose expected consequences change depending on the level of social acceptance.
[16] More generally, all the graphs she uses to illustrate rule-utilitarianism have *expected* social utility on the y-axis.

realizations) will result, and we can use this proportion as a guide to the respective probabilities, given a large number of trials and sufficiently representative participants. In fact, economists have conducted experiments that are tantalisingly similar to what I am envisioning (for an example, see Cadsby and Maynes[17]). Such experiments are not perfectly analogous to the rule-utilitarian's case, but they at least indicate how the probabilities used by my solution might be obtained.[18]

Of course, Smith might continue to press the objection in situations where we cannot identify the probabilities involved.[19] I think that such situations are highly unlikely, given that there will always be some relevant empirical evidence that bears on the probabilities involved. (Recall that these probabilities are simply probability estimates justified by the available evidence. Thus as long as there is some available evidence, we will be able to identify the probabilities – however imprecise our estimates might be.) But suppose such a situation obtains – the rule-utilitarians might still be able to reasonably evaluate moral rules in other ways. For instance, they might follows Rawls and take a maximin approach – where we choose the set of moral rules whose worst possible outcome creates more utility than the worst possible outcome of any other set – which does not need the probabilities of any of the possible outcomes.[20]

Regardless of how rule-utilitarians deal with the situation where the probabilities cannot be identified, it is important to note that problems pertaining to probabilities are not specific to the compliance issues that Smith raises – rather, they are general problems that arise from the expected utility approach. We could just as easily formulate the same problems in terms of the probabilities of the various ways the world could turn out, for instance. Thus I believe that Smith

---

[17] Charles Bram Cadsby and Elizabeth Maynes, 'Voluntary Provision of Threshold Public Goods with Continuous Contributions: Experimental Evidence', *Journal of Public Economics* 71.1 (1999) pp. 53–73.
[18] The key difference is that participants of these experiments are not told to follow any particular moral code.
[19] Thanks to the Editor-in-Chief for urging me to address this.
[20] John Rawls, *A Theory of Justice,* Rev. edn. (Oxford, 1999), pp. 132–4. Thanks to the Editor-in-Chief for suggesting this.

is in fact committed to some way of resolving such problems, given that she already allows using expected utility to deal with uncertainty about how the world would turn out.

Smith's second problem raises a different kind of indeterminacy, and it is as follows. Suppose we have a moral code C2 that requires agents to repay their debts, and we want to find its expected value at 60 per cent acceptance (this just means the case where 60 per cent of all agents accept C2's requirement that they repay their debts). The standard way to do this is to multiply the value of each possible outcome resulting from 60 per cent acceptance of C2 with that outcome's probability of occurring, and then to add up all the products of these multiplications. However, if 60 per cent of agents accept C2, it means that 40 per cent of agents *reject* it. And this 40 per cent will likely reject the code in different ways – some might reject it by paying less than they owe, some by paying nothing at all, and some by killing the person to whom money is owed. Depending on which pattern of rejection obtains, a different set of consequences will result, changing the overall consequences of 60 per cent acceptance. The rule-utilitarian, who is concerned with such overall consequences, thus also has to determine which pattern of rejection will actually obtain. But Smith states that it is impossible to do so, hence creating an additional problem of indeterminacy.[21]

My reply is similar – the rule-utilitarian simply has to perform another iteration of expected value calculations, this time over the possible patterns of rejection. From this we can obtain the expected value of 40 per cent rejection, and add it to the expected value of 60 per cent acceptance to get the overall expected value. As with my treatment of the first problem, we could conduct experiments to guide us in assigning the respective probabilities, or, failing that, use some other reasonable approach like maximin.

---

[21] Smith, 'Measuring the Consequences', pp. 426–428. Despite the way I have presented it, this problem actually arises at 100 per cent compliance too. To see this, imagine a population that is fully compliant with some moral code. They still have to make some decisions between options that are all permissible according to this code – such as the decision between taking the bus or train to work, for instance. Choosing one option or another will have different effects on the overall consequences. But it is indeterminate as to how they would choose, and so Smith's second problem similarly applies.

In general, the rule-utilitarian can probably use expected utility in the face of any kind of uncertainty. Such uncertainty is likely to arise in many areas, extending to everything outside the rule-utilitarian's direct control. Teaching a code to 100 per cent of the population will not ensure that any determinate percentage accepts it, acceptance does not translate into any determinate percentage of compliance, and, as Smith's two problems show, compliance does not translate into any determinate set of actual consequences. The appropriate response to all this uncertainty is to use expected utility, with iteration if necessary. Computationally, this will be extremely complicated – especially because the effects extend across time, and because each agent can have an effect on others through her acceptance and compliance. At some point, I expect that rule-utilitarians will no longer perform simple calculations and instead resort to statistical simulations of moral agents being taught one moral code or another. Much like in weather simulations, these moral code simulations will yield a set of outcomes with the respective frequencies, which can then be used in rule-utilitarian calculations. The idea, however, is the same – we know that a prior state (of teaching, accepting, compliance, or whatever) could lead to several possible outcomes with the respective probabilities, and we use that knowledge to evaluate various moral rules as best we can.

Independent of the solutions I offered earlier, there is one further reason to dismiss Smith's problems, and it is that they prove too much. This is because moral rules are not the only things that fall prey to Smith's two problems – laws are susceptible too. For instance, Smith's first problem would apply equally to a law requiring at least one parent to take ten days of parental leave each year.[22] Her second problem would similarly apply to a law which states that offenders of a smoking ban have to either pay $500 or go to jail.[23] Such examples could be

[22] Even with universal acceptance of and compliance with the law, three different patterns of action might emerge – only parent A takes leave, or only parent B takes leave, or both A and B take leave. And it is epistemically indeterminate as to which one will actually obtain.

[23] Suppose 60 per cent of people do not flout the ban – this means that 40 per cent do flout it. Some proportion of this 40 per cent will choose to pay $500, and the rest will go to jail – but it is epistemically indeterminate as to what the proportions might be.

multiplied with ease – in general, the evaluation of various laws seems plagued by the same epistemic indeterminacy that Smith finds problematic, along with the intertemporal and interpersonal effects that exacerbate such indeterminacy. This indeterminacy also extends across the stages of teaching, acceptance, and compliance, just like in the case of moral rules. And yet, I suspect that most moral philosophers do not take such indeterminacy to fatally undermine the evaluation of laws. How, then, could we think that it compromises the rule-utilitarian evaluation of moral rules? [24]

[24] Many thanks to Weng Hong Tang and Christopher Anthony Brown for the comments and discussion. I am also grateful to the Editor-in-Chief of this journal for his helpful suggestions. Any errors in this paper are my own.