Rule Consequentialism and the Problem of Ties

Johan E. Gustafsson*

Draft
October 15, 2023

ABSTRACT. According to Rule Consequentialism, an act ought to be done if it is required by a rule that is better than any other rule. This formulation, however, does not tell us what to do if there are multiple optimal rules. One way to handle such ties is the Indifference Criterion, which says that an act is right if and only if it is permitted by some optimal rule. This proposal, however, may lead to outcomes with unnecessarily low expected value. The Convention Criterion breaks ties with the help of conventional morality and the Salience Criterion does so with salience. These proposals depart significantly from consequentialism and still fail to break all ties between optimal rules. This paper proposes the Second-Order Indifference Criterion, which improves on the Indifference Criterion in that it may lead to outcomes with a greater expected value and never to outcomes with lower expected value. And the Second-Order Indifference Criterion does not depart significantly from consequentialism.

Rule Consequentialism may seem to have an advantage over Act Consequentialism in that it handles co-ordination cases in a better way.[1] But, as we shall see, Rule Consequentialism is also vulnerable to problems with co-ordination in case there is a tie between several optimal rules (optimal in the sense that they are at least as good as any alternative rule).[2]

In their subjective, expectational form, Act and Rule Consequentialism can be stated as follows:[3]

* I would be grateful for any thoughts or comments on this paper, which can be sent to me at *johan.eric.gustafsson@gmail.com*.

[1] Act Consequentialism is the oldest form of consequentialism. It is the form defended by Bentham (1970, pp. 12–13). Rule Consequentialism was first suggested by Harrod (1936, pp. 148–9). The 'Act'/'Rule' terminology is due to Brandt (1959, pp. 380, 396).

[2] Following Sen's (1997, p. 756; 2017, p. xxix) definition of optimality.

[3] Timmons (2013, pp. 117, 155) states Act and Rule Consequentialism with an added 'and because' after 'if and only if'. I am leaving it out, since we may accept the biconditional and still hold that the moral status of acts is justified in some other way. (See Gustafsson 2021, pp. 264–5.) This won't matter for our discussion, however.

Act Consequentialism An act is right if and only if the expected value of its outcome is at least as good as the expected value of the outcome of any alternative act.

Rule Consequentialism The value of a rule is equal to the expected value of that rule being universally accepted.[4] An act ought to be done if it is required by a rule that is better than any other rule.

(It will soon become clear why Rule Consequentialism is, not yet, stated in terms of necessary and sufficient condition for rightness.) The problem we will discuss in this paper also occur for objective versions of Rule Consequentialism where rules are evaluated by the value of what would be the outcome if the rule were universally accepted.[5]

As mentioned, one of the main attractions of Rule Consequentialism is that it seems to do better than Act Consequentialism in co-ordination cases — that is, it seems to do better by the Consequentialist Tenet:[6]

[4] Some versions of Rule Consequentialism, such as Hooker's (2000, p. 32) version, assess rules not by the outcomes of their being universally accepted but by something less universal, such as the outcome of their 'internalization by the overwhelming majority of everyone everywhere in each new generation'. Smith (2010, p. 418) avoids the vagueness and arbitrariness of 'overwhelming majority' by letting the relevant acceptance level be the optimal acceptance level. But these complications won't matter in the two-agent examples we will consider. Moreover, some version of Rule Consequentialism focus on sets of rules or codes of rules. (See, for example, Barnes 1971, p. 57 and Hooker 2000, p. 32.) Here, we assume that these sets (or codes) of rules can be combined into a single rule that only requires that agents follow all the rules in the set. Finally, it matters a great deal for Rule Consequentialism whether we consider the outcomes of the rules being universally *accepted* or universally *followed* when we assess their optimality. See, for example, Parfit 2011, pp. 405–7. For the discussion in this paper, however, the distinction doesn't matter. We can collapse the distinction in at least one direction by assuming that, in the examples we will consider, the agents would follow a rule if they were to accept it.

[5] And the solution proposed in this paper should work equally well for the objective versions.

[6] Harsanyi 1977a, pp. 36, 38–41; 1985, p. 48, Regan 1980, p. 83, and Parfit 1986, p. 867. For the Consequentialist Tenet, see Parfit 1984, p. 24. You may, of course, accept Rule Consequentialism for other reasons than this tenet. You may reject the tenet and take Rule Consequentialism to be justified by the way it puts your considered judgements in reflective equilibrium. Hooker (2007, pp. 517–18), for example, rejects any overriding commitment to maximizing the good in his version of Rule Consequentialism. Still, Hooker (2007, p. 517) accepts an overriding commitment to prevent disasters. But then, as Hooker (2007, pp. 83–4) points out, we still need a solution to the Problem of Ties. To see this, note that we can make the cost of non-coordination between optimal rules as disastrous as we like, by replacing the 0:s in the cases we will discuss by an arbitrarily

The Consequentialist Tenet The ultimate moral aim is that things go as well as possible.

Consider the following case, where agents Row and Column have a choice between phi-ing and psi-ing:[7]

Case One

		Column	
		phi-ing	psi-ing
Row	phi-ing	4	0
	psi-ing	0	3

In this case, we get an outcome of 4 units of value if both agents choose to phi, and we get an outcome of 3 units of value if both agents choose to psi. But, if one of the agents chooses to phi while the other chooses to psi, we get an outcome of 0 units of value.

(For simplicity, we will only apply rules on single-shot games between two agents. You can assume that the rules gives the same results in all other interactions except the case under consideration or that we have a two-person universe where these games represent the only choices that will ever be made.[8] The underlying issues with the Problem of Ties, and its solution, will be the same if we apply rules to all choices in a world or a history without these assumptions, but the examples would be unnecessarily complex.)

Suppose that both Row and Column choose to psi in this case. Row and Column act independently. So, if one of them had chosen to phi, the other agent would still have chosen to psi. Hence, according to Act Consequentialism, they both did what they ought to do — in fact, it was the case for both of them that psi ought to be done and that phi is wrong.[9]

Psi-ing may seem like the wrong recommendation: Row and Column, we may assume, are two conscientious consequentialists who are trying to achieve optimal outcomes. So one may think that consequentialism

low number (see note 23). In what follows, however, we will grant that Rule Consequentialism is motivated by the Consequentialist Tenet.

[7] Luce and Raiffa 1957, p. 107, Schelling 1960, p. 293, Gibbard 1965, pp. 214–15, Barnes 1971, p. 61, Regan 1980, p. 18, and Parfit 1984, p. 72; 1986, p. 867. This case, like the other cases we will discuss, is a *pure coordination game*, since all agents have exactly the same goals. See Schelling 1960, pp. 84–5, Lewis 1969, p. 14, and Binmore 1992, p. 296.

[8] Following Gibbard 1965, p. 214.

[9] Gibbard 1965, p. 215, Barnes 1971, p. 61, and Regan 1980, pp. 18–19.

should guide them both towards phi-ing.[10] Rule Consequentialism gets us this result. The best rule in Case One is *Rule Phi* that prescribes phi-ing.[11] Accordingly, Rule Consequentialism entails that both Row and Column ought to phi, which would result in the optimal outcome.[12]

But the above formulation of Rule Consequentialism is incomplete. It doesn't tell us what to do in case there are two or more optimal rules.[13] This is the *Problem of Ties*: What should we do when two optimal rules prescribe different, incompatible acts?[14] What makes this problem all the

[10] In defence of Act Consequentialism, we should also note that, given that the other agent acts wrongly (perhaps by mistake), following the prescription to phi would have bad consequences. And then it seems that it is Act Consequentialism, rather than Rule Consequentialism, that does better by the Consequentialist Tenet. See Gibbard 1965, pp. 217–18 and Parfit 2011, p. 313. So, in Case One, if Row knows (or has reason to believe) that Column will psi, then it seems that Row should also psi. Rule Consequentialism, however, may also yield this result if we include Row's knowledge in the description of Row's situation so that the rules can take this knowledge into account. Then it may be that, given this knowledge in Case One, the optimal rules may imply that Row should psi. Yet, if the rules take into account the agents' knowledge or credences about what the other agents will do, there is a threat of a collapse of Rule Consequentialism into Act Consequentialism; see Lyons 1965, pp. 136–9. To avoid collapse, we will assume that the rules don't take into account the agents' knowledge or credences about what the other agents will do.

[11] Given that there are limits on the complexity of the rules we can accept, the best rule in one situation could conflict with the rule that is best generally for all situations. In order to let us focus on simple cases, we will ignore this limit on complexity in this paper. It doesn't make any crucial difference for the Problem of Ties. The difference is that we would have to consider ties between rules considered across all situations. But the problem, and the solution, would be the same. (Alternatively, we could assume that each of the cases we will consider represents the only choices that will ever be made in some possible universe.)

[12] Gibbard 1965, pp. 215–16 and Barnes 1971, p. 62. This rebuts Brandt's (1963, p. 121) claim that Rule Consequentialism would be extensionally equivalent to Act Consequentialism. Yet see note 10.

[13] Parfit (2011, p. 378) raises the issue, but he (2011, pp. 407–8) punts on the question of how to solve it.

[14] There is an analogous problem for Kant's (*GMS* 4:421; 2011, p. 71)

> *Formula of Universal Law* Act only according to that maxim through which you can at the same time will that it become a universal law.

It's unclear what you should do in case there are two or more incompatible maxims each of which you could will that it become a universal law; see Singer 1961, p. 296. Likewise, there is an analogous problem for *Functional Decision Theory*, which (very roughly) tells you to follow the decision algorithm that, compared to any other decision algorithm, would have have the greatest expected utility if it were followed generally. But what should you do in case there are two or more decision algorithms that are optimal

more pressing is that it is, in fact, very likely that there are two or more optimal rules. To see this, consider the following case, where Row and Column have a choice between phi-ing and psi-ing:[15]

Case Two

		Column	
		phi-ing	chi-ing
Row	phi-ing	4	0
	chi-ing	0	4

It seems that an optimal rule must either prescribe phi-ing or prescribe chi-ing; but it cannot permit each of phi-ing and chi-ing. Yet, given that an optimal rule prescribes one of these acts, there has to be a second rule that is just like the first except that it prescribes the opposite act in cases structured like Case Two. Given that the first rule is optimal, the second rule should be so too. So there is no uniquely optimal rule.[16]

One may think that there is a trivial fix to the Problem of Ties, namely,

> *The Indifference Criterion* An act is right if and only if it is permitted by a rule that is least as good as any other rule.[17]

The trouble is that the Indifference Criterion allows a lack of co-ordination between the optimal rules, which may have disastrous consequences.[18] Consider following case, where Row and Column have a choice between phi-ing, chi-ing, and psi-ing:[19]

Case Three

		Column		
		phi-ing	chi-ing	psi-ing
	phi-ing	4	0	0
Row	chi-ing	0	4	0
	psi-ing	0	0	3

in this sense but which result in different incompatible decisions? It is beyond the scope of this paper whether my proposal can be adapted to handle the Problem of Ties for these other theories.

[15] Schelling 1960, p. 294, Gibbard 1965, p. 218, and Parfit 1986, p. 867.

[16] Sobel 1968, pp. 153–6.

[17] Sobel 1968, p. 156, Barnes 1971, p. 57, and Card 2007, p. 253.

[18] Regan 1980, p. 91 and Hooker 2008, pp. 83–4.

[19] Schelling 1960, p. 296 and Regan 1980, p. 194.

Here, there are two optimal rules: Rule Phi, which prescribes phi-ing, and *Rule Chi* that prescribes chi-ing. For each of these rules, universal acceptance of the rule leads to an outcome of 4 units of value. But, since the Indifference Criterion allows that agents follow either of these rules, it doesn't guarantee that Row and Column follow the same rule. Assuming that Row and Column act independently and that they're equally likely to perform any right act, the expected value of everyone following the Indifference Criterion is just 2 units.[20] This is suboptimal, because there is a rule that uniquely leads to an outcome of 3 units of value, namely, *Rule Psi* that prescribes psi-ing.

Can we handle ties in a better way? One idea for how to break the tie between Rule Phi and Rule Chi is to discriminate between them by their closeness to established conventions. Consider

> *The Convention Criterion* An act is right if and only if it is permitted by a rule that is better than any other rule or, if there are two or more optimal rules, the act is permitted by an optimal rule that is as close to conventional morality as any other optimal rule.[21]

Even though this proposal may break ties between many rules, it cannot help us in case two or more optimal rules are at least as close to conventional morality as any optimal rule or in case there are two or more incompatible conventions that seem equally relevant.[22]

Another worry is that letting conventional morality (or any other convention) determine what acts are right is a significant departure from consequentialism.[23] Consequentialism is here taken to be the view (echoing the Consequentialist Tenet) that our evaluative focal points (act/rules)

[20] Harsanyi (1987, p. 17) suggests, similarly, that one uses the mixed strategy with an equal probability of each optimal strategy. In a fully symmetric case, we may plausibly assume that each agent is equally likely to follow each of the permitted rules.

[21] Hooker 2000, pp. 2–3, 32.

[22] Bykvist 2010, pp. 151–2.

[23] Card 2007, p. 253 and Bykvist 2010, p. 152. Given Hooker's (2007, pp. 517–18) rejection of any overriding commitment to maximizing the good, he may not think this departure is a problem. But, since he accept a commitment to disaster prevention, he is still open to the objection that there may also be a tie between the optimal rules in their closeness to conventional morality. Consider the following variation of Case Three where we suppose that the two optimal rules, Rule Phi and Rule Chi, are equally close to conventional morality:

are assessed morally based on the value of their potential outcomes.[24] Conventional morality is not based on the value of the consequences of rules or actions. So making the rightness of acts depend on conventional morality makes the theory non-consequentialist.

A related idea is to let salience serve as a tie-breaker:[25]

> *The Salience Criterion* An act is right if and only if it is required by a rule that is better than any other rule or, if there are two or more optimal rules, the act is required by one of the optimal rules that is at least salient as any other optimal rule.

Consider following case where Row and Column have a choice between surrendering, charging, ambushing, and waiting:

Case Four

		Column			
		surrendering	charging	ambushing	waiting
Row	surrendering	3	0	0	0
	charging	0	3	0	0
	ambushing	0	0	3	0
	waiting	0	0	0	2

*Case Three**

		Column		
		phi-ing	chi-ing	psi-ing
Row	phi-ing	4	−100	0
	chi-ing	−100	4	0
	psi-ing	0	0	3

Here, since phi-ing and chi-ing both satisfy the Convention Criterion, the agents may follow either of these rules. Assuming that Row and Column act independently and that they're equally likely to choose any of the two permitted acts, the expected value of the Convention Criterion is −48 units of value, which we may count as a disaster. Alternatively, we can say that the Convention Criterion has a one-in-two risk of a outcome of −100 units of value, which we may count as a one-in-two chance of disaster. But, if we instead follow Rule Psi as the Second-Order Indifference Criterion (presented later) prescribes, there would be no chance of disaster (that is, there would be no chance of a negative outcome) and the expected value would be 3 units of value. Hence the Convention Criterion conflicts with Hooker's commitment to disaster prevention.

[24] The term 'evaluative focal points' is due to Kagan (2000, p. 134).

[25] Schelling 1960, pp. 54–8, Lewis 1969, p. 35, Gauthier 1975, pp. 207–13, and Regan 1980, pp. 194–5.

Here, there are three optimal rules: *Rule Surrender* that prescribes surrendering, *Rule Charge* that prescribes charging, and *Rule Ambush* that prescribes ambushing. One of these rules sticks out. Among the optimal rules, Rule Surrender is the only peaceful rule. Therefore, since this rule is salient, Row and Column ought to surrender according to the Salience Criterion, which results in an optimal outcome.

A problem with this solution is that (like the Convention Criterion) it departs from consequentialism. Whether a rule is salient (or peaceful) shouldn't matter in a consequentialist theory.[26] Moreover, our intuitive sense of what is salient seems hard to codify in a precise manner.[27] So the Salience Criterion would introduce a lot of vagueness to the recommendations of Rule Consequentialism. Furthermore — like earlier criteria — the Salience Criterion is open to the problem that some ties aren't broken as two optimal rules can be equally salient.[28]

Note that we're discussing the subjective, expectational form of Rule Utilitarianism. Hence we wouldn't want to solve the Problem of Ties with the following objective criterion:

> *The Objective Criterion* An act is right if and only if it is permitted by a rule that is better than any other rule or, if there are two or more optimal rules, an optimal rule that would have an at least as good outcome given what other agents will in fact do as any other optimal rule.

This proposal depends on information that agents needn't possess, namely, what other agents will in fact do. Assuming that, in the cases

[26] For the same reason, we can rule out tie-breaking rules, based on the labels or the presentation of the alternatives, as suggested by Schelling (1960, pp. 300–1) in a different context.

[27] Regan 1980, p. 193.

[28] Another idea would be to borrow some ideas from the related problem in game theory of how to single out one of many *Nash equilibria* — combinations of strategies such that each agent's strategy is optimal for them given the other agents' strategies (see Nash 1950b, p. 49). Harsanyi's tracing procedure (in Harsanyi 1975; 1977b, pp. 214–219; 1980, pp. 196–201; 1987 and Harsanyi and Selten 1988, pp. 131–241) is one potential candidate. While the idea of using this procedure to resolve the Problem of Ties for Rule Consequentialism is intriguing, it's doesn't seem like the tracing procedure fits with the motivation for Rule Consequentialism. It depends crucially on the agents' initial credence distribution in other agents' strategies. For instance, unless you have a 3/7 credence that the other agent will phi, the tracing procedure will recommend that you chi in Case One. This is worse than the other variants of Rule Consequentialism we discuss, which all invariably recommend that you phi in Case One.

we discuss, the agents do not know what the other agents will do, they cannot rely on the Objective Criterion to break ties between rules.[29]

Nevertheless, we can get around this problem by instead relying on the agent's expectations about what other agents will do. That is, we could use the subjective form of Act Consequentialism as a tie-breaker:[30]

> *The Act-Consequentialist Tie-Breaker Criterion* An act is right if and only if either (i) it is permitted by all optimal rules or (ii) it is permitted an optimal rule and the expected value of the outcome of the act is at least as good as the expected value of the outcome of any alternative act that is permitted by an optimal rule.

The trouble with this principle is similar to that of earlier criteria. It need not break all ties. Consider, for instance, Case Three. Given the symmetrical structure of this case, there seems to be no reason for either agent to believe that the other agent is more likely to phi rather than to chi, and vice versa. Yet, given that Row finds it equally likely that Column will phi as that Column will chi, the expected value of the outcome of phi-ing for Row will be the same as the expected value of the outcome of chi-ing. And the same holds for Column, changing what needs to be changed. Hence we still have a tie between Rule Phi and Rule Chi. Accordingly, the Act-Consequentialist Tie-Breaker Criterion is open to the same objection as the Indifference Criterion.

To avoid these problems and still achieve better outcomes than the Indifference Criterion, I suggest a second-order approach. Consider

[29] See also note 10.

[30] Smart 1956, p. 345 and Singer 1961, p. 205.

> *The Second-Order Indifference Criterion* Partition all rules into groups by their value.[31] And, for each group, replace the rules in that group by the rule that permits an act if and only if that act is permitted by some rule in the group. Now, consider the best of these replacement rules. Among these optimal replacement rules, let the *prime replacement rule* be the optimal replacement rule that replaced the best rules. An act is right if and only if it is permitted by the prime replacement rule.

The idea behind this criterion, like the idea behind the Indifference Criterion, is that, when two rules are equally good, we have no consequentialist grounds for favouring one of them over the other. The Indifference Criterion then tells us, essentially, to follow the rule that permits an act if and only if it is permitted by one of the optimal rules. The twist for this second-order criterion is that we then go on to check whether this rule is dominated by some other rule that, for a certain value, permits an act if and only if it is permitted by one of the rules of that value. Then if there is a tie in this second-order comparison, we break the tie with the first-order comparisons. Hence we're able to break all second-order ties without regress.

To see how the Second-Order Indifference Criterion works, consider Case Three. The optimal rules are Rule Phi and Rule Chi, each of which results in an outcome of 4 units of value. The third best rule is Rule Psi which results in an outcome of 3 units of value. And the fourth best rule is *Rule Phi-or-Chi* that prescribes either phi-ing or chi-ing (that is, each of phi and chi is permitted, but psi is prohibited). Assuming that Row and Column act independently and that they're equally likely to choose any permitted act, the expected value of everyone following Rule Phi-or-Chi is 2 units. (The assumption that each permitted act is equally likely to be chosen may seem arbitrary. If we had access to some non-arbitriry probabilites, we could use them instead. But, if not, an equal distribution seems the least arbitrary.[32]) We partition all rules by their value, so we

[31] A variation worth considering is to partition all rules into groups not only by their value but also by whether their prescriptions corresponds to a Nash equilibrium (see note 28) given that the agents' utilities matches the moral value of the outcomes. So, on this *Nash-Differentiating* variant, rules are in the same group if and only if they have the same value and either all of them are equilibria or none of them are equilibria. If there is a tie between the replacement rules for two groups of rules of the same value, the replacement for the rules that are equilibria is favoured.

[32] As long as we have a non-arbitrary way of individuating alternative acts, we should

have the groups {Rule Phi, Rule Chi}, {Rule Psi}, {Rule Phi-or-Chi}, and so on. Now, for each group, we replace the rules in that group with the rule that permits an act if and only if the act is permitted by some rule in the group. If there is only one rule in a group, the rule is trivially replaced by itself. So we replace the rules in group {Rule Phi, Rule Chi} with Rule Phi-or-Chi. We trivially replace the rule in group {Rule Psi} with Rule Psi. Likewise, we trivially replace the rule in group {Rule Phi-or-Chi} with Rule Phi-or-Chi. And so on for all groups. Considering just the resulting new set of rules, we have that Rule Psi is the best rule (followed by Rule Phi-or-Chi). Accordingly, Rule Psi is the prime replacement rule. So we find that both Row and Column ought to psi, which results in an outcome of 3 units of value. Hence the expected value of the Second-Order Indifference Criterion is greater than expected value of the Indifference Criterion, whose outcome had an expected value of just 2 units.

Note that the expected value of the Second-Order Indifference Criterion in Case Three still falls short of the expected value of the optimal rules, Rule Psi and Rule Chi.[33] But, since there is no consequentialist basis for singling out one of these rules rather than the other, this ideal is unattainable given the resources of any version Rule Consequentialism that doesn't depart from consequentialism.

It may be objected that we should also have considered rules that prescribe probabilistic mixtures of phi, chi, and psi — that is, rules that prescribe that the agents choose to randomly do one these acts with some specific probability distribution.[34] But, if the agents were able to perform such probabilistic acts, those probabilistic acts should be represented as further available acts, rather than as probabilistic rules over regular acts.[35] (Still, the Second-Order Indifference Criterion would also work if we allowed probabilistic rules, but including those rules would complicate the examples.[36])

---

avoid the standard counter-examples to the Principle of Indifference. See Keynes 1921, pp. 42–44, and van Fraassen 1989, pp. 302–317.

[33] This may make Rule Consequentialism, to some extent, self defeating. Rule Consequentialism gives us the aim to follow optimal rules. But, if Rule Consequentialism is itself treated as a rule, it wouldn't be optimal. For a discussion of self-defeating theories, see Parfit 1984, pp. 3–114.

[34] See, for example, Regan 1980, pp. 196–197.

[35] In game theory, following von Neumann and Morgenstern (1944, pp. 143–144), probabilistic acts are called 'mixed strategies'.

[36] If we include probabilistic rules, then the Nash-Differentiating Second-Order Indifference Criterion is more plausible (see note 31). In Case Three, for example, there

To see why we may need to break ties between replacement rules, consider the following case where Row and Column have a choice between phi-ing, chi-ing, psi-ing, xi-ing, and pi-ing:

Case Five

		Column				
		phi-ing	chi-ing	psi-ing	xi-ing	pi-ing
	phi-ing	75	0	0	0	0
	chi-ing	0	75	0	0	0
Row	psi-ing	0	0	75	0	0
	xi-ing	0	0	0	50	0
	pi-ing	0	0	0	0	50

Here, there are three optimal rules: Rule Phi, Rule Chi, and Rule Psi, which have an outcome of 75 units of value. The next best rules are *Rule Xi* that prescribes xi-ing and *Rule Pi* that prescribes pi-ing, which have an outcome of 50 units of value. We partition all rules by their value, so we have the groups {Rule Phi, Rule Chi, Rule Psi}, {Rule Xi, Rule Pi}, and so on. We replace {Rule Phi, Rule Chi, Rule Psi} with *Rule Phi-or-Chi-or-Psi* that permits Phi, Chi, and Psi but no other acts. We replace {Rule Xi, Rule Pi} with *Rule Xi-or-Pi* that permits Xi and Pi but no other acts. And so on for all groups. Comparing these replacement rules, we have two optimal rules: Rule Phi-or-Chi-or-Psi and Rule Xi-or-Pi, each of which has an expected value of 25 units. Allowing the agents to follow both these rules is equivalent to allowing them to follow the rule that permits any act in this case, which has an expected value of 13 units. This is a lower expected value than that of requiring the agents to following just one of these rules. So we should break this tie. We do so by favouring the optimal replacement rule that replaced the best rules.[37] Hence we favour Rule Phi-or-Chi-or-Psi since it replaced better rules than Rule Xi-or-Pi.

would be lots of probabilistic rules that prescribe mixtures of phi, chi, and psi which would be equally good as Rule Psi. But, unlike Rule Psi, they wouldn't correspond to a Nash equilibrium. So the Nash-Differentiating variant of the Second-Order Indifference Criterion would still entail that the agents ought to psi — whereas, on the standard version with probabilistic rules, both phi and chi would be right.

[37] We could, of course, break the tie the other way by favouring the the optimal replacement rule that replaced the worst rules. Doing so would have an outcome of the same expected value. So why break the tie one way rather than another? It seems more fitting, in a non-arbitrary way given consequentialism, to favour the replacement rule that replaced (and corresponds to) the better rules. We are then using first-order consequentialist dominance to break second-order consequentialist ties.

The Second-Order Indifference Criterion has one peculiar property that the Indifference Criterion lacks: The Second-Order Indifference criterion violates

> *Contraction Consistency* If an act is right in a situation, it would still be right if we removed some of the other available acts.[38]

To see this, compare Case One and Case Three. Case One is just like Case Three except that chi-ing is unavailable. According to the Second-Order Indifference Criterion, phi-ing is right in Case Three but not right in Case One, which may seem strange.

But note that violations of Contraction Consistency are unavoidable if we accept each of the following claims:

(1) acts that are prescribed by uniquely optimal rules are right,

(2) the rightness of acts only depends on the value of the consequences of acts and rules, and,

(3) if randomly following one of multiple optimal rules has a worse expectation than following the uniquely next-best rule, then it is not the case that each act that is prescribed by one of these optimal rules is right.

From (1), it follows that phi-ing is right in Case One. From (2), it follows find that phi-ing and chi-ing in Case Three are either both right or both not right. From (3), it follows find that phi-ing and chi-ing in Case Three are not both right. Hence phi-ing is not right in Case Three. But then phi-ing is right in Case One but not in Case Three, which violates Contraction Consistency. Rule Utilitarianism commits us to (1), consequentialism commits us to (2), and, to avoid doing worse in expectation than the Second-Order Indifference Criterion, we need (3).

Why should we favour the Second-Order Indifference Criterion, which violates Contraction Consistency, over the Indifference Criterion, which violates (3)? From a consequentialist perspective, the Second-Order Indifference Criterion has a significant advantage over the Indifference Criterion: The Second-Order Indifference Criterion may lead to outcomes with a greater expected value and never to outcomes with lower expected value than the Indifference Criterion.

[38] Nash 1950a, p. 159.

And the advantage of the Second-Order Indifference Criterion over the Convention Criterion and the Salience Criterion is that, unlike the latter two, the Second-Order Indifference Criterion does not depart significantly from consequentialism.

I wish to thank Krister Bykvist, Tomi Francis, Frederik Van De Putte, Wlodek Rabinowicz, and Alexandros Rigos for valuable comments.

References

Barnes, Gerald (1971) 'Utilitarianisms', *Ethics* 82 (1): 56–64.

Bentham, Jeremy (1970) *An Introduction to the Principles of Morals and Legislation*, eds. J. H. Burns and H. L. A. Hart, The Collected Works of Jeremy Bentham, London: Athlone.

Binmore, Ken (1992) *Fun and Games: A Text on Game Theory*, Lexington, MA: Heath.

Brandt, Richard B. (1959) *Ethical Theory: The Problems of Normative and Critical Ethics*, Englewood Cliffs, NJ: Prentice-Hall.

— (1963) 'Toward a Credible Form of Utilitarianism', in Hector-Neri Castañeda and George Nakhnikian, eds., *Morality and the Language of Conduct*, pp. 107–143, Detroit: Wayne State University Press.

Bykvist, Krister (2010) *Utilitarianism: A Guide for the Perplexed*, London: Continuum.

Card, Robert F. (2007) 'Inconsistency and the Theoretical Commitments of Hooker's Rule-Consequentialism', *Utilitas* 19 (2): 243–258.

Gauthier, David (1975) 'Coordination', *Dialogue* 14 (2): 195–221.

Gibbard, Allan F. (1965) 'Rule-Utilitarianism: Merely an Illusory Alternative?', *Australasian Journal of Philosophy* 43 (2): 211–220.

Gustafsson, Johan E. (2021) 'Utilitarianism without Moral Aggregation', *Canadian Journal of Philosophy* 51 (4): 256–269.

Harrod, R. F. (1936) 'Utilitarianism Revised', *Mind* 45 (178): 137–156.

Harsanyi, John C. (1975) 'The Tracing Procedure: A Bayesian Approach to Defining a Solution for *n*-Person Noncooperative Games', *Journal of Game Theory* 4 (2): 61–94.

— (1977a) 'Rule Utilitarianism and Decision Theory', *Erkenntnis* 11:25–53.

— (1977b) 'A Solution Concept for *n*-Person Noncooperative Games', *International Journal of Game Theory* 5 (4): 211–255.

— (1980) 'Uses of Bayesian Probability Models in Game Theory', in D. H.

Mellor, ed., *Science, Belief and Behaviour: Essays in Honour of R. B. Braithwaite*, pp. 189–201, Cambridge: Cambridge University Press.
— (1985) 'Does Reason Tell Us What Moral Code to Follow and, Indeed, to Follow Any Moral Code at All?', *Ethics* 96 (1): 42–55.
— (1987) 'The Tracing Procedure: A Self-Correcting Reasoning Procedure', *Theory and Decision* 23 (1): 7–23.
Harsanyi, John C. and Reinhard Selten (1988) *A General Theory of Equilibrium Selection in Games*, Cambridge, MA: MIT Press.
Hooker, Brad (2000) *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*, Oxford: Clarendon Press.
— (2007) 'Rule-consequentialism and Internal Consistency: A Reply to Card', *Utilitas* 19 (4): 514–519.
— (2008) 'Rule-Consequentialism versus Act-Consequentialism', *Politeia* 24 (90): 75–85.
Kagan, Shelly (2000) 'Evaluative Focal Points', in Brad Hooker, Elinor Mason, and Dale E. Miller, eds., *Morality, Rules, and Consequences: A Critical Reader*, pp. 134–155, Edinburgh: Edinburgh University Press.
Kant, Immanuel (2011) *Groundwork of the Metaphysics of Morals: A German-English Edition*, ed. Jens Timmermann, Cambridge: Cambridge University Press.
Keynes, John Maynard (1921) *A Treatise on Probability*, London: Macmillan.
Lewis, David (1969) *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.
Luce, R. Duncan and Howard Raiffa (1957) *Games and Decisions: Introduction and Critical Survey*, New York: Wiley.
Lyons, David (1965) *Forms and Limits of Utilitarianism*, Oxford: Clarendon Press.
Nash, John F. Jr. (1950a) 'The Bargaining Problem', *Econometrica* 18 (2): 155–162.
— (1950b) 'Equilibrium Points in *n*-Person Games', *Proceedings of the National Academy of Sciences of the United States of America* 36 (1): 48–49.
Parfit, Derek (1984) *Reasons and Persons*, Oxford: Clarendon Press.
— (1986) 'Comments', *Ethics* 96 (4): 832–872.
— (2011) *On What Matters Volume One*, Oxford: Oxford University Press.
Regan, Donald (1980) *Utilitarianism and Co-Operation*, Oxford: Clarendon Press.
Schelling, Thomas S. (1960) *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.

Sen, Amartya K. (1997) 'Maximization and the Act of Choice', *Econometrica* 65 (4): 745–779.

— (2017) *Collective Choice and Social Welfare: An Expanded Edition*, Cambridge, MA: Harvard University Press.

Singer, Marcus George (1961) *Generalization in Ethics: An Essay in the Logic of Ethics, with the Rudiments of a System of Moral Philosophy*, New York: Alfred A. Knopf.

Smart, J. J. C. (1956) 'Extreme and Restricted Utilitarianism', *The Philosophical Quarterly* 6 (25): 344–354.

Smith, Holly M. (2010) 'Measuring the Consequences of Rules', *Utilitas* 22 (4): 413–433.

Sobel, Jordan Howard (1968) 'Rule-Utilitarianism', *Australasian Journal of Philosophy* 46 (2): 146–165.

Timmons, Mark (2013) *Moral Theory: An Introduction*, Lanham: Rowman & Littlefield, second edn.

van Fraassen, Bas C. (1989) *Laws and Symmetry*, Oxford: Clarendon Press.

von Neumann, John and Oskar Morgenstern (1944) *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.