Utilitas (2025), 1–14
doi:10.1017/S0953820825000111

CAMBRIDGE UNIVERSITY PRESS

ARTICLES

Rule Consequentialism and Demandingness: The Wrong Solution(s)?

Andrea Sauchelli[1,2]

[1]Department of Philosophy, Lingnan University, Hong Kong and [2]Hong Kong Catastrophic Risk Centre, Lingnan University, Hong Kong
Email: andreasauchelli@ln.edu.hk

Abstract

A textbook objection to consequentialism is that it is too demanding—on the assumption that a moral theory which is excessively demanding thereby loses plausibility. In this paper, I assess whether the mechanisms employed by two versions of rule consequentialism, those of Brad Hooker and Tim Mulgan, adequately meet the requirement of not being too demanding. I also examine whether the concept of human nature might help determine what should count as demanding for a moral theory. While this suggestion also faces significant challenges, I contend that prescribing less partiality towards the present generation may not be a drawback for the consequentialist frameworks under consideration.

Keywords: rule consequentialism; demandingness objection; partiality; catastrophic risks; human nature

'Enough is enough' applies also to morality, at least in the sense that, according to some philosophers, a plausible moral theory should not make excessive demands. The plausibility of this intuition depends on how we understand 'enough' and 'excessive', and upon whom these demands are being made. Still, an objection based on it is familiar among the staple criticisms of utilitarianism and other forms of consequentialism. More precisely, critics of this family of normative theories think that injunctions to be fully impartial or to maximise the (expected) goodness our actions produce – typical features of consequentialist theories – are, at least in some cases, too demanding. For example, it would be excessive to prescribe that agents should be completely impartial and maximise the overall good in all of their everyday choices. As such, normative theories including such excessive prescriptions are implausible, or at least less plausible than alternative views. This is the basic idea of the demandingness objection to consequentialism. In this paper, I discuss this influential objection with a particular focus on rule consequentialism (RC), which some philosophers consider capable of answering the objection.[1] More

[1]D. Miller (2014) and Hooker (2023) are excellent recent introductions to rule utilitarianism and to RC more generally. In this paper, I understand the former as a version of the latter. I will not assume any specific

specifically, I focus on Brad Hooker's and Tim Mulgan's versions of RC and on the mechanisms their theories employ to ensure that the moral verdicts they generate are not too demanding.[2]

According to both Hooker's and Mulgan's versions of RC, the deontic status of what can be legitimately evaluated from a moral perspective is determined by a code of rules or a moral outlook.[3] Their views differ, however, on what determines which code or outlook is the best or optimific one.[4] To a first approximation, in Hooker's view the optimific code is determined by comparing the results of cost–benefit analyses that are based on the *acceptance* of various possible codes by the majority of a relevant group. The code that would have the best consequences if accepted by the vast majority of a target group is the one that determines the moral status of actions, character traits, and so on. Given its reliance on the notion of acceptance (more on this later), I call this a variety of acceptance-based RC (AB-RC).

On the other hand, Mulgan's RC suggests that the best outlook is the one it would be best to *teach* to the next generation: a version of teaching-based RC, or TB-RC. Both theories, as forms of consequentialism, have to address the demandingness objection as they include features that may expose them to it: for example, prescriptions to maximise the good. Both theories have mechanisms to answer this objection; Section 2 and its subsections outline them. Section 1 clarifies the distinction between two dimensions of demandingness and their application to RC.

In Section 3, I discuss three arguments against the claim that the two theories' mechanisms successfully address the objection. The conclusion of the first argument is that, given the ambitions of at least Mulgan's view, the theories may turn out to be demanding after all. The second objection questions the mechanisms' effectiveness from a different angle: in particular, I elaborate on the thought that accepting or learning a morally demanding view may not be as costly as the forms of RC in question may suggest. In the third and last objection, I challenge the relevance of the mechanisms themselves for determining what should count as demanding. In particular, I suggest that it is at most unclear whether these mechanisms track what should be properly regarded as a cost in the calculation of what is demanding. I then propose to deploy the notion of human nature to make the mechanisms salient (§3.2), although this would seem to ameliorate the problem of demandingness only partially.[5] Still, an underlying thought, further articulated in the conclusions, is that, given the severe challenges posed

theory of the good or of well-being, making the assumption that most of what is said would turn out to be compatible with almost any plausible version of such theories.

[2]Hooker has defended his preferred version of RC in Hooker (2000) and several other papers (e.g. Hooker and Fletcher 2008; Hooker 2020). For Mulgan's view, see Mulgan (2006), (2017), (2020). Other works focusing on these versions of RC include Miller (2021), Mosquera (2022), Woollard (2022), and Sauchelli (2025).

[3]Hooker uses the notion of a 'code of rules', whereas Mulgan deploys the more inclusive notion of a moral outlook. I will use the two terms interchangeably, leaving open the possibility that an outlook may include not only a code of rules but also, say, a set of virtues. I will leave aside the possible complication that an outlook may be more demanding to teach than a code of rules.

[4]I will understand 'the best or optimific' code or outlook as the code or outlook that maximise overall well-being.

[5]Most appeals (including mine) to such a notion will leave at least some philosophers dissatisfied: given the history of the concept and the variety of approaches used to study it, it is almost inevitable that some important studies will not be mentioned. See Kronfeldner et al (2014) for a recent survey.

https://doi.org/10.1017/S0953820825000111 Published online by Cambridge University Press

by the real possibility of 'broken futures', a more demanding morality is warranted – at least with respect to a significant reduction of our partiality for the present generation.[6]

1. Disentangling the demandingness objection

Although the demandingness objection is frequently interpreted as a reason to reject consequentialism (Hooker 2009; Mulgan 2020, p. 14; Woodard 2019, pp. 21–25), the objection (i) also applies to several other moral theories (e.g. Kantian ethics), and, in some cases, (ii) depends on rejecting principles that some consequentialists regard as independently plausible. The latter point is dialectically relevant when the objection is deployed as an argument to convince consequentialists that their theory is wrong. In fact, if much of the plausibility of the idea that consequentialism is too demanding relies on the implicit rejection of some independently plausible principles (e.g., impartiality), arguments that presuppose such a rejection are questionable. Instead of arguing that consequentialism should be rejected because it is too demanding, proponents of the demandingness objection may be better off arguing directly that these other underlying principles of consequentialism (e.g. that morality should be impartial) are implausible.[7] Still, even if we agree with Sobel (2007, 2020) that the demandingness objection may not be a persuasive critique of consequentialism – or, at least, that other theories may handle a similar criticism no better – the objection constitutes a helpful reminder for any plausible moral theory: Any moral theory should take into account the nature of the moral agents to which it purports to apply. More specifically, the demandingness objection would be helpful to indirectly highlight that any moral theory for human agents should not imply prescriptions, permissions, obligations, etc. that they cannot follow.

To evaluate the demandingness objection as a non-comparative argument against consequentialism, it is necessary to specify why certain moral prescriptions would be too demanding. In particular, to determine what is too demanding we should specify the different kinds of costs that a moral theory may legitimately impose (Sobel 2007; Woollard 2016). For instance, a prescription considered too demanding may demand that one perform a certain action-type (e.g. killing one's children), or just perform an action-type to a certain degree (e.g. donating all of one's own possessions to charity). In addition, a moral theory may impose (probable) costs upon moral patients as a result of what it permits – and such costs may be considered impermissible on independent grounds if they involve impermissible action types, or permissible actions but to an impermissible degree.

In some cases, supporters of the demandingness objection do not always clearly distinguish the various ways in which prescriptions, permission, etc. can be regarded as too demanding. In fact, sometimes consequentialism is considered too demanding because it cannot accommodate cases of allegedly permissible (or even required) partiality. For example, to the extent that consequentialism is committed to indifference

[6]See Mogensen (2021) for the claim that the moral demandingness objection levelled against utilitarianism has been misunderstood – at least in light of what the proper demands of future generations are reasonably expected to be.

[7]See Ashford (2003) and Hills (2010) for the claim that contractualist moral obligations are at least as demanding as those suggested by utilitarianism. See Formosa and Sticker (2019) for a Kantian conception of beneficence that seems to strike a balance between the extremes of over-demandingness and un-demandingness.

https://doi.org/10.1017/S0953820825000111 Published online by Cambridge University Press

in all cases between, say, saving one stranger and saving one's family member, it would be intolerably demanding.[8] This way of specifying why consequentialism may be too demanding should be distinguished from one of its other typical features: the prescription to maximise the good.[9] On this other understanding of demandingness, direct forms of consequentialism (other than, for example, satisficing consequentialism) would be too demanding because, for any situation we can influence, they require us to maximise the good. So, there are at least two ways in which consequentialism can be too demanding: because of its prescription to be (always) impartial – a demandingness of selection – or because of its prescription to maximise the good – a demandingness of measure.

These two specifications are orthogonal. For example, someone may object that consequentialism would be too demanding even if it were to allow significant degrees of partiality – including a significant degree of partiality towards oneself. In particular, a maximising form of consequentialism that allows significant partiality may still require the maximising of, say, one's own good – and this may be regarded as too demanding. Surely, the critic may argue, it should be morally permissible to refuse to maximise my own well-being, at least on some occasions. Similarly, a non-maximising form of consequentialism may be considered too demanding because of its impartiality – because of its restriction in selecting, for example, the recipients of our beneficence. Certainly, complete impartiality in a non-maximal distribution of resources between a mass murderer and one's own child – assuming a similar level of generated overall well-being – would demand too much of a parent.

2. Acceptance-based and teaching-based consequentialist answers

2.1 Acceptance-based consequentialism

Versions of RC deploy different strategies to rebut the demandingness objection. For example, Hooker's RC, ideal code RC or ICRC, includes two 'mechanisms': (i) the acceptance filter and (ii) a limit to the disaster condition.

The first mechanism works as follows: which ideal code is optimific depends on the benefits that arise from its acceptance by most individuals (90%) in a certain population (Hooker 1998, 2000, 2020).[10] More specifically, the ideal code for a group is the code that would produce the most overall well-being when accepted by most members. Crucially, ICRC stipulates that members of the relevant group are ordinary people, that is, people not equipped with super-human or saintly moral dispositions and capacities (e.g. the capacity to empathise with everybody at the same level). So, Hooker suggests, it is only to be expected that the benefits of the large acceptance of an ideal code for 'saints' may not

[8]Some degree of partiality towards one's own family members (an 'agent-centred prerogative') is just one of the many possible forms of partiality that are allegedly legitimate. Others include partiality towards colleagues, compatriots, team members, allies in war, party members and so on. Clearly, not all such forms of partiality are equally morally justifiable (see Hooker 2013; Jeske 2020).

[9]See Horta et al. (2022) on recent definitions of consequentialism.

[10]Both Brandt (1979) and Hooker (2000) choose an interpretation of acceptance whereby accepting or adopting a moral code entails its internalisation: being motivated to adhere to it, disapproving of those who violate it, believing that acting in line with the code is significant, and, overall, exhibiting a range of positive attitudes towards it that go beyond simply complying with it.

https://doi.org/10.1017/S0953820825000111 Published online by Cambridge University Press

make it up for the costs of having ordinary persons accept it.[11] The idea is that it might take such an exorbitant effort to ensure that the majority of a group of ordinary human beings acquired certain saintly dispositions (if it were even possible) that a less demanding code would turn out to have a greater net benefit overall.

Of course, were human agents different, a saintly code would perhaps be the optimific one. However, Hooker suggests that, in calculating which code is optimific, we should not consider the expected benefits of implementing such a code among *tabulae rasae* capable of moral sainthood, but instead among 'ordinary' human agents with a variable but limited amount of moral plasticity. In a recent exposition of his view, Hooker uses such terms as 'natural inclinations' to refer to aspects of our nature that should be considered when calculating the costs of implementing an ideal code (Hooker 2020, pp. 450–451). From his discussion, it seems that these relevant human inclinations include certain dispositions to be partial towards those we care about, as well as instincts of self-preservation and self-interest, at least to a certain degree. So, in an overall calculation of the expected value of a moral code's widespread acceptance, a code that strongly contrasts with these 'natural inclinations' is unlikely to be optimific.

Hooker also claims that any plausible version of consequentialism (and possibly any plausible moral theory) should take into account how much goodness some deviations from the prescribed code may produce. For instance, ICRC may imply that we have to commit a certain amount of time to helping others but, when non-compliance is higher than 10%, we may have to commit more than the rules would normally prescribe, at least when some additional sacrifice would produce a significant amount of overall benefit. The magnitude of these additional obligations is regulated by a requirement to amend the rules when doing so would 'prevent disaster' (Hooker 2000, p. 98; cf. Brandt 1992, p. 151); however, Hooker believes that this requirement should be restricted as well. This thought brings us to ICRC's second mechanism for curbing its alleged demandingness. Hooker claims that internalising a code that includes an unrestricted 'preventing disaster' requirement would be extremely costly – so much so that it may outweigh the benefits of including it. Deploying Cullity's (1995) distinction between aggregative and iterative cost, Hooker suggests that ICRC would include this rule or principle:

> Over time agents should help those in greater need, especially the worse off, even if the personal sacrifices involved in helping them add up to a significant cost to the agents. The cost to the agents is to be assessed aggregatively, not iteratively. (Hooker 2000, p. 166)

The last part of this rule is what would limit the demandingness of the 'preventing disaster' requirement: the costs of beneficence on any specific occasion, properly calculated, include the aggregated costs of what the agent has already done, and thus not solely the costs to the agent of a single action on a single occasion (cf. Dougherty 2017; Forcehimes and Semrau 2019; Murphy 1993). The idea is that small costs may add up to an unreasonable one during a lifetime. Thus, in calculating the costs of a particular action – even those aimed at preventing some disaster – we ought to include all the other sacrifices the agent has already made (Hooker 2000, p. 167; cf. Thomson 1971).

[11]Wolf (1982, p. 420) defines a moral saint as someone whose life is 'dominated by a commitment to improving the welfare of others or of society as a whole'.

https://doi.org/10.1017/S0953820825000111 Published online by Cambridge University Press

2.2 Teaching-based consequentialism

Tim Mulgan's teaching-based RC identifies the Ideal Moral Outlook (IMO) through a process of mapping 'moral outlooks' onto possible futures. First, moral outlooks are ranked according to the expected value of the consequences if we tried to *teach* them. Among all the outlooks mapped into possible futures, the best is the one that maximises the total expected well-being, after the costs of teaching it to the next generation are considered (Mulgan 2017, p. 291). A related RC theory ('Ideal Moral Outlook Consequentialism', or IMOC) would suggest that 'the right act is an act that would be performed by someone who had internalised the ideal code' (Mulgan 2017, p. 291). According to Mulgan, this formulation 'sets aside' the costs of changing existing moral beliefs, which would otherwise make some moral outlooks too demanding.

The main mechanism this version of TB-RC offers to curb the demandingness of the ideal code or outlook is similar to ICRC's: as there are limits to what the next (or any) generation of normal human beings can learn, the ideal outlook would not comprise rules that are too demanding. In fact, the reasoning goes, attempts to teach excessively demanding moral codes may require so great an effort and so many resources that less morally ambitious codes would bring about more well-being overall.

3. Demandingness, acceptability and teachability

3.1 Problems for ICRC and IMOC

ICRC and IMOC may turn out to be demanding for different reasons. For example, given the ambitions of Mulgan's IMOC, this theory may turn out to be demanding because it implies that certain forms of partiality are forbidden, or because it may require agents to maximise the overall good 'excessively'. In fact, although Mulgan claims that the learnability of certain outlooks imposes limits that curb the view's demandingness, the ideal outlook is still supposed to be justifiable to people living in 'broken futures' – credible scenarios in which basic resources have become scarce (Mulgan 2015, 2017, 2018; cf. Mogensen 2021). We may understand these scenarios – for example, a possible future in which climate change has made water and food scarce – as states of permanent near-disaster or emergency (Gardiner 2011). Now, it seems plausible that the practices we presently justify by our strong partial preferences for 'conspicuous consumption' are not ones we could imagine justifying to people living in futures of scarcity generated (at least partly) by such practices. In fact, in his recent work, Mulgan seems to accept that, given the risks humanity faces in the near future, significant aspects of our partiality for the present generation may be unjustifiable. It seems plausible that Mulgan would not think that forms of strong generational partiality could be justifiable to proximate future generations living in conditions of scarcity resulting from our actions.

In what follows, I will assume that the imaginative exercise of justifying our practices to people living in broken futures reveals the degree of reasonable partiality which a given theory tolerates, at least at an intergenerational level. The limits to what is teachable to individuals, on the other hand, constitute a mechanism that would limit both of the ways in which demandingness may vary. In light of these considerations, IMOC may face the demandingness objection head on: it would say that if the best moral theory turns out to be demanding in scenarios involving broken futures, then so much the worse for common-sense standards that are faulty and lax enough to allow us to be strongly partial towards the present generation – see the conclusions of this paper for an elaboration on this idea. Still, although Mulgan recognises that his theory may depart

https://doi.org/10.1017/S0953820825000111 Published online by Cambridge University Press

from common-sense morality, at least if common sense justifies a certain unjustifiable degree of partiality towards the present generation, he claims that his IMOC would not deviate too much from what common-sense morality seems to imply (Mulgan 2017, p. 298). The reason, again, is that the IMO would have to be taught, and human beings are not so indefinitely pliable that they will learn any sort of behaviour. As such, the IMO would at least involve teaching moral dispositions such as honesty, generosity, aversion to murder, and so on (Mulgan 2017, p. 298). However, critics of RC may reiterate their point by claiming that, at the least, ambitious versions of RC (such as Mulgan's) would indeed require a demanding form of (intergenerational) impartiality.

Another problem for ICRC and IMOC is that, when the alleged limits to teachability and acceptability are considered more carefully, these mechanisms may not deliver the expected results: they might not significantly limit the demandingness of their respective codes. Some examples suggest that human beings really can be made to believe, and do, what would appear to be against their interests – let alone against their partiality and, arguably, in favour of maximising even their own good. Novice Buddhists are taught an impersonal ethics that is based on the metaphysical belief that there is no underlying and enduring self (Goodman 2009; Siderits 2007). According to several Buddhist schools, personal identity is a sort of illusion: it is literally untrue that you or I persist over even a day. This belief may motivate a more impersonal, and thus impartial, approach to ethics (cf. Parfit 1984/87). This is hardly the most extravagant metaphysical or moral belief that is held by a considerable number of people. After all, if atheists are right, the vast majority of humanity is deluded about a central aspect of their lives.[12]

People have been made to believe not only counterintuitive metaphysical claims, but enact moral practices that are apparently self-denying or even morally depraved: the long list of atrocities includes the obligation to perform human sacrifices, the legitimacy of killing innocent Jews, the moral legitimacy of mass killings for ideological or utopian reasons, suicide attacks, witch hunting and the Inquisition. In addition, it is at best unclear whether learning these beliefs and the practices based on them required significant amounts of effort, other than perhaps starting at a very early age.

It may be argued that moral transformations – such as one involving a reduction of generational partiality – do not involve merely the transmission of beliefs, but also the cultivation of certain intellectual capacities. In his recent work (2025), Mulgan discusses at length, in the context of a thought experiment in which humanity would be facing extinction in two hundred years, the feasibility of multigenerational projects involving moral transformation. More specifically, Mulgan wants to defend the claim that certain multigenerational projects aimed at reorienting certain moral traditions and practices are feasible – even in the context of his thought experiment. Now, while conventional moral progress involves expanding existing moral frameworks or improving compliance, more radical moral transformations require also fundamentally reimagining moral concepts. In describing the intellectual skills required for moral transformation, Mulgan introduces two distinctions, the first between 'imaginativeness' (i.e., the ability to recognize or invent new possibilities) and 'imagination' (i.e., the ability to visualise or entertain objects or states of affairs) (Mulgan 2025, p. 204), and the second

[12]The conclusion that at least a great part of humanity is deluded may follow even if atheists are wrong and one (exclusivist) religion is right: if only one religion can be right about existential claims pertaining to deities, then other religions' existential claims are false. All but believers in the true religion would hold crucial but mistaken beliefs about the existence of divine entities that, in most cases, influence their moral choices as well.

https://doi.org/10.1017/S0953820825000111 Published online by Cambridge University Press

between the morally imaginative (i.e., someone who invents new ethical possibilities) and the morally creative person (i.e., someone who puts new ethical possibilities into practice) (Mulgan 2025, p. 205). Mulgan also argues that moral imaginativeness and creativity – which are central to the kind of moral transformation required in certain multigenerational projects, including those reducing generational partiality – are inherently collective (Mulgan 2025, pp. 206–207). In fact, the application of these skills requires an inherited foundation in existing moral concepts developed by others, interpersonal justification and debate, social validation of innovations, and supportive institutional frameworks.

Assuming that the above account is largely correct, we may argue that moral transformations involving, for example, an increase in intergenerational impartiality would require considerable costs only in those contexts in which the prerequisite moral concepts or institutional framework are completely or significantly absent. If the costs of such changes are relevant in considering which code is optimific, then a code requiring such changes may be discouraged, and thus the resulting moral prescriptions of the code may not be too demanding after all. However, in those contexts in which the moral changes required to assume a more impartial ethics do not necessitate significant costs to be taught, then the relevant code that would turn out to be optimific may turn out to be the one prescribing, say, a higher level of impartiality. In general, it does not seem that, at least in some contexts, certain dispositions towards partiality would be extraordinarily difficult to replace. In these contexts, it therefore seems questionable that teaching a new generation to be more impartial would be particularly costly or difficult to reconcile with human nature. Different contexts, on the other hand, may require the introduction of certain educational practices aimed at improving skills such as moral imaginativeness. Such an introduction may turn out to be considerably costly. Whereas this may be true in some societies, it seems that most liberal societies do have the educational practices and structures required to foster certain skills conducive to more intergenerational impartiality. If this is correct, then after all IMOC or even ICRC may still prescribe moral codes that may seem demanding to common sense (now) – at least in those contexts in which their teaching would not be comparatively too costly to implement.

To summarise the main gist of this second line of objection: it is at best unclear whether it would be that costly to have the next generation accept or learn a more impartial ethics; with some exceptions, vast numbers of human agents seem receptive to many metaphysical and moral beliefs that are seemingly counter-intuitive and even self-denying. If it turns out not to be significantly costly to accept or learn an impartial and thus (apparently) demanding code, then the optimific code ICRC or IMOC prescribes may turn out to be one that is demanding by the present standards of common-sense morality.[13]

In general, the previous objections both show that the limiting mechanisms may fail to ensure that the forms of RC under consideration will not prescribe codes that are too demanding – demanding according to some presumed standards of common-sense morality. However, there is another, more fundamental problem with the proposed mechanisms: neither ICRC nor IMOC provides sufficient reasons for believing that these mechanisms align with what can reasonably be considered the proper limits of our impartiality and of our obligations to maximise the good. For instance, we are not given sufficient or at least explicit reasons to believe that our 'natural inclinations' against

[13]This point is not meant to deny or downplay other challenges posed, for instance, by the psychological difficulties related to a demanding moral code – see McElwee (2023).

https://doi.org/10.1017/S0953820825000111 Published online by Cambridge University Press

accepting a code, our resistance to learning an outlook, or the costs of implementing a successful moral education are the right sort of mechanisms or filters to help us judge whether prescriptions, obligations and so on are aligned with *proper* limitations on our impartiality, or on the maximisation of the good. In other words, we lack a good explanation of why reluctance to accept or learn a moral code would align with what we should properly regard as too demanding.

One approach to provide an explanation is to consider how limits to impartiality and maximising in ICRC and IMOC might be explained by the mechanisms that are supposed to align with them. For example, consider again IMOC. Suppose that the resulting IMO would include a certain limit to impartiality – that it is permissible, when the resulting loss of overall well-being is minimal, to dedicate part of one's spare time to oneself rather than, say, helping one's local community. Suppose that the agent's loss of well-being would not be significant either, so their sacrifice would not be of a kind that, if protracted and universalised among the population, would significantly decrease overall well-being. Now, suppose too that IMOC implied that it is permissible just to count the blades of grass in my lawn for 15 minutes rather than helping the local community. Can the mechanism adopted by IMOC be used to explain why this is the case? Applying it would give us the result that it is permissible, in this case, to count the blades of grass because it would be too costly, overall, to teach the contrary. However, the explanation seems neither convincing nor one that tracks the relevant moral facts – why would the fact that it is hard to learn to abstain from that action be what explains its permissibility? That the basis of its limiting mechanism lacks relevant explanatory value does not imply that IMOC has no other resources for explaining why an obligation to maximise overall well-being in this case would be too demanding. It is striking, however, that the basis of the main mechanism devised to weed out what is too demanding does not always seem to track the right or any explanatory facts. So, it is at least mysterious why we should trust these mechanisms to produce the intended results. It may well be true that, due to the difficulty of teaching a moral outlook, the overall costs of teaching it may decrease the overall benefit of doing so. However, it is at best unclear why costs of that kind should be regarded as the right kind of costs to determine what is properly demanding. It may be true that a requirement for an ideal code to be ideal is that it must be one that can be accepted or taught. Still, such a requirement may not be the main factor determining what is too costly in a cost–benefit analysis. After all, difficulty in learning or accepting an outlook may be due to entirely contingent factors that have little or nothing to do with what ought to be permissible or required. It may be the case that some of these difficulties do have explanatory value in determining what is permissible, but not all difficulties seem equally relevant.

3.2 Demandingness and human nature

In the previous section, I argued that a fundamental problem for both ICRC and IMOC is that how difficult it is to accept or learn a code is not always a relevant factor in a proper cost–benefit analysis comparing different codes. In this section, I outline one way of supplementing ICRC and IMOC that may partly ameliorate this problem.

A plausible starting point is the following question: what would need to be added to ICRC and IMOC to ensure that acceptability or teachability were (always) explanatorily relevant? The reply I sketch here develops what Hooker seems to have in mind when he deploys the expression 'natural dispositions'. In particular, I understand it to refer, at least implicitly, to the notion of human nature, in the sense that the relevant natural

https://doi.org/10.1017/S0953820825000111 Published online by Cambridge University Press

dispositions are those dispositions which would be included in a plausible account of our nature. More specifically, I will articulate the idea that acceptability and learnability are (always) relevant factors in a cost–benefit analysis of the kind at issue because, and insofar as, the difficulties in accepting or learning a code would be contrary to human nature, including our psychological capacities. The explanatory burden would thus shift to those features of our nature which make accepting or learning difficult, rather than acceptability and learnability in themselves. Acceptability or learnability would still be theoretically relevant factors because they would be tangible expressions of certain limits imposed by our nature.

One initial problem with this suggestion is that the notion of human nature is notoriously controversial; it is risky to attempt to clarify one central aspect of RC by deploying a notion that may turn out to be on an even less secure theoretical footing. Still, in recent years, plausible attempts to clarify the notion of human nature have been abundant. Among them, Edouard Machery's nomological account (Machery 2008, 2017, 2018) provides a promising expansion of what seem to be the 'natural dispositions' to which Hooker alludes.[14] Machery's account is called 'nomological' because it encompasses law-like (nomological) generalisations about humans – what humans tend to possess or exhibit because of their evolutionary history. This account regards as relevant those regularities in human traits that are grounded in our evolutionary past and are therefore relatively stable and predictable, similar to how scientific laws describe regular patterns in nature. In fact, on Machery's view, 'human nature' stands for the set of traits that 'human beings tend to possess as a result of the evolution of their species' (Machery 2008, p. 323). This definition combines two key elements: the universality proposal (traits must be typical of human beings) and the evolution proposal (traits must have evolved). These traits, in turn, are not exclusively instantiated by human beings – that is, for a property to be part of human nature, it need not be one that distinguishes humans from non-humans. In addition, the traits making up our nature have only to be typical; none of them needs to be possessed by all humans (Machery 2018, p. 20). Bimodal traits (e.g. different male and female mating psychologies) would not be included, as they are not widely shared by humans in general (Machery 2008, p. 324), but the ability to speak would be included, as would fear reactions to unexpected noise, biparental investment in children and being bipedal, even though not all of these traits are exclusively human and even if not all human beings have each one (Machery 2008, p. 323). Arguably, among these traits we can also include an instinct for self-preservation and some tendency towards partiality (Bloom 2013), and other psychological limits to our capacity to concentrate, patience, perseverance and so on (Chappell 2019; p. 255, McElwee 2023, pp. 924–925). According to Tomasello (2019), another typical trait of human beings is our capacity for shared intentionality – the ability to create shared goals, attention, knowledge, and attitudes with others. This capacity emerges in two key stages during development: joint intentionality (at around nine months), which enables infants to engage in joint attention, cooperative communication, and collaboration with individual partners; and collective intentionality (at around three years), which enables children to identify with and conform to cultural group norms and conventions. The properties and limits that belong to human nature

[14]There are other accounts that, properly adjusted, could serve a similar purpose. For instance, Martha Nussbaum has made a protracted attempt to ground her 'capabilities approach' to well-being and flourishing on our common human nature (see Nussbaum 1992). See also Antony (2000) for a sophisticated investigation of the idea.

https://doi.org/10.1017/S0953820825000111 Published online by Cambridge University Press

are those most humans possess and have because of a causal process: that of evolution.[15] Human nature is not itself a cause, but it is the result of a causal evolutionary process.

Now, this account can be combined with IMOC or ICRC by claiming that any prescriptions, obligations and so on that clash with traits that belong to human nature are at least always relevant when the comparative costs of a code including them are calculated. In fact, we may plausibly claim that a moral theory should not require actions that the relevant agents to which it applies *cannot* perform.[16] So, the costs of people's accepting or learning to behave in ways that would defy their typical nature – neglecting their children, complete sexual abstinence, immense psychological efforts, deprivation from participating in shared goals and activities, etc. – should always count when determining whether a code is too demanding. The appeal to human nature would be the theoretical tool against the claim that certain difficulties rather than others are arbitrary: there does not seem anything arbitrary in claiming that the relevant moral agents can hardly be expected to do what they cannot do given the kind of agents they are. This solution would suggest that difficulties in accepting or learning a code are always relevant in comparative cost–benefit analyses if they arise from violations of our human nature. Arguably, such difficulties also have explanatory value.

Unfortunately, this attempt to amend RC still faces certain limits. First, it may be questioned why one specific individual's actions should be regulated by what is typical for members of their species, rather than, say, their own particular nature.[17] In response to this objection, we may suggest that the relevant aspects of human nature are not all those that are typical of human beings, but those of the relevant individual agent(s). On this modified view, what is relevant for a cost–benefit analysis of a code is the particular nature (and limits) that each individual in the relevant group typically displays (and has) separately and in relation to each required part of the code. In fact, we may argue that, within certain limits, rules should be 'individualised' to make them responsive to individuals' natural capacities and limits. This possible solution has several drawbacks, not least a risk of engendering codes of implausible complexity. In fact, codes would have to contain rules addressed to specific classes of individuals. Still, even considering this drawback, it seems to me that having a more complex code that responds to each person's natural capacities and limits is the better option – at least insofar as one of the purposes of RC is that of maximising overall well-being. Another possible criticism of this solution is that some aspects of our nature (e.g., some violent dispositions of some individuals) should not have to be counted as relevant in assessing whether a theory is demanding in case the theory requires us to curb them. For example, a code that would be optimific other than for the fact that it requires difficult processes for learning how to reduce some individuals' natural aggressive behaviour should not be preferred to a code the only comparative advantage of which is making more lenient requests to particularly violent individuals. So, although it is plausible that some moral rules may be too demanding because they violate certain aspects of our human nature – and that difficulty in learning or accepting them is a manifestation of this – not all of our typical traits seem to be relevant to, or deserve to count in, a comparative cost–benefit analysis determining which code is optimific.[18] If this is correct, then a version of ICRC and IMOC which is based on a nomological account of human nature would at most be

[15]See also Griffiths (1999) and Tomasello (2019).

[16]The kind of modality implied in this use of 'cannot' is nomological or biological.

[17]This point is persuasively discussed, in a different context, by Norcross (2004, p. 240).

[18]See Antony's (2000) discussion of Nussbaum's proposal for even more sceptical considerations.

https://doi.org/10.1017/S0953820825000111 Published online by Cambridge University Press

in need of further supplementation – in particular, a method for determining which natural dispositions or traits are relevant to the comparative cost–benefit analysis of codes or outlooks. Still, an appeal to human nature would at least make less arbitrary some claims regarding the demandingness of a moral code.

4. Conclusion

Although the last section concluded with a series of doubts about my attempt to solve the third objection by including a specification of human nature in RC, I do not think that ICRC or IMOC is significantly less plausible because their limiting mechanisms are not sufficiently strong – which would result in the theories being relatively demanding, at least for the present generation. Indeed, it can be argued that it is not substantially damaging to a moral theory's plausibility that the theory may prove more demanding towards some generations than some in those same generations might pre-theoretically expect, especially considering what Mulgan calls 'broken futures'. For example, suppose our considerations of what is demanding include the costs future generations will incur if we do not decrease our partiality for our present generation enough to reduce certain catastrophic risks, such as those related to climate change. These future costs may be in the currency of suffering, loss of opportunity, resources to be deployed to curb the ill effects of climate change, and so on. If we include such costs in a consideration of what is demanding, we may consider theories that permit us to expose future generations to risks generating such costs to be even more demanding (at least for future generations) than those theories requiring the present generation to make sacrifices which would reduce such risks. A moral theory sufficiently ambitious to address the possibility of broken futures may end up having to violate certain intuitions of common-sense morality, such as those supporting a substantial partiality towards present generations over future ones.

Financial support. This research was supported by the grant FRG#101919, Faculty of Arts, Lingnan University.

References

Antony Louise, 'Nature and Norms', *Ethics*, **111** (2000), 8–36.
Ashford Elizabeth, 'The Demandingness of Scanlon's Contractualism', *Ethics*, **113** (2003), 273–302.
Bloom Paul, *Just Babies* (New York: Random House, 2013).
Brandt Richard, *A Theory of the Good and the Right* (Oxford: Oxford University Press, 1979).
Brandt Richard, *Morality, Utilitarianism, and Rights* (Cambridge: Cambridge University Press, 1992).
Chappell Richard, 'Willpower Satisfacing', *Noûs*, **53** (2019), 251–265.
Cullity Gareth, 'Moral Character and the Iteration Problem', *Utilitas*, **7** (1995), 289–299.
Dougherty Tom, 'Altruism and Ambition in the Dynamic Moral Life', *Australasian Journal of Philosophy*, **95** (2017), 716–729.
Forcehimes Andrew T. and Luke Semrau, 'Beneficence: Does Agglomeration Matter?', *Journal of Applied Philosophy*, **36** (2019), 17–33.
Formosa Paul and Martin Sticker, 'Kant and the Demandingness of the Virtue of Beneficence', *European Journal of Philosophy*, **27** (2019), 625–642.
Gardiner Stephen, *A Perfect Moral Storm* (Oxford: Oxford University Press, 2011).
Goodman Charles, *Consequences of Compassion* (Oxford: Oxford University Press, 2009).
Griffiths Paul E., 'Squaring the Circle: Natural Kinds with Historical Essences', in *Species: New Interdisciplinary Essays*, ed. by R. A. Wilson (Cambridge, MA: MIT Press, 1999), pp. 209–228.

https://doi.org/10.1017/S0953820825000111 Published online by Cambridge University Press

Hills Alison, 'Utilitarianism, Contractualism, and Demandingness', *Philosophical Quarterly*, **60** (2010), 225–242.
Hooker Brad, 'Rule-Consequentialism and Obligations Toward the Needy', *Pacific Philosophical Quarterly*, **79** (1998), 19–33.
Hooker Brad, *Ideal Code, Real World* (Oxford: Clarendon Press, 2000).
Hooker Brad, 'The Demandingness Objection', in *The Problem of Moral Demandingness*, ed. by T. Chappell (London: Palgrave, 2009), pp. 148–162.
Hooker Brad, 'Egoism, Partiality, Impartiality', in *Oxford Handbook on the History of Ethics*, ed. by Roger Crisp (Oxford: Oxford University Press, 2013), pp. 710–728.
Hooker Brad, 'The Role(s) of Rules in Rule Consequentialism', in *The Oxford Handbook of Consequentialism*, ed. by D. Portmore (Oxford: Oxford University Press, 2020), pp. 441–462.
Hooker Brad, 'Rule Consequentialism', in *Stanford Encyclopedia of Philosophy*, Spring 2023 Edition, ed. by Edward N. Zalta and Uri Nodelman (2023). <https://plato.stanford.edu/archives/spr2023/entries/consequentialism-rule/>
Hooker Brad and Guy Fletcher, 'Variable versus Fixed-Rate Rule-Utilitarianism', *Philosophical Quarterly*, **58** (2008), 344–352.
Horta Oscar, Gary O'Brien and Dayron Teran, 'The Definition of Consequentialism', *Utilitas*, **34** (2022), 368–385.
Jeske Diana, 'Consequentialism and Partiality', in *The Oxford Handbook of Consequentialism*, ed. by D. Portmore (Oxford: Oxford University Press, 2020), pp. 238–252.
Kronfeldner Maria, Neil Roughley and Georg Toepfer, 'Recent Work on Human Nature', *Philosophy Compass*, **9** (2014), 642–652.
Machery Eduard, 'A Plea for Human Nature', *Philosophical Psychology*, **21** (2008), 321–330.
Machery Eduard, 'Human Nature', in *How Biology Shapes Philosophy*, ed. by D. Livingstone Smith (Cambridge: Cambridge University Press, 2017), pp. 204–226.
Machery Eduard, 'Doubling Down on the Nomological Notion of Human Nature', in *Why We Disagree About Human Nature*, ed. by E. Hannon and T. Lewens (Oxford: Oxford University Press, 2018), pp. 19–40.
McElwee Brian, 'Cost and Psychological Difficulty', *Australasian Journal of Philosophy*, **101** (2023), 920–935.
Miller Dale E., 'Rule Utilitarianism', in *Cambridge Companion to Utilitarianism*, ed. by B. Eggleston and D. E. Miller (Cambridge: Cambridge University Press, 2014), pp. 146–165.
Miller Dale E., 'Moral Education and Rule Consequentialism', *Philosophical Quarterly*, **71** (2021), 120–140.
Miller Timothy, 'From Compliance, to Acceptance, to Teaching: On Relocating Rule Consequentialism's Stipulations', *Utilitas*, **33** (2021), 204–220.
Mogensen Andreas, 'Moral Demands and the Far Future', *Philosophy and Phenomenological Research*, **103** (2021), 567–585.
Mosquera Julia, 'Rule-Consequentialism, Procreative Freedom, and Future Generations', *Ratio*, **35** (2022), 333–343.
Mulgan Tim, *Future People* (Oxford: Oxford University Press, 2006).
Mulgan Tim, 'Utilitarianism for a Broken World', *Utilitas*, **27** (2015), 92–114.
Mulgan Tim, 'How Should Utilitarians Think About the Future?', *Canadian Journal of Philosophy*, **47** (2017), 290–312.
Mulgan Tim, 'Answering to Future People', *Journal of Applied Philosophy*, **35** (2018), 532–548.
Mulgan Tim, *Utilitarianism* (Cambridge: Cambridge University Press, 2020).
Mulgan Tim, *Philosophy for an Ending World* (Oxford: Oxford University Press, 2025).
Murphy Liam, 'The Demands of Beneficence', *Philosophy and Public Affairs*, **22** (1993), 267–292.
Norcross Alastair, 'Puppies, Pigs, and People', *Philosophical Perspectives*, **18** (2004), 229–245.
Nussbaum Martha, 'Human Functioning and Social Justice: In Defense of Aristotelian Essentialism', *Political Theory*, **20** (1992), 202–246.
Parfit Derek, *Reasons and Persons* (Oxford: Clarendon Press, 1984/87)).
Sauchelli Andrea, 'Acceptance-Based or Teaching-Based Rule Consequentialism?', *Ratio*, **38** (2025), 63–70.
Siderits Mark, *Buddhism as Philosophy* (Indianapolis, IN: Hackett, 2007).
Sobel David, 'The Importance of the Demandingness Objection', *Philosophers' Imprint*, **7** (2007), 1–17.

Sobel David, 'Understanding the Demandingness Objection', in *The Oxford Handbook of Consequentialism*, ed. by D. Portmore (Oxford: Oxford University Press, 2020), pp. 220–237.
Thomson Judith J., 'A Defence of Abortion', *Philosophy and Public Affairs*, **1** (1971), 47–66.
Tomasello Michael, *Becoming Human* (Cambridge, MA: Harvard University Press, 2019).
Wolf Susan, 'Moral Saints', *Journal of Philosophy*, **79** (1982), 419–439.
Woodard Christopher, *Taking Utilitarianism Seriously* (Oxford: Oxford University Press, 2019).
Woollard Fiona, 'Dimensions of Demandingness', *Proceedings of the Aristotelian Society*, **116** (2016), 89–106.
Woollard Fiona, 'Hooker's Rule-Consequentialism, Disasters, Demandingness, and Arbitrary Distinctions', *Ratio*, **35** (2022), 289–300.

Cite this article: Sauchelli, Andrea. 2025. Rule Consequentialism and Demandingness: The Wrong Solution(s)?. *Utilitas* 1–14. https://doi.org/10.1017/S0953820825000111