Consequentialism and Collective Action

Brian Hedden

Abstract

Many consequentialists argue that you ought to do your part in collective action problems like climate change mitigation and ending factory farming because (i) all such problems are triggering cases, in which there is a threshold number of people such that the outcome will be worse if at least that many people act in a given way than if fewer do, and (ii) doing your part in a triggering case maximizes expected value. I show that both (i) and (ii) are false: Some triggering cases cannot be solved by appeal to expected value, since they involve infinities, and some collective action problems are not triggering cases, since they involve parity. However, I argue that consequentialism can still generally prohibit failure to do your part in those collective action problems where we believe that so acting would be impermissible.

Keywords: Consequentialism, collective action, parity, incommensurability, climate change

1 Collective Action Problems

Collective action problems take many forms. I will focus on a kind of collective action problem in which a much better outcome would result if all or most people performed a given action than if few or none did so, and yet it is tempting to think that no particular individual could make any difference by acting one way or the other.

Examples: A much better outcome would result if everyone reduced her carbon footprint than if no one did. But it is tempting to think that no single individual could make a

difference by reducing her carbon footprint. Average global temperatures would be the same regardless of whether one drove a hybrid or a gas-guzzler.

If everyone were to refrain from buying factory farmed meat, things would be much better than if no one did so. But it is tempting to think that no single purchase would affect how many animals are raised and slaughtered in factory farms. The supply chain just isn't that sensitive to individual purchases.

If everyone were to vote for the better candidate, then that candidate would win. If no one were to do so, then either the worse candidate would win, or else democracy might collapse (if no one voted at all). But it is tempting to think that no single individual would make a difference to the outcome. Only tiny races for dog-catcher are ever decided by a single vote.

To fix terminology, say that an *exceptional* act is one such that things would be better if few or no people did it than if all or most did, but where arguably no single such act would make things worse. The term is supposed to evoke the idea that in performing such an act, one seems to be making an exception for oneself. Not voting, buying factory-farmed meat, and not reducing one's carbon footprint are examples of exceptional acts.

My focus in this paper is whether and to what extent consequentialism permits performing exceptional acts. *Prima facie*, it might seem that consequentialism will often permit such exceptional acts. For consequentialism is concerned with bringing about the best outcome possible, and if any particular exceptional act would make no difference to how good the outcome is, it seems that that act would be permitted by consequentialism. This implication might constitute a serious objection to consequentialism. But my concern is not with whether consequentialism's treatment of collective action problems should make us more, or less, confident in the truth of consequentialism. Rather, I am concerned simply with whether and to what extent consequentialism permits exceptional acts. I leave it to others to decide whether consequentialism's verdicts, and reasons for those verdicts, are adequate.

Many consequentialists have argued that, in fact, consequentialism does *not* permit ex-

ceptional acts (Kagan 2011; see also Singer 1980; Parfit 1984; Norcross 1997, 2004). Following Kagan, say that in a *triggering case* there is some threshold number of people k such that the outcome would be worse if at least k people were to perform a given exceptional act than if fewer were to (there may be several such thresholds). These consequentialists then make two claims: First, in triggering cases, the exceptional act is impermissible because it has sub-maximal *expected* value.[1] It has a high probability of making things slightly worse, or a low probability of making things *much* worse, or something in between, but in any case its expected value will be lower than that of not performing the exceptional act. Second, all collective action problems are triggering cases.

This standard treatment is attractive and elegant. But it has also been met with skepticism (Nefsky 2011; Budolsfon 2018). I argue that both of its component claims are in fact false, but not for the reasons emphasized by critics, and, more importantly, not for reasons that significantly threaten the overall verdict. In each case, a slight tweak saves the verdict that consequentialism generally prohibits exceptional acts.

Here is the plan: In §2, I show that expected value theory provably cannot prohibit exceptional acts in some infinitary triggering cases but argue (*contra* Budolfson) that it is likely to do so in more realistic cases. In §3, I turn to the second claim. While conceding the flaws in previous attempts to show that all collective action problems are triggering cases, I argue that a better strategy is available, based on the claim that all relations of the form *is exactly as F as* are equivalence relations. Hence there cannot be a sequence of outcomes (ordered by the number of people who perform the exceptional act) with the first better than the last but each exactly as good as its predecessor. But as I discuss in §4, this is not sufficient to show that all collective action problems are triggering cases. Parity lurks. While there cannot be a sequence of outcomes with the first better than the last and each

[1]The expected value of an act is the result of adding up, for each possible outcome, the product of the value of that outcome and the probability that the act will result in that outcome. In symbols: EV(A) $= \sum_i P(O_i \mid A)V(O_i)$. The consequentialist theory then says that an act is permissible just in case no alternative act has higher expected value.

outcome *exactly as good as* its predecessor, the possibility of parity means that there can be a sequence of outcomes with the first better than the last and each outcome *not worse than* its predecessor. But I show that one attractive theory of decision-making under parity allows consequentialism to prohibit exceptional acts even in parity-laden, non-triggering collective action problems. If this decision theory is correct, then even though not all collective action problems are triggering cases (because of parity) and expected value theory can't prohibit exceptional acts in all triggering cases (because of infinities), consequentialism still prohibits exceptional acts in most of the kinds of collective action problems with which we are most concerned, and where we tend to judge that exceptional acts would be impermissible.

2 The Appeal to Expected Value

Let's start with the bad news. The standard consequentialist treatment of collective action runs aground in infinitary cases. Suppose there are (countably) infinitely many people, each facing a switch. If only finitely many people flip their switches, then everyone spends eternity in heaven. But if infinitely many people do so, then everyone spends eternity in hell.

This is a triggering case, where the threshold number is $\aleph_0$, the cardinality of any countably infinite set. The outcome would be much worse if at least $\aleph_0$ people flipped their switches than if fewer did so. But here the appeal to expected value is impotent. For no single switch can mean the difference between infinitely many switches being flipped and only finitely many being flipped. Thus each person has probability 0 of making the outcome worse by flipping her switch, and flipping the switch does not have lower expected value than not flipping it.[2]

[2]See also Arntzenius, Elga, and Hawthorne (2004) for a related but importantly different case. In a group version of their Satan's Apple, there are infinitely many people, each with a slice of apple. If infinitely many eat their slices, everyone goes to hell, while if finitely many do so, everyone goes to heaven. It is good for each person to eat her slice, but the goodness of the gustatory pleasure is far outweighed by the badness of hell. In this case, for each person, the outcome would be better if she were to eat her slice than if she were not to, holding fixed what others do. Hence she would be required to eat on consequentialist grounds. This yields the puzzling result that every combination of acts is worse than some alternative combination. In my case, for each person, the outcome would be exactly as good whether or not she were to flip her switch.

Infinities are weird, and in this case I think we have no considered judgment that the exceptional act is indeed impermissible. Consequentialists may therefore be willing to concede that they cannot prohibit exceptional acts in *all* triggering cases, hoping instead to do so in those finite cases where we tend to judge that the exceptional act is impermissible.[3]

Budolfson (2018) argues that even this more modest goal is unattainable. Focusing on the case of factory farming, Kagan (2011) begins by conceding that it is unlikely that each chicken purchased causes one more chicken to be raised and slaughtered.[4] But there must still be some threshold, such that if that many chickens are purchased in a given period, approximately the same number of additional chickens will be raised and slaughtered.

> Thus we know that there is some triggering number T (more or less), such that every Tth purchase (more or less) triggers the order of another T chickens (more or less). I don't have any idea what that number is, but I do know that whatever it is, I have a 1 in T chance of triggering the suffering of another T chickens (more or less). And so in terms of chicken suffering, my act of purchasing a chicken still has an expected disutility equivalent to one chicken's suffering. And since, by hypothesis, this is greater than the pleasure I will get from eating the chicken, the net expected utility of my purchase remains negative. As I walk to the butcher counter, then, not only don't I know whether my act will have bad results, I don't even know what the *chances* are that my act is a triggering act. But I do know, for all that, that the net expected results of my act are bad. So I should not buy a chicken. (2011, 124).

As Budolfson explains, this reasoning relies on the assumption that the expected effects of the relevant act are approximately equal to the average effects of that sort of act. Given modest assumptions about the efficiency of the marketplace, each act of purchasing a chicken has, on average, the effect of one chicken being raised and slaughtered. And, Kagan suggests,

Hence consequentialist considerations seem to permit, but not require, each person to flip. And here, there are many optimal combinations of acts, namely all those where only finitely many people flip.

[3]This concession does mean denying that it is part of the nature of morality that it satisfies the 'Principle of Moral Harmony,' which states that 'when all the members of a social group do what they morally ought to do, the group as a whole does benefit more than it would have from the performance of any worse alternative set of actions' (Feldman 1980, 167). Feldman gives independent reasons to doubt this principle, however. But see Portmore (2018) for a defense of a modified principle of moral harmony.

[4]See Broome (2018) for discussion of thresholds in the case of climate change.

if you are ignorant about the location of the thresholds and about how others will act, the expected number of extra dead chickens resulting from an individual chicken purchase will be approximately equal to that average effect. (As a side note, it is worth mentioning that consequentialism will still prohibit buying a chicken even if that act has an expected increase in the number of chickens produced of less than one, provided that the suffering of each chicken far outweighs the difference in the pleasure one gets from eating a chicken and the pleasure one would get from alternative vegetarian meals. If the suffering of a chicken in a factory farm still outweighs, say, the aggregate net pleasure of eating 100 chickens, as it plausibly does, then buying a chicken would be prohibited even if its expected increase in the number of chickens produced were only around 0.01.)

Budolfson (2018) argues that Kagan is overly optimistic. This is because we are in a position to know about the presence of 'buffers' in the supply chain which reduce both the probability of an individual act making a difference as well as the size of the difference it will make, if it makes any difference at all (see also Nefsky 2011). A buffer is anything that makes production less sensitive to individual consumer acts. As one example, a chicken wholesaler might have the option of selling unsold meat at cost to a dog food manufacturer or rendering plant, with the result that a small reduction in demand from ordinary consumers will not yield a concomitant reduction in the number of chickens the wholesaler purchases from producers. Budolfson argues that these buffers make the expected effect of an individual act significantly lower than the average effect of acts of that type, with the result that the relevant exceptional act does not have sub-maximal expected value.

Does our knowledge of these buffers scuttle the appeal to expected value? I do not think so. It is important to emphasise that buffers in a supply chain do not *eliminate* the existence of threshold numbers. Instead they help determine what those threshold numbers are, in particular by placing the thresholds farther apart than they would otherwise have been. But Kagan has already conceded that the thresholds may be very few and far between, and that

the probability of an act making any difference at all may therefore be very low. So why should the existence of buffers pose a threat to the appeal to expected value? Expected value calculations take into account all probabilities, no matter how small, meaning that there is no positive number such that the probability of an act making a difference must be above that number in order for it to be prohibited on expected value grounds. What matters is the relation between the probability of making a given difference and the size of that potential difference, where an increase in one can compensate for a decrease in the other. And Kagan makes the plausible assumption that, in general, buffers which reduce the probability of an individual act making a difference will yield a compensating increase in the size of the effect that an individual act will have, in the unlikely event that it does make a difference. If this is right, then chicken-purchasing will be prohibited on expected value grounds, regardless of the size and effectiveness of the buffers in question.

Budolfson argues that this assumption is mistaken and that buffers can reduce the probability of an act making a difference without yielding a compensating increase in the size of the effect that the act will have, if it does happen to make a difference. He writes (p. 8):

> even in the very unlikely event that, say, an individual purchase of meat really did succeed in making the price of animals at one point at a production end of the supply chain $0.01 higher than it otherwise would have been, that would not make the dramatic difference to the number of animals that are brought into existence that it would have to make in order for the possibility of such a threshold effect to drive the expected effect toward the average effect, in part because the number of animals that are brought into existence is suprisingly insensitive to very small changes in price at that location for a variety of reasons.

In a footnote, Budfolson supports this contention by observing, for the case of cattle raising, that 'insofar as ranchers judge that capital should be invested in raising cattle rather than other investments, they will tend to raise as many cattle as they can afford to breed and feed within that budget, letting the ultimate extent of their profits fall where it may at the feedlot' and that many ranchers 'use the nutritional well-being of their herd as a buffer

to absorb changes in market conditions, feeding their cattle less and less to whatever point maximizes the new expectation of profits as adverse conditions develop' (ibid).

However, this observation at most shows that individual acts are unlikely to have any large effects *in the short term*. But consequentialists care about an act's *long-term* effects.[5] Perhaps, over the course of the next year, ranchers will raise and slaughter the same number of cattle regardless of any (relatively small) price changes. But profits one year will affect how things go the year after, and the year after that. At the margins, lower profits discourage new entries into the industry and may lead some existing ranchers to abandon cattle production altogether or to diversify their investments, for instance by shifting toward raising sheep for wool. This makes evident that quite a bit of work is being done by Budolfson's caveat about 'insofar as ranchers judge that capital should be invested in raising cattle.' We cannot hold fixed people's judgments about where to invest capital, since these may themselves be affected by consumer acts. Indeed, it is by influencing investment decisions that individual acts may be most likely to have large effects. Lower profits also mean that ranchers who persist in cattle raising will have less capital the next year. They might still 'raise as many cattle as they can afford to breed and feed within that budget,' but the budget will be lower, possibly resulting in fewer cattle bred and fed. Of course, individual acts are unlikely to have such dramatic effects, but that has already been conceded by consequentialists like Kagan. The point is that there is still a small probability of their having such large long-term effects. I am therefore unconvinced by Budolfson's contention that the effects of individual acts will be either null or too small to have any hope of being prohibited on expected value grounds.

Let me now turn to an example Budolfson gives to illustrate the importance of buffers:

> Richard makes paper T-shirts in his basement that say 'HOORAY FOR CONSEQUENTIALISM!', which he then sells online. The T-shirts are incredibly cheap to produce and very profitable to sell and Richard doesn't care about waste per se, and so he produces far more T-shirts than he is likely to need each month, and

[5] This focus on the long term raises Lenman's (2000) famous 'cluelessness' problem, however.

> then sells the excess at a nearly break-even amount at the end of each month to his hippie neighbor, who burns them in his wood-burning stove. For many years Richard has always sold between 14,000 and 16,000 T-shirts each month, and he's always printed 20,000 T-shirts at the beginning of each month. Nonetheless, there is a conceivable increase in sales that would cause him to produce more T-shirts—in particular, if he sells over 18,000 this month, he'll produce 25,000 T-shirts at the beginning of next month; otherwise he'll produce 20,000 like he always does. So, the system is genuinely sensitive to a precise tipping point—in particular, the difference between 18,000 purchases and the 'magic number' of 18,001. (2018, 6)

Budolfson argues that, given the facts about buffers in the T-shirt supply chain (the option of selling excess merchandise at cost) and about the historical trends in consumer purchasing decisions, the expected effect on T-shirt production of a single act of purchasing a T-shirt is 'essentially zero' because 'there is virtually no chance that exactly 18,001 people are going to buy Richard's T-shirts this month and trigger a dramatic threshold effect' (ibid, 6). Thus the expected effect of buying a T-shirt is much lower than the average effect of consumers' acts of buying T-shirts. He concludes that the problem with Kagan-style reasoning is 'that it overlooks the fact that we can know enough about the supply chains...to know that threshold effects are not sufficiently likely and are not of sufficient magnitude to drive the expected effect of consumption anywhere close to the average effect' (ibid, 7).

Now, we must concede that there can be no decisive, *a priori* argument that the expected value of purchasing a chicken (or other exceptional acts) will be sub-maximal, because as Budolfson rightly notes, 'the knowledge available about the mechanisms at play in such situations matters greatly' (ibid, 11). After all, it is rational subjective probabilities that matter in calculating expected values, and rational subjective probabilities depend on the agent's evidence. We can even imagine an evil demon planting misleading evidence to suggest to each person that they are nowhere near any thresholds, in which case expected value theory will not prohibit the exceptional act.[6]

[6]Note also that in cases where we know we are nowhere near any thresholds, we often do not judge that the exceptional act would be impermissible. For instance, deciding not to engage in any food production is

But it is important to be clear about what is going on in Budolfson's example. As noted above, buffers in the supply chain do not eliminate the existence of thresholds, but instead help determine where they are. This means that, when you possess detailed information about the exact workings of buffers, this could in principle provide evidence about what the threshold numbers are. And when you also possess information about historical trends in consumer purchasing decisions, this provides evidence about how many others will perform the relevant act. Now, it is not surprising that if you have evidence about what the threshold numbers are and about how many others will perform the relevant act, the expected effects of your act will probably be lower than if you didn't possess all that evidence. Think of it this way: In cases where threshold numbers are few and far between, we already know that it is very likely that your act will make no difference. Hence it is very likely that, in the limiting case where you are fully informed about all aspects of the situation, and, in particular, about the exact (post-buffer) threshold numbers and about how everyone else will act, the expected effects of your act will be null. More generally, as you gain more and more evidence about what the threshold numbers are (e.g., by learning more and more about the buffers) and how others will act (e.g., by learning more and more about historical demand), the expected effects of your act will *probably* become lower and lower, the exception being the rare case in which you are in fact right at the threshold number, in which case the expected effects of your act will actually increase as you become more informed.

But in order for the expected effect of buying a chicken on the number of chickens raised and slaughtered to be 'essentially zero,' it is not enough to know *that* there are certain buffers in the supply chain and *that* there are historical trends in consumer decisions. This is because mere knowledge *that* there are certain buffers provides little or no evidence about where the new threshold numbers are (or about the possible magnitudes of an individual act's effects, assuming the previous argument is correct), and because mere knowledge *that*

an exceptional act, as things would be worse if everyone did this than if no one did. But it is permissible for me not to produce food, since I know I am nowhere near a threshold where doing so will make things worse.

there are historical trends in consumer decisions provides little or no evidence about how many others will perform the relevant act.

In order for the the appeal to expected value to fail, you would also need detailed evidence about what the trends in consumer decisions in fact are, as well as detailed evidence about the exact workings of these buffers, so as to be able to better locate where you sit in relation to any thresholds. These facts were simply given to us in Budolfson's T-shirt case. But in real-life we have no such knowledge. I have no idea even approximately how many chickens are consumed worldwide each year. And while I believe that consumption is increasing, with the effects of a growing population and people rising out of poverty outweighing increased vegetarianism, I am ignorant about its rate of increase. Now, some evidence about consumer trends is available online. But more to the point, while I have some idea about the nature of buffers in the global chicken supply chain (e.g., that some excess is sold to rendering plants), I have no idea even approximately what the new threshold numbers are that result from the operation of these buffers. And given the complexity of global economic forces, not even industry experts could determine even roughly what the threshold numbers are that result from these buffers, especially given that those thresholds concern the number of chickens raised and slaughtered over the long run.

In real-life cases, then, we are in roughly the situation that Kagan and others suppose. Namely, we are very ignorant about where the thresholds are and about how many others will perform the relevant exceptional act, and so we are also very ignorant about how close or far we may be from hitting the threshold.[7] Along with my previous argument that buffers do

[7]Budolfson might concede that you should have a roughly uniform probability distribution over hypotheses about exactly how many others will buy chickens, as well as a roughly uniform probability distribution over hypotheses abvout what all the threshold numbers are. However, he could rightly point out that this is not enough to vindicate the appeal to expected value. For you might have a non-uniform probability distribution about what the threshold numbers are, conditional on any given hypothesis about how many chickens will be purchased. That is, even if you have no idea what demand will be or where the new, post-buffer thresholds are, you might nonetheless think that the two are correlated, such that the thresholds and anticipated demand, whatever they are, are likely to be far apart. In the T-shirt case, this would be the case if you thought that Richard sets up his supply chain with the aim of ensuring that demand will not approach the new thresholds. But I see little reason to think that demand and the new, post-buffer thresholds will be correlated in this

not prevent individual acts from having large long-term effects, this ignorance about where we sit in relation to any thresholds largely vindicates the appeal to expected value.[8]

In conclusion, I think Budolfson is greatly overstating things when he writes that 'in the real world we generally have access to additional evidence that makes it empirically indefensible to equate the expected marginal effect and average effect in such a way, and that makes it similarly indefensible to assign a probability to making a difference that would be sufficiently high to vindicate the conclusions of the [expected value response to triggering cases]' (ibid, 10-11). Budolfson is correct that the expected value of purchasing a chicken (or any other exceptional act) will not necessarily be sub-maximal regardless of what one's evidence might be, but wrong in thinking that the expected value will not be sub-maximal given our actual evidence.[9] I conclude that expected value considerations will still prohibit exceptional acts in most of the triggering cases with which we are most concerned.

way in real-world cases like factory farming, given that these cases involve a large and fluctuating number of producers, operating independently, none of whom has the power to unilaterally determine global production and none of whom is likely to care much about exact global demand.

[8]In this respect, the cases of factory farming and climate change are importantly different from that of voting, *contra* Budolfson (2018, 8). Polling data and knowledge of the voting rule allow citizens to locate approximately where they sit in relation to the relevant thresholds. In cases where the race isn't close, this will justify a tiny probability of one's vote making a difference. And this probability must be further discounted by one's confidence that one has accurately identified the candidate who is in fact better, as Lomasky and Brennan (2000) note. So consequentialism will not always require voting. But when the race is close, the stakes are huge, and one candidate is clearly better, as in the case of the last several US presidential elections, say, consequentialism may require voting despite the still tiny probability of making a difference. See also Barnett (ms) for an argument that, given two modest assumptions which are often met in real-life, voting will be required on consequentialist, difference-making grounds.

[9]As a reviewer noted, Budolfson might be interpreted not as making a claim about expected effects, given a typical consumer's actual evidence, but rather as providing us with additional evidence such that, in light of that evidence, we see that the expected effects of the exceptional act are too small for it to be prohibited on expected value grounds. But as noted, the complexity of economic forces means that even industry experts will be ignorant of where the thresholds are and how large the long-term effects of a given consumer act might be. So, even relative to such experts' evidence, the expected effects of purchasing a chicken will be approximately equal to one additional chicken produced. And it would not help to interpret Budolfson as making a claim about expected effects relative to the objective chance function. For the global economy is arguably not a physically chancy system, meaning that the objective chance of an individual act making a difference is either 0 or 1, depending on whether we are in fact right at a threshold, but that we don't know which it is. And if it is the latter, then the expected effect of purchasing a chicken, relative to the objective chance function, will be much greater than the average effect of chicken-purchasing acts.

3 Imperceptible Harms

Turn now to the second component of the standard consequentialist treatment of collective action cases. This is the claim that all collective action problems are triggering cases: there is always some threshold number of people k such that the outcome would be worse if at least k people performed a given exceptional act than if fewer did so. (As noted, there may be multiple thresholds, and indeed it could be that every additional exceptional act makes the outcome worse.) This claim is intuitively compelling in the cases of voting and factory farming. But it is less obvious in other cases, like that of climate change.

Let us consider a famous case which puts pressure on this claim, namely Parfit's (1984, 80) case of the harmless torturers. There is a patient hooked up to a torture machine. Other than the patient, there are n people, including you, each of whom has a switch in front of her. Flipping that switch will slightly increase the voltage going into the patient. If no one flips her switch, the patient will receive no voltage and experience no pain. If everyone flips her switch, the patient will receive a very high voltage and experience great pain. But for all j, the patient cannot tell the difference between the pain involved in the outcome O_j in which exactly j people flip their switches and the pain involved in the outcome O_{j+1} in which exactly $j+1$ people flip their switches. Hence it seems like each possible outcome O_{j+1} is just as painful—and therefore just as good—as its predecessor O_j. Thus, the harmless torturers case seems like a non-triggering collective action problem.

I begin by giving my own response before explaining how it avoids the problems facing previous consequentialist-friendly responses. I claim that for all F, the relation *is exactly as F as* is an equivalence relation.[10] (Indeed, I think this fact is a conceptual truth, though I

[10]This seems obvious to me, but oddly, it has been scarcely defended in the literature. Broome (2004, 151-2) does claim that *equally as good as* is transitive. This follows from his defintion, on which A is equally as good as B just in case (i) A is neither better nor worse than B, and (ii) for any C, C is better (worse) than A if and only if C is better (worse) than B (ibid, 20). He also notes that it would follow from analysis on A is equally as good as B just in case the degree of A's goodness is identical to the degree of B's goodness, given the transitivity of identity. Broome would presumably thinks that the same holds if we substitute any other predicate for 'good.' I do not, however, commit myself to either of Broome's proposed analyses.

do not need this stronger claim here.) Being an equivalence relation, it is transitive. Hence there cannot be a sequence of outcomes such that the last is F-er than the first and yet each outcome is exactly as F as its predecessor.

In the harmless torturers case, the relevant F is *painful*. From my general claim, it follows that the relation *is exactly as painful as* is an equivalence relation, and hence transitive. Thus, there cannot be a sequence of states such that the last is more painful (for the patient) than the first and yet each state is exactly as painful its predecessor. Given that O_n is indeed more painful than O_0, it follows that there must be at least one state that is not exactly as painful as its predecessor.[11]

This does not mean that the patient can tell the difference between any two adjacent states. The two states may be indiscriminable in the sense that the patient is not in a position to know whether they are exactly as painful as each other (Williamson 2013 (1990)). Indiscriminability is non-transitive. This should not be surprising. Given our limited powers of discrimination, it should not be assumed that one is always in a position to tell whether two states are *exactly*, as opposed to merely *almost* exactly, the same as each other, even with respect to some phenomenal property. This is especially true in cases like this one, where the states cannot be experienced simultaneously (meaning that the comparisons rely on memory), not to mention that extreme pain interferes with one's cognitive capacities.[12]

[11] See also Barnett (2018) for a different innovative and, in my view, compelling argument that cases like the harmless torturers must be triggering cases.

[12] Compare Graff Fara (2001) and Mills (2002), who defend the fallibilist claim that we are not always in a position to know whether two things look the same to us. Graff Fara (2001) rebuts an argument that limited powers of discrimination mean that looking the same, understood as sameness of visual phenomenology, is non-transitive. She considers two possible ways of cashing out the claim that our powers of discrimination are limited. First way: 'For some sufficiently slight amount of change (in colour, sound, position, etc.), when we perceive an object for the entirety of an interval during which it changes by less than that amount, we perceive it as not having changed at all during that interval' (917). But this claim is false, for it entails that we never misperceive an object as having changed in the relevant respect when it has in fact not changed at all. Second way: 'For some sufficiently slight amount of change (in colour, sound, position, etc.), we cannot perceive an object as having changed by less than that amount unless we perceive it as not having changed at all (as having changed by a zero amount)' (917). But this claim does not entail that phenomenal sameness is non-transitive. For ruling out the possibility of an interval during which the object appears to change by some amount below the threshold leaves open the possibility that the object will appear to change discretely at some point in an interval where it in fact changes continuously.

And *defining* what it is for one state to be exactly as painful as another in terms of the impossibility of discriminating between them in a pairwise comparison smacks of the crude operationalism that has long since fallen into disrepute in the philosophy of science; we would not, for instance, define what it is for one thing to be exactly as hot as another in terms of the impossibility of some thermometer's giving different readings for them.

(It may also be that for each state, it is indeterminate, and not merely unknowable, whether it is exactly as painful as its predecessor, even though it is determinately true that not every state is exactly as painful as its predecessor. Then, it would be determinately the case the harmless torturers case is a triggering case, but indeterminate where the thresholds are. This indeterminacy-laden case can be treated along consequentialist lines by appeal to a decision theory for indeterminacy which is analogous to the decision theory for parity that I explore in the next section. See also footnote 26.)

What about the claim that phenomenal properties are response-dependent in the sense that judging that the property applies in a given case makes it the case that it so applies? If the patient judges that each state is exactly as painful as its predecessor, might that make it the case that they are exactly as painful as each other? I do not need to deny that the monadic property of being painful is response-dependent (though I am skeptical of this claim). It may be that whenever a subject judges, of the state she is currently in, that it is painful, then it is painful. But such response-dependence is implausible for relations of comparative painfulness, unless further constraints are imposed. For we can imagine a subject who judges that state S_1 is more painful than S_2, and also judges that S_2 is more painful than S_1; an unconstrained response-dependence thesis for *more painful than* would then entail, falsely, that it is non-asymmetric. Worse, we can imagine a subject who judges that S_1 is more painful than S_1; unconstrained response-dependence would then entail, again falsely, that the relation is non-irreflexive. Thus, any response-dependence thesis for the relation *more painful than* must impose constraints that ensure that it satisfies various structural constraints such

as irreflexivity, asymmetry, and transitivity. Once these constraints are imposed, it is unclear why the response-dependence theorist would reject constraints which ensure the reflexivity, symmetry, and transitivity of *exactly as painful as*.

Let me consider three objections.[13] The first is that I am simply dismissing the Sorites. Nefsky (2011, 383-9) levels this charge at Kagan, whose argument we will briefly consider below. The standard Sorites, applied to the case at hand, involves the three (classically) jointly inconsistent claims that O_0 is not painful, that O_n is painful, and that for all j, if O_j is not painful, then neither is O_{j+1}. This is a genuine paradox, and I offer no solution here. Nor do I need to, for what matters is not (or at least not only) whether a given state is painful, but rather (or in addition) how painful it is.[14] And even if no single switch flipped can change the outcome from not painful to painful (a difference with respect to a vague, monadic property), this does not mean that no single switch flipped can affect the morally significant underlying dimension of how painful it is.

Now consider a different Sorites-like paradox, involving the four jointly inconsistent claims that O_0 is not painful, that O_n is painful, that if one state is exactly as painful as another then one is painful just in case the other is, and that for all j, O_j is exactly as painful as O_{j+1}. Here, the last claim is not intuitively compelling, once we distinguish between states being exactly as painful as each other, and their being merely almost exactly as painful as each other. Of course, one can make a theoretical argument in favor of this last claim, but it does not have the same intuitive pull as the third claim of the standard Sorites. Thus, I am

[13] A brief comment on a fourth objection: I am not reifying 'feels,' in the way that Dennett's (1978, xix-xx) imaginary society reifies 'fatigues.' That is, I do not focus on the relation *feels the same as*, analyze it as *has the same feel as*, and appeal to the fact that identity (and hence identity of feels) is an equivalence relation (see Williamson (1994, 179) for discussion).

[14] Bacon (2018) argues for the stronger conclusion that it is irrational to care intrinsically about the vague. For example, it is irrational to care about whether one is bald, over and above all the underlying facts relevant to baldness, such as how many hairs one has, how they are distributed, how people react to you, and so on. I am sympathetic to Bacon's claim, and to the analogous view that vague properties are not intrinsically morally significant. But for my purposes, I need only the weaker claim that it is irrational to care exclusively about the vague, and that facts involving vague properties do not exhaust what is morally significant, to the exclusion of the underlying more precise properties and relations.

not dismissing the Sorites, but only this latter psuedo-Sorites, which is no paradox at all.

The second objection is inspired by Temkin (2012, 164), who considers and rejects the claim that it is a conceptual truth that for all F, *is F-er than* is a transitive relation:[15]

> Consider the following example. Let us define the relation "larger than" as follows: for any two people a and b, a is *larger than* b if a is heavier than b *or* if a is taller than b. Clearly, so defined, a might be larger than b, because heavier, and b might be larger than c, because taller, yet a might *not* be larger than c, as c might be both heavier and taller than a. So it appears than one *could* have a "...er than" relation that is not transitive.

If Temkin is right, that would cast doubt on my claim that for any F, the relation *is exactly as F as* is an equivalence relation. For we could imagine defining *is exactly as large as* thus: for any two objects a and b, a is exactly as large as b if and only if either a is exactly as heavy as b or a is exactly as tall as b. Defined thus, we might have a, b, and c such that a is exactly as large as b, b is exactly as large as c, and yet a is not exactly as large as c.

My response is flatfooted: neither the comparative 'larger,' nor the relation it expresses, work in the way Temkin is imagining. And it is even clearer that neither 'is exactly as large as,' nor the relation it expresses, work in the way I just sketched. Admittedly, this dispute is difficult to settle. We are close to bedrock. But standard linguistic treatments of comparatives (Kennedy 2007; Schwarzschild 2008; see also Kamp 1975, 145) agree that 'is exactly as F as' and 'is F-er than' always express transitive relations. Indeed, it is hard to see how to devise a plausible compositional semantics for comparatives, including for the morpheme '-er,' the modifier 'exactly' (and the contrasting modifiers 'approximately' and 'roughly'), and especially for the positive form 'is F,' which rejects these claims (see Nebel 2018).[16] While I do not take this to decisively settle the matter, I think that it remains

[15] See also Temkin (1996) and Rachels (1998).

[16] A note on compositionality and 'exactly.' The expression 'is exactly as painful as' is not an idiom; it is not like 'kick the bucket,' where we understand the expression by learning it as a whole rather than by understanding the meanings of the component words and how they are put together. Therefore, our semantics should have it that 'exactly' means the same thing in expressions like 'is exactly as painful as' as

overwhelmingly plausible that 'is F-er than' and, more importantly for my purposes, 'is exactly as F as' always express transitive relations.

The third objection is that discriminability by humans may be necessary in order for two states to differ in a morally significant way. Differences in painfulness that cannot be detected by humans are not morally significant. On this view, while *is exactly as painful as* may be transitive, *is exactly as painful in the morally relevant sense as* is non-transitive, since indiscriminability is non-transitive.[17]

There are two points to make in response to this objection. The first is that this view entails that *is exactly as good as* is also non-transitive, which conflicts with the claim that *all* relations of the form *is exactly as F as* are equivalence relations, regardless of whether the relevant *F* is *painful*, *good*, or anything else. The second is that there is good reason to doubt that discriminability is necessary for the difference between two states to be morally significant. It is important to distinguish between cognitive (or belief-like) judgments about painfulness and the underlying painfulness itself. As defined above, indiscriminability is understood in terms of cognitive judgments: two states are indiscriminable (for an agent) with respect to painfulness just in case the agent is not in a position to know (or, perhaps, to reliably judge) that they differ in their painfulness. This is the sense in which indiscriminability may be non-transitive. But why privilege these cognitive judgments about phenomenology over the underlying phenomenology itself? If two states differ in how painful they are, why should the agent's inability to have knowledge of their differing painfulness mean that this

in expressions like 'has exactly two children' and 'arrived at exactly noon.' It also means that expressions of the form 'is exactly as *F* as' should work the same way regardless of whether '*F*' expresses a phenomenal property like *painful*, a non-phenomenal, descriptive property like *tall*, or a normative property like *good*. And they should work the same way regardless of whether '*F*' expresses a unidimensional property like *tall* or a multidimensional property like *large*. This suggests, for example, that 'is exactly as painful as' cannot mean the same as 'is indiscriminable with respect to pain from' (even setting aside the non-transitivity of indiscriminability). For an expression like 'is exactly as tall as' does not mean the same as 'is indiscriminable with respect to height from.' Perhaps it is impossible tell the difference between two things differing in height by a Planck length and their not differing in height at all (and we can even imagine this to be a nomological impossibility and not merely a practical one); hence two things could indiscriminable with respect to height even if their heights differ by a Planck length, but they would then not be exactly as tall as each other.

[17]Thanks to an anonymous reviewer for pressing me on this objection.

difference is morally insignificant?

To press the point further, consider a creature with less capacity for fine-grained introspective knowledge than humans. Certain animals might well fit the bill. Perhaps this creature has no capacity for cognitive judgments whatsover. In this case, all states count as indiscriminable for that creature. But, assuming that the creature can feel pain, and different levels of pain, it is implausible to say that none of the creature's possible pain states differ in their moral significance. Alternatively, we can imagine that the creature has the capacity for introspective knowledge, but only of a very limited and coarse-grained sort. While there are many different levels of pain that the creature can feel, it is only ever in a position to know that two pain states differ when one is very slight and the other very intense. Again, it is implausible that this epistemic limitation means that the difference between slight and moderate pain, or between moderate and very intense pain, is morally insignificant. The lesson is clear: differences in painfulness must sometimes be morally significant even when the subject is not in a position to have knowledge of their differing painfulness.[18]

Now I want to argue that my approach is superior to existing consequentialist treatments of harmless torturers-style cases. The first reason is that my approach does not rely on contentious claims about verbal reports and their relation to phenomenal states. Kagan (2011) notes that if asked in O_0 whether she is in pain, the patient will answer 'no,' while

[18]The objector might respond that the sense in which the tiny differences between adjacent states of the machine are undetectable is that they feel the same to the agent, and that if two states feel the same, they must be equally morally valuable. I have avoided talking in terms of the locution 'feels the same as' and instead focused on the relation *is exactly as painful as*, since the former locution is unhelpfully ambiguous. As Keefe (2011) notes, it has at least two distinct readings. First, there is a purely phenomenal reading, on which 'S_1 feels the same as S_2' can be glossed as 'The feel of S_1 is the same as the feel of S_2.' On this reading, 'feels the same as' expresses a transitive relation, since identity is transitive. However, one might doubt the legitimacy of reifying feels in this way, as noted in footnote 13. Second, there is a reading of 'feels the same as' on which it means 'feels as though they are the same.' This is a cognitive reading, on which two states feel the same roughly when the agent judges (or is inclined to judge), on the basis of introspection, that they are the same in the relevant respect. So understood, 'feels the same as' expresses a non-transitive relation. But the holding of this relation does not suffice for two states to be equally morally valuable, since as argued above, it is implausible to privilege cognitive judgments about, or based on, phenomenology to the exclusion of the underlying phenomenology itself. Absent some other proposed reading of 'feels the same,' I conclude that it never expresses a relation that is both non-transitive and such that its holding between two states suffices for them to be equally morally valuable.

if asked in O_n whether she is in pain, she will answer 'yes.' Hence there must be adjacent outcomes which differ with respect to the patient's answer to the question whether she is in pain. Kagan says that the two outcomes must therefore feel different.

McCarthy and Arntzenius (1997) previously gave a more sophisticated version of this argument. They imagine allowing the patient infinite time to play around with the machine, trying out each of the states multiple (even infinitely many) times, and each time recording her best description of how painful it felt, using whatever language she likes. Allowing the patient to try out each state multiple times mitigates worries about the possible instability of her responses and gives an accurate record of her overall dispositions with respect to how to describe her experience. But the basic argument is the same as Kagan's. The patient's verbal response disposition for O_0 clearly differs from her verbal response disposition for O_n. Thus, there must be two adjacent outcomes that differ, if only slightly, with respect to the verbal response dispositions they yield. And this means that those two adjacent outcomes must not feel the same to the patient.

This strategy has two significant limitations. First, verbal reports, and even long-run verbal report dispositions, need not accurately reflect underlying phenomenal states (Nefsky 2011). It might be that the patient is more disposed to report painfulness in O_j than in O_{j+1} even though they feel exactly the same. This is because one's verbal reports could be influenced not only by the underlying phenomenal states, but also by non-phenomenal states like tissue damage. The fact that two adjacent states yield differing verbal dispositions does mean that the agent is sensitive to some differences between the two outcome states, but not that those differences show up in her phenomenology.[19]

[19]Kagan (2011, 136) is alert to this problem and replies that 'it is important to bear in mind that these [reports] are indeed immediate and spontaneous reports concerning the qualitative aspects of the victim's experiences. The victim is simply reporting how the state *feels* to him, with regard to whether it involves pain, or whether the amount of pain differs from that involved in other states.' But it is not at all clear that even if we try, we can make our verbal reports sensitive only to our introspective phenomenology. Of course, Kagan can just stipulate that the patient is responding only to her phenomenology, but then the argument could be simpler, and indeed more similar to my own. He could leave out the verbal reports and simply point out that the painfulness of O_0 and the painfulness of O_n differ and that *feels the same as* is an equivalence

Second, even if we assume that differing verbal reports mean a difference in how the states feel, this would only show that there must be two adjacent states that don't feel the same. It would not show that there must be two adjacent states that fail to be exactly as painful as each other. For it is at least logically possible for one state to be exactly as painful as another state even though the two don't feel the same. Similarly, it is logically possible for two visibly different paintings to be exactly as beautiful as each other, or for two things to be (or look) exactly as red as each other without looking the same, for instance if one is a bit greenish and the other a bit blueish, but they are equally far from pure red.[20]

My strategy also has the advantage of not being tied exclusively to phenomenal properties. Nefsky (2011, 374) considers a consequentialist view on which fairness is morally relevant:

> Now, imagine that there is a large supply of clean water that two impoverished communities, A and B, have equal claim to and that will be distributed by an international committee. The fair outcome would be for the water to be divided approximately evenly between the two communities. *Approximately* evenly because—I think we can say—fairness is not, in this case, an extremely precise matter. A few drops of water more or less on one side does not make the distribution unfair (or even any less fair) in any morally relevant sense of the term.

But *is exactly as F as* must always be an equivalence relation, regardless of whether F is a phenomenal property, a normative property, or anything else. Hence *is exactly as fair as* is an equivalence relation as well. More generally, *is exactly as good as* is an equivalence relation, and so no matter what is morally valuable, there cannot be a sequence of states with the first better than the last but each exactly as good as its predecessor.

---

relation, and hence there must be adjacent states that don't feel the same.

[20]Even functionalists about pain will grant that pain involves not only verbal dispositions, but other dispositions as well, such as a disposition to grimace. We can then imagine two states S_1 and S_2, such that one is slightly more disposed to report being in pain in S_1 than in S_2, but one is slightly more disposed to grimace in S_2 than in S_1. These differing dispositions may suffice for the two states to feel different, but if they exactly balance each other, the states will nonetheless count as exactly as painful as each other.

One might object that in the cases I've discussed, the two things cannot be *exactly* as beautiful, as red, or as painful as each other if they look or feel different. Instead they can only be 'on a par' with respect to beauty or redness. I discuss parity in the next section.

4 Parity

But wait! I have argued that *is exactly as F as* is always an equivalence relation. But the fact that *is exactly as good as* is an equivalence relation does not suffice to show that all collective action problems are triggering cases. All that this fact shows is that there cannot be a sequence of states with the first better than the last and each state exactly as good as its predecessor. But a triggering case, as defined above, is one involving a sequence of states with the first better than the last and where some state is *worse* than its predecessor.

Arguably, one thing can be neither better, nor worse, nor exactly as good as another. Some philosophy job P might be neither better nor worse than some journalism job J, which is also neither better nor worse than the philosophy job with an extra \$100 ($P+$). Given the transitivity of *is exactly as good as*, P and $P+$ cannot each be exactly as good as J, since $P+$ is better than P. Instead, at least one of them must be *on a par* with J (Chang 2002).

Hence, it is compatible with the view I defended in the previous section that there be a sequence of states such that the first is better than the last and yet each state is neither better nor worse nor exactly as good as its predecessor in the sequence.[21]

The harmless torturers case is not such a case. Parity with respect to F-ness can arise only when there are multiple factors relevant to how F a thing is, but no precise way of assigning weights to those factors so as to enable precise trade-offs. For instance, if how large something is depends on both how heavy and how tall it is, but there are no precise weights assigned to the two dimensions of heaviness and tallness, there can be cases where one thing is neither larger nor smaller nor exactly as large as another. But in the harmless torturers case, the kind of pain involved in each state is the same, and the only thing that varies is its intensity. This means there will be no parity between any of the states.

But climate change may be a case in which there is a sequence of possible outcomes

[21] Compare Nefsky's (2011, 382) complaint that Kagan shows at most that some state must feel *different* than predecessor, and not that it feels *worse*. She does not discuss parity, however.

where the first is better than the last, but none is worse than its predecessor. Things would be better if everyone reduced emissions than if no one did so. But it may be that no tiny increment in emissions would make the outcome worse; instead, it would leave it exactly as good as, or on a par with, how it would otherwise have been.[22] After all, even where a small increment in emissions causes a morally relevant difference, the difference needn't be all bad: some people will feel less comfortable, others more so; some animals will have less food, others more; and so on. And there may be no way of assigning precise weights to these various harms and benefits. Indeed, in many collective action problems, our acts may affect the number and identities of the people who exist, and the kinds of harms and benefits that accrue to different people; these are the sorts of differences that may yield parity between outcomes. We can also modify the harmless torturers case to introduce parity:

> **Harmless Torturers (Parity Version)**
>
> There is a patient hooked up to a torture machine and n other people (including you), each facing a switch. If no one flips, the patient feels almost no pain. If everyone flips, the patient feels excrutiating pain. But there are two kinds of pain: burning pain and throbbing pain. And as more switches are flipped, the pain intensity increases, but alternates between burning pain and throbbing pain. So, O_0 means throbbing pain of intensity 1, O_1 means burning pain of intensity 2, O_2 means throbbing pain of intensity 3, and so on. In addition, there is parity among types of pain, such that burning pain of intensity x is neither better nor worse nor exactly as bad as throbbing pain of intensity $x \pm 1$.

Here, the outcome will be much better if no one flips their switch than if everyone does, but no outcome is worse than its predecessor. Thus, you know that, regardless of how many others flip, your flipping will not bring about a worse outcome than not flipping.

Does that mean that it is permissible for you to flip your switch? That depends on the correct theory of decision making under uncertainty *and parity*. Schoenfield (2014, 267) endorses a principle for decision-making under parity which says that it is permissible for you

[22] cf. Andreou (2006), who likens cases of pollution to Quinn's (1990) case of the self-torturer with intransitive preferences, though she does not claim that betterness itself is non-transitive (Andreou 2018).

to flip. For her LINK principle says, in part, that 'If you are rationally certain that neither of the two options [A and B] will bring about greater value than the other, it's not required that you choose A, and it's not required that you choose B.'

But a different decision theory, *Prospectism* (Hare 2010; see also Weirich 2004), prohibits flipping, at least given complete uncertainty about how many other people will flip. With parity, the betterness ordering can be incomplete and hence not representable by a value function that assigns one outcome a greater real number than another outcome if and only if the former is better than the latter. For it may be that O_1 is not better than O_2, O_2 is not better than O_3, and yet O_1 is better than O_3. But there are no real numbers such that $x \leq y$, $y \leq z$, and yet $x > y$. We can, however, consider value functions representing *coherent completions* of the betterness ordering, where V represents a coherent completion of that betterness ordering just in case, for all O_i and O_j, $V(O_i) > V(O_j)$ if (but not only if) O_i is better than O_j. (Think of a coherent completion of a betterness ordering as one that respects the original ordering's *better than* relations but also eliminates parity by taking each instance of some O_i being on a par with O_j and replacing it with O_i's being either better than, worse than, or equally good as O_j.) Prospectism then says:

> **Prospectism**: It is permissible to perform an action if and only if, for some value function V that represents a coherent completion of the betterness ordering, no alternative action has higher expected value relative to V.

My aim is not to provide a defense of Prospectism, but simply to show that it gives consequentialism a way of prohibiting exceptional acts in parity-laden collective action problems.[23] And let me first flag that my argument will rely only on the left-to-right (or 'only if')

[23]The debate over decision-making under parity and uncertainty has focused on a different kind of case (Hare 2010; Schoenfield 2014). Suppose A and B are on a par. As such, mildly sweetening one of them (say, by adding \$5 to it), converting it to $A+$ or $B+$, would not make it better than the other. Now suppose there are two opaque boxes. One contains A and the other contains B, with the arrangement determined by a fair coin. And you can see left-hand box has a \$5 note on top of it. Question: Are you required to take the (sweetened) left-hand box? Prospectism says 'yes.' For taking the left-hand box is associated with the prospect {0.5 chance of $A+$, 0.5 chance of $B+$}, while taking the right-hand box is associated with the prospect

direction of Prospectism, meaning that any decision theory which agrees with its necessary condition for permissibility will allow consequentialism to get the same desired result.

Now, suppose you are completely uncertain about how many other people will flip their switches, such that you have a uniform probability distribution over the states $S_0, ..., S_{n-1}$ (where S_j is the state where exactly j-many other people flip). In this case, Prospectism prohibits flipping. This is because each 'intermediate' outcome $O_1, ..., O_{n-1}$ has the same $(\frac{1}{n})$ probability of resulting from you flipping as from you not flipping and hence can be ignored; they cannot make a difference to the relative expected values of flipping and not flipping. (To illustrate, if $n > 17$, outcome O_{17} would result from your flipping if S_{16} is actual, and would result from your not flipping if S_{17} is actual, but by our setup S_{16} and S_{17} are equiprobable.) As for the non-intermediate outcomes O_0 and O_n, flipping has probability 0 of yielding O_0 and probability $\frac{1}{n}$ of yielding O_n, while not flipping has probability $\frac{1}{n}$ of yielding O_0 and probability 0 of yielding O_n. But since O_0 is better than O_n, each value function representing a coherent completion of the betterness ordering assigns the former a higher number than the latter. Therefore, relative to each value function representing a coherent completion of the betterness ordering, flipping has lower expected value than not flipping.[24]

Another way to put this is that flipping is *stochastically dominated* by not flipping, where A stochastically dominates B just in case for each outcome O, the probability of yielding an outcome at least as good as O is at least as great for A as for B, and for some outcome O^*,

{0.5 chance of A, 0.5 chance of B}, and relative to any value function representing a coherent completion of the betterness ordering, the former prospect has higher expected value than the latter. Schoenfield's LINK principle says 'no,' since you know that, no matter how the coin landed, the contents of the left-hand box are not better than the contents of the right-hand one (since $A+$ is not better than B, nor is $B+$ better than A). Without going fully into the arguments for and against each response to the two opaque boxes case, let me add that it is a messy affair designing a full decision theory that entails LINK. The only one I am aware of is Hare's *Deferentialism*, which is considerably more complex than Prospectism. If parsimony is a virtue in normative theorizing as well as in empirical inquiry, this may be one reason to favor Prospectism.

[24]Put in terms of symbols, the point is that for all V_i:
$EV_i(\neg\text{flip}) = \frac{1}{n} \times V_i(O_0) + \frac{1}{n} \times V_i(O_1) + ... + \frac{1}{n} \times V_i(O_{n-1}) + 0 \times V_i(O_n)$
$EV_i(\text{flip}) = 0 \times V_i(O_0) + \frac{1}{n} \times V_i(O_1) + ... + \frac{1}{n} \times V_i(O_{n-1}) + \frac{1}{n} \times V_i(O_n)$
Hence, $EV_i(\neg\text{flip}) > EV_i(\text{flip})$ iff $V_i(O_0) > V_i(O_n)$
But, since O_0 is better than O_n, $V_i(O_0) > V_i(O_n)$, and so for all V_i, $EV_i(\neg\text{flip}) > EV_i(\text{flip})$

the probability of yielding an outcome at least as good as O^* is strictly greater for A than for B. And Prospectism prohibits stochastically dominated actions (Bader 2018). (Note that flipping is also stochastically dominated in the original harmless torturers case, provided you have a uniform probability distribution over hypotheses about how many others will flip. Therefore, even if you are unconvinced by my treatment of that case, consequentialism will still prohibit flipping, given any decision theory that prohibits stochastically dominated acts.)

Thus, if Prospectism is correct, consequentialism can prohibit the exceptional act of flipping even in this parity-laden, non-triggering case. This result holds provided you are uncertain how others will act. Now, if you knew exactly how many others would flip, Prospectism would permit flipping. As for intermediate cases, where you are neither completely uncertain nor completely knowledgable about how others will act, the devil is in the details; whether Prospectism prohibits flipping will depend on exactly how far your probability distribution deviates from uniformity and 'how much' parity there is in the betterness ordering.[25] But

[25]If you deviate from a uniform probability distribution over the states S_0,...S_{n-1}, flipping is no longer stochastically dominated and hence *may* be permitted by Prospectism. For there will be some intermediate outcomes that are more probable if you flip than if you don't. If a value function assigns sufficiently high value to those outcomes, flipping will then have highest expected value. But whether such a value function is admissible will depend on the details of both the betterness ordering and your probability function.

The point is best seen with a modification of the opaque boxes case in footnote 23. In this new version, suppose that there is probability $n \neq 0.5$ that A was placed in the (sweetened) left-hand box. Taking the left-hand box (L) is associated with the prospect $\{n$ chance of $A+$, $1-n$ chance of $B+\}$, while taking the right-hand box (R) is associated with the prospect $\{1-n$ chance of A, n chance of $B\}$. So, for any value function V_i, $EV_i(L) = n \times V_i(A+) + (1-n) \times V_i(B+)$, while $EV_i(R) = (1-n) \times V_i(A) + n \times V_i(B)$.

Without loss of generality, let $n > 0.5$. Then, $EV_i(R) \geq EV_i(L)$ only if both (a) $V_i(B) > V_i(A+)$ and (b) $\frac{V_i(B)-V_i(A+)}{V_i(B+)-V_i(A)} \geq \frac{1-n}{n}$. Such an admissible value function will exist if the only facts about the betterness ordering are that $A+$ is better than A and that $B+$ is better than B. But in more realistic cases, there will be additional facts about the betterness ordering will constrain admissible value functions. When (a) obtains, $\frac{V_i(B)-V_i(A+)}{V_i(B+)-V_i(A)}$ increases (with an asymptotic upper bound of 1) as $V_i(A+)-V_i(A)$ and $V_i(B+)-V_i(B)$ decrease and as $V_i(B)-V_i(A+)$ increases. But additional facts about the betterness ordering may put a lower bound on how far $A+$ and $B+$ must be ranked above A and B, respectively, and an upper bound on how far B can be ranked above $A+$. How much by way of additional constraints are needed to render R impermissible depends on the value of n, with more constraints needed as n approaches 1. This is because $\frac{1-n}{n}$ decreases as n increases, thereby making it 'easier' to find an admissible value function that jointly satisfies (a) and (b).

The details are messy, but the moral is simple: In both the opaque boxes case and the parity version of the harmless torturers case, Prospectism prohibits taking the right-hand box or flipping your switch if your probability distribution over the relevant states is uniform. Otherwise, it *may* permit such acts, though this will depend on the details of the probability distribution and the betterness ordering involved.

this limitation may not be terribly serious, for we have already seen that consequentialism will not prohibit exceptional acts in *all* collective action problems. At most, it will do so in realistic cases in which you're very uncertain how others will act. I have shown that parity need not threaten this more modest claim.[26]

Is Prospectism compatible with consequentialism? There is one way of characterizing consequentialism on which the two are not compatible. On this gloss, consequentialism says that an act is permissible just in case there is no alternative act which would yield a better outcome. But this gloss is incompatible with *any* decision theoretic version of consequentialism, for it makes no reference to the agent's subjective probabilities. Fortunately, there is another standard gloss that leaves room for a decision theoretic version of consequentialism. On this gloss, consequentialism says that the good is prior to the right, such that an act's permissibility depends only on the values of the possible outcomes of the available acts, as well as the agent's (rational) subjective probabilities. This gloss is neutral with respect to which decision theory is correct and is therefore compatible with Prospectism.[27]

The issue, then, is not whether Prospectism is compatible with consequentialism, but simply whether it is correct. It is certainly somewhat counterintuitive that Prospectism sometimes prohibits an act which it is known will not yield a worse outcome than its alternatives. But while this constitutes probably the most significant objection to Prospectism, it does not mean that Prospectism is false, for there are also significant objections to non-Prospectist

[26]Broome (1997) regards incommensurability not as parity, but as vagueness or indeterminacy. Even if he is wrong, it may also be indeterminate what the likely outcome of some action would be, for instance if it is indeterminate what the threshold numbers are in a triggering case, or if it is indeterminate how precisely you would perform the act in question. To deal with indeterminacy, consequentialists would do well to adopt some Prospectism-like theory on which an act is prohibited if if it has sub-maximal expected value on to every admissible way of resolving any indeterminacy. (For epistemicists (Williamson 1994), indeterminacy is a kind of unknowability and hence rational uncertainty, and as such may not require any modification of expected value theory.) See Hare (2011), Moss (2015), and Williams (2017) for further discussion of decision-making given indeterminacy.

[27]Still, (act) consequentialism may be incompatible with certain ways of *motivating* decision theories. In particular, it would be in tension with defending a decision theory on the basis of its tending to yield better overall results in a sequence of choice situations (whether faced by a single agent at different times or by different agents). Without going into details, however, let me just say that all of the defenses of Prospectism cited below are based on grounds other than the results it is likely to yield when followed repeatedly.

theories that avoid this oddity (Hare, 2010; Bader, 2018; Doody, 2019; Rabinowicz, ms). Settling the matter requires evaluating all the arguments for and against Prospectism and its competitors, which is beyond the scope of this paper. Suffice it to say that Prospectism is a live possibility but that my conclusion is a conditional one: If Prospectism (or at least its necessary condition for permissibility) is correct, then consequentialism can prohibit exceptional acts even in parity-laden, non-triggering collective action problems.

5 Conclusion

Consequentialists have responded to a standard sort of collective action problem by arguing (i) that all such cases are triggering cases, and (ii) that exceptional acts are prohibited in triggering cases by virtue of having have sub-maximal expected value. Unfortunately, both claims are false. Expected value theory won't prohibit exceptional acts in some triggering cases, since they involve infinities. And some collective action problems are not triggering cases, since they involve parity.

Nonetheless, I conclude that many intuitively impermissible exceptional acts can still be prohibited on consequentialist grounds. First, while consequentialism cannot prohibit exceptional acts in some infinitary triggering cases, it likely does so in more realistic cases where we have strong judgments of impermissibility. In such triggering cases, most of us are sufficiently ignorant—both about the mechanisms involved and about how others will act—that consequentialism will prohibit the exceptional act on expected value grounds.

Second, while I have argued that there cannot be a sequence of outcomes with the first better than the last and each exactly as good as its predecessor, parity means there can be a sequence of outcomes with the first better than the last and each *not worse than* its predecessor. This means that not all collective action problems are triggering cases. Nonetheless, consequentialism can still prohibit exceptional acts in such cases, provided that

Prospectism (or, again, at least its necessary condition for permissibility) is correct.

I conclude that consequentialism can prohibit exceptional acts in many, if not all, of the sorts of collective action problems where we tend to judge that the exceptional act is indeed impermissible. While consequentialists will likely welcome this conclusion, it does not entail that the consequentialist treatment of collective action problems (let alone consequentialism more generally) is correct. Whether consequentialism gives the correct verdict about permissibility in *all* collective action problems, and whether it gives the correct *explanation* of these verdicts, are topics for another paper.[28]

[28]For helpful feedback, I would like to thank John Broome, Stephanie Collins, Mark Colyvan, Kevin Dorst, Luke Elson, Daniel Greco, Alan Hàjek, Caspar Hare, Shelly Kagan, Daniel Muñoz, Miriam Schoenfield, Roger Schwarzschild, Sam Shpall, Nicholas JJ Smith, Jack Spencer, and Daniel Wodak, as well as audiences at the University of Colorado-Boulder, MIT, Yale, the Australian National University, the Australian Catholic University, the University of Adelaide, and the University of Sydney.

References

Andreou, Chrisoula. 2006. 'Environmental Damage and the Puzzle of the Self-Torturer.' *Philosophy and Public Affairs* 34 (1): 95–108.

Andreou, Chrisoula. 2018. 'Better Than.' *Philosophical Studies*. Early view online.

Arntzenius, Frank, Elga, Adam, and Hawthorne, John. 2004. 'Bayesianism, Infinite Decisions, and Binding.' *Mind* 113 (450): 251–83.

Bacon, Andrew. 2018. *Vagueness and Thought*. New York: Oxford University Press.

Bader, Ralf. 2018. 'Stochastic Dominance and Opaque Sweetening.' *Australasian Journal of Philosophy* 96 (3):498–507.

Barnett, Zach. 2018. 'No Free Lunch: The Significance of Tiny Contributions.' *Analysis* 78 (1): 3–13.

Barnett, Zach. 'Voting to Change the Outcome is Rational.' Unpublished manuscript.

Broome, John. 1997. 'Is Incommensurability Vagueness.' In R. Chang (ed.), *Incommensurability, Incomparability and Practical Reason*. Cambridge, MA: Harvard University Press, 67–89.

Broome, John. 2004. *Weighing Lives*. Oxford: Oxford University Press.

Broome, John. 2018. 'Against Denialism.' *The Monist* 102 (1): 110–29.

Budolfson, Mark. 2018. 'The Inefficacy Objection to Consequentialism and the Problem with the Expected Consequences Response.' *Philosophical Studies*. Early view online.

Chang, Ruth. 2002. 'The Possibility of Parity.' *Ethics* 112 (4): 659–88.

Dennett, Daniel. 1978. *Brainstorms*. Montgomery, VT: Bradford Books.

Doody, Ryan. 2019. 'Opaque Sweetening and Transitivity.' *Australasian Journal of Philosophy*. https://www.tandfonline.com/doi/abs/10.1080/00048402.2018.1520269?journalCode=rajp20

Feldman, Fred. 1980. 'The Principle of Moral Harmony.' *Journal of Philosophy* 77 (3): 166–79.

Graff Fara, Delia. 2001. 'Phenomenal Continua and the Sorites.' *Mind* 110 (440): 905–35.

Hare, Caspar. 2010. 'Take the Sugar.' *Analysis* 70 (2): 237–47.

Hare, Caspar. 2011. 'Obligation and Regret When There is No Fact of the Matter About What Would have Happened If You Had Not Done What You Did.' *Noûs* 45 (1): 190–206.

Kagan, Shelly. 2011. 'Do I Make a Difference?' *Philosophy and Public Affairs* 39 (2): 105–41.

Kamp, Hans. 1975. 'Two Theories about Adjectives.' In E. Keenan (ed.), *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press, 123–55.

Keefe, Rosanna. 2011. 'Phenomenal Sorites Paradoxes and Looking the Same.' *Dialectica* 65 (3): 327–44.

Kennedy, Christopher. 2007. 'Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives.' *Linguistics and Philosophy* 30 (1): 1–45.

Lenman, James. 2000. 'Consequentialism and Cluelessness.' *Philosophy and Public Affairs* 29 (4): 342–70.

Lomasky, Loren and Geoffrey Brennan. 2000. 'Is There a Duty to Vote?' *Social Philosophy and Policy* 17 (1): 62–86.

McCarthy, David, and Arntzenius, Frank. 1997. 'Self-Torture and Group Beneficence.' *Erkenntnis* 47: 129–44.

Mills, Eugene. 2002. 'Fallibility and the phenomenal sorites.' *Noûs* 36 (3): 384–407.

Moss, Sarah. 2015. 'Time-Slice Epistemology and Action Under Indeterminacy.' In J. Hawthorne and T. Szabó Gendler (eds.) *Oxford Studies in Epistemology* vol. 5. Oxford: Oxford University Press, 172–94.

Nebel, Jacob. 2018. 'The Good, the Bad, and the Transitivity of *Better Than*.' *Noûs* 52 (4): 874–99.

Nefsky, Julia. 2011. 'Consequentialism and the Problem of Collective Harm: A Reply to Kagan.' *Philosophy and Public Affairs* 39 (4): 364–395.

Norcross, Alastair. 1997. 'Comparing Harms: Headaches and Human Lives.' *Philosophy and Public Affairs* 26 (2): 135–67.

Norcross, Alastair. 2004. 'Puppies, pigs, and people: Eating meat and marginal cases.' *Philosophical Perspectives* 18 (1): 229–45.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Portmore, Douglas. 2018. 'Maximalism and Moral Harmony.' *Philosophy and Phenomenological Research* 96 (2): 318–41.

Quinn, Warren. 1990. 'The Puzzle of the Self-Torturer.' *Philosophical Studies* 59 (1): 79–90.

Rabinowicz, Wlodek. 'Incommensurability Meets Risk.' Unpublished manuscript. Available: https://www.york.ac.uk/media/ppe/documents/Incommensurability%20meets%20risk.pdf

Rachels, Stuart. 1998. 'Counterexamples to the Transitivity of *Better Than*.' *Australasian Journal of Philosophy*. 76 (1): 71–83.

Schoenfield, Miriam. 2014. 'Decision Making in the Face of Parity.' *Philosophical Perspec-*

tives 28 (1): 263–77.

Schwartzschild, Roger. 2008. 'The Semantics of Comparatives and Other Degree Constructions.' *Language and Linguistics Compass* 2 (2): 308–31.

Singer, Peter. 1980. 'Utilitarianism and Vegetarianism.' *Philosophy and Public Affairs* 9 (4): 325–37.

Temkin, Larry. 1996. 'A Continuum Argument for Intransitivity.' *Philosophy and Public Affairs* 25 (3): 175–210.

Temkin, Larry. 2012. *Rethinking the Good*. New York: Oxford University Press.

Weirich, Paul. 2004. *Realistic Decision Theory*. Oxford: Oxford University Press.

Williams, J. Robert. 2017. 'Indeterminate Oughts.' *Ethics* 127 (3): 645–73.

Williamson, Timothy. 1994. *Vagueness*. New York: Routledge.

Williamson, Timothy. 2013 (1990). *Identity and Discrimination*. Revised and Updated Edition. Oxford: Wiley-Blackwell.