SHELLY KAGAN

Do I Make a Difference?

I

Let's start with a story from the "good old days."[1]

Imagine that I used to run a factory, from which I periodically released some of the poisonous waste (by-products of the manufacturing process) into the stream which ran behind the plant. The stream traveled down to the nearby village, where some of the water was drunk by a girl who died from the poison.

Obviously enough, this is not at all a good story from the standpoint of morality, nor from the standpoint of the victim. But it is, in its way, a satisfying story from the perspective of consequentialist moral theory, for it is an example which can be straightforwardly handled in familiar consequentialist terms. Had I not released the waste into the stream, the girl would not have drunk poison, and so would not have died prematurely. Which is to say: the results would have been better had I acted differently; so my act was wrong.

To be sure, not everyone finds this consequentialist analysis an adequate account of why it was wrong for me to act as I did. But my concern here is not to convince anyone of the general plausibility of this basic consequentialist account of right and wrong. Rather, my goal is to examine a different sort of case, a case where it is typically thought that the standard consequentialist analysis yields what is clearly the wrong answer. We'll turn to a few examples of the sort of case I have in mind in a moment. But before that, let me clarify a few points.

1. As will soon be obvious to all who know the book, this essay is very heavily indebted to Chapter 3 of Derek Parfit's *Reasons and Persons* (Oxford: Oxford University Press, 1984). Indeed, it seems fair to say that my entire paper is a long-percolating reaction to that incredibly stimulating discussion, which I first read more than thirty years ago.

 Philosophy & Public Affairs 39, no. 2

The problem, in effect, is this: consequentialism condemns my act only when my act makes a difference. But in the kind of cases we are imagining, my act *makes* no difference, and so cannot be condemned by consequentialism—even though it remains true that when enough such acts are performed the results are bad. Thus consequentialism fails to condemn my act.

In cases of this sort, therefore, consequentialism seems to fail even by its own lights. For here—unlike the deontological cases that purport to show that something else matters besides results—the act seems wrong precisely because of the *bad results* of everyone's doing acts of the same sort. Yet consequentialism still cannot condemn the act. Apparently, then, consequentialism fails to handle a kind of case that even consequentialists admit it ought to be able to handle.

III

It may be helpful to have a few examples of this new type of problem before us. (In point of fact I believe that many of these cases are typically misdescribed. But for the time being I am going to speak with the vulgar—and a good many of the learned.) Consider, first, an "updated" version of our pollution story: imagine, as before, that I run a polluting factory, but suppose now that my factory releases its toxins into the air through a smokestack. And imagine that the smokestack is sufficiently tall that the pollutants are swept up into the stratosphere, where they are so scattered by the winds that when the toxins do come back down to the surface of the earth they are spread over a very wide area—indeed, spread so thin that no single individual ever breathes in more than a single molecule from my plant.

Now it seems that we can easily imagine that the following is the case as well: a single molecule of the toxin makes no difference to anyone's health. To be sure, if enough molecules are taken in, the result is sickness or death; but one molecule, more or less, simply doesn't make any difference at all to anyone's health.

Imagine, next, that there are thousands, or tens of thousands, of similarly polluting factories around the nation (or the world). Each scatters its toxins so widely that no single individual ever takes in more than a single molecule from any single plant. But because there are indeed thousands of such factories, many people do take in enough of the toxin

to become ill. This is clearly a bad result, and we can certainly imagine that it is a bad enough result to outweigh whatever good is done by running the factories in this way (as opposed to some alternative way that would dispose of the toxins more safely). But for all that, it seems as though each factory owner can truthfully say to himself that it makes no difference whether or not *he* pollutes, for his decision puts at most one extra molecule of toxin in any given individual, and by hypothesis a single molecule, more or less, simply doesn't make a difference to anyone's health. When I think about my own decision whether or not to pollute, then, I have to admit that my polluting doesn't actually harm anyone, since it doesn't make a difference to anyone's health. And this means that consequentialism cannot condemn my act. The results would not be better if I didn't pollute: they would be the same (or perhaps slightly worse, given the lost profit, say, from running the factory in a less polluting manner).

Here, then, we seem to have a case of the sort under consideration. If enough factory owners pollute, the results are bad overall; yet each can truthfully say that his *own* act of polluting does not have bad results, since it makes no difference whether or not he pollutes. Thus consequentialism appears incapable of condemning the actions of the individual factory owners—even though the results of everyone's polluting are stipulated to be bad overall.

Some find it difficult to see how it could be true that each *individual* act makes no difference at all (regardless of how many other such acts are occurring) even though the cumulative effect of many such acts is undeniably bad. But this may be easier to grasp if we shift the pollution case yet again, so that the pollutants are not poisonous, but simply darken the sky. When enough such particles are released, the skies will be gloomy and gray—lowering the pleasure of those who are forced to live in such haze. But for all that, if a given factory's pollution is sufficiently widespread, no single person might ever have more than a single particle from a given factory in their visual field; and it seems easy enough to see how the presence or absence of a single particle might not be able to make any difference at all to the visual experience of a given viewer. Here then we have a particularly compelling case of the sort we are looking for: although the result of everyone's polluting is bad, my own decision whether or not to pollute makes no difference at all. And the same thing might be true of everyone.

It is sometimes suggested that cases that have this structure are a peculiarly modern problem, turning, as they do, on the cumulative effects of perhaps thousands—or even millions!—of agents, each of whose acts is individually too "small" to make a difference (acts which, nonetheless, produce undeniably bad results when combined in large enough aggregates). In the good old days, with which our essay began, if an individual produced harmful results he did at least some of this on his own; but the modern condition is such that we together often produce bad results even though none of us as individuals make any difference at all.

But whether or not cases of this sort are more prevalent nowadays than in the simpler times past, I hope it is clear that one hardly needs complicated technology to generate such examples and that at least some cases of this sort can arise in fairly "primitive" cultures. Thus, for example, a fishing village might depend for its livelihood on the small lake nearby. We can imagine that if each fisher restricts their catch to a given, fixed number of fish, the small but stable fish population in the lake will be able to reproduce and sustain itself, and the village's lifestyle can continue indefinitely. Furthermore, however many fish are taken, one fish more or less would not make a difference to the ability of the fish to successfully reproduce. But if several dozen people take an extra fish each, the fish population will crash, and the villagers will all suffer. Here, then, we have a case where if everyone overfishes by a single fish the results are bad, yet each can, it seems, correctly say that their own act of taking an extra fish simply makes no difference to the outcome at all: even if they hadn't taken the extra fish, the result would have been just as bad.

Here is a somewhat similar case in more modern dress. When I go to the supermarket, there are a number of dead chicken carcasses available for my purchase at the butcher's counter. Let us suppose, as seems to me to be the case, that the suffering a given chicken undergoes in being raised and slaughtered under contemporary factory farming methods is greater than the pleasure I get from eating the chicken (rather than some vegetarian alternative). Does this mean that consequentialism can condemn my purchase of a chicken at the grocer? Unfortunately, apparently not, for it seems to be the case that whether or not I buy a chicken makes no difference at all to how many chickens are ordered by the store—and thus no difference in the lives of any chickens. To be sure,

when hundreds of thousands of us each buy a chicken this week, this does make a difference—for if several hundred thousand fewer chickens were sold this week, the chicken industry would dramatically reduce the number of chickens it tortures. Thus the overall result of everyone's buying chickens is bad. But for all that, it seems true that it makes no difference at all whether or not *I* buy a chicken; even if I don't buy one, the results are no better.

Clearly, examples like these can easily be multiplied. In each case, there is a bad result overall when a large enough number of us perform some act. Precisely because of these bad results, we want to condemn the acts in question. So these ought to be among the kinds of cases that consequentialism can readily handle. Unfortunately, however, in each case it also appears to be true that no *individual* act makes any difference at all. But if it makes no difference whether or not I act, then consequentialism cannot condemn my act. Thus consequentialism appears unable to condemn any of the acts in question. Accordingly, even those sympathetic to consequentialism should find such cases troubling.

IV

I believe that appearances are misleading. I think that the cases we have been considering do not pose a genuine difficulty for consequentialism. The *collective action problem*—as we might dub it—can be straightforwardly solved along consequentialist lines. (There are several other problems also known as "collective action problems." But for the purposes of this paper, let us restrict the term to cases of the sort we have been describing.) Appearances to the contrary notwithstanding, it isn't really true that in collective action cases each individual's act makes no difference. Thus I am going to argue that we can indeed condemn the given acts for familiar consequentialist reasons.

But before turning to that argument, it may be helpful to note some of the other ways that one might propose to try to solve the collective action problem. Each of these represents some sort of departure from the standard consequentialist approach that we have considered up till now (though the departure grows progressively smaller), and each faces its own difficulties. I won't attempt a systematic comparison of the various advantages and disadvantages of each alternative. Indeed, for the most part, I'll barely pause to indicate what those difficulties might be. My goal

here is simply the more limited one of indicating some possible solutions of which I won't avail myself.

First, then, faced with collective action cases one might appeal to some sort of universalizability test, asking, "what if everyone did that?" If an act is forbidden when the results would be bad were everyone to act in the same way—even if the specific act token in question would not (or does not) actually have bad results at all—then we can condemn the relevant acts in the various examples we have considered. For example, even if my individual act of polluting has no bad results, the results are certainly bad when everyone pollutes, and so my individual act of polluting will be forbidden as well.

A second, similar solution is offered if we shift our allegiance from consequentialism at the normative level to consequentialism at the foundational level, and accept *rule* consequentialism, which insists that right and wrong is not a matter of the actual results of the given individual act, but rather a matter of conformity to the optimal rules—where rules are evaluated in terms of the overall results were everyone to conform to them. Presumably, a set of rules permitting pollution will have worse results than one that forbids polluting; and so my individual act of polluting will be forbidden, even if it makes no difference at all whether or not I pollute.

Despite the appeal to consequences that admittedly plays a role in both of these approaches, neither approach is likely to be attractive to those sympathetic to the kind of normative consequentialism that is our concern.[2] For both approaches are in tension with the thought that the rightness or wrongness of a given act should depend upon the consequences of *that* act.

Accordingly, we can do somewhat better from the perspective of (normative) consequentialism if we turn instead to a third possible solution, where we supplement the consequentialist's account of individual duty with an account of *group* duties. Here, the idea is that in addition to looking at the individual results of individual acts, we must also look at the collective results of actions done by groups. Since we together can do

2. In saying this, I am obviously presupposing a distinction between consequentialism at the foundational level, and consequentialism at the normative, or factoral, level. See "The Structure of Normative Ethics," *Philosophical Perspectives* 6 (1992): 223–42, or *Normative Ethics* (Boulder, Colorado: Westview, 1998).

what I by myself cannot, we must also look to see whether group actions make a difference, and we can then condemn those cases where the results would be better were the group to have acted differently.[3] Presumably this allows us to condemn what is done in the case where we all pollute: for even if it is true that each of us meets his own individual consequentialist duties in this situation (for it would have made no difference had I acted differently), we all fail to meet our *collective* consequentialist duties (for it would make a significant difference indeed had we all acted differently).

While some sympathetic to consequentialism may find this appeal to group duties a congenial one, it is not altogether transparent how this sort of approach is actually to be applied to specific cases, for it is not altogether clear how such collective moral duties should impinge on the decision making of a given individual. Even if the group has a duty to act differently, I am not the group, but only a member of the group. Should we say, then, that each of the members of the group individually inherits the duties of the group? If not, then we cannot condemn any of the individual acts of pollution, and so we would not yet have an adequate response to the collective action problem. But even if group duties *can* be inherited in this way, it is not yet clear what one should do when individual duties and group duties conflict. If my individual duties pull me one way, and my (inherited) collective duties pull another, where does my overall duty lie?

I don't want to insist that adequate answers to these questions cannot be provided. But perhaps the difficulties raised by these questions suffice to give us reason to look for a consequentialist solution to the collective action problem that focuses on individual duties alone.

A fourth possible solution suggests that we can solve the collective action problem given an appropriate method of "bookkeeping." In the collective action cases, there is a large bad result caused by cumulative effects of n acts of the given kind. Perhaps then we should say that each individual act is "responsible" for $1/n$th of the bad results. This would suffice to allow us to condemn each act of pollution along individualistic consequentialist grounds. (After all, by hypothesis, when we all pollute, the bad results of doing this, B, are greater than the good results (if any)

3. The possibility of such collective consequentialist duties is considered by Parfit in Section 26 of *Reasons and Persons*.

of doing this, g.[4] But if $B > g$, then $B/n > g/n$. Thus if each individual can look at his share of the total results as the difference his act makes, then each of us can correctly say that his act does more bad than good overall. And if the overall results of my act of polluting are bad, then (individualistic) consequentialism can condemn it after all.)

Unfortunately, the creative bookkeeping endorsed by this approach is problematic. Even if it gives us the answer we want in some cases, in other cases it will give unacceptable answers.[5] But regardless of this difficulty, it is not clear that consequentialists can legitimately avail themselves of this proposal, for it seems to me that it actually abandons the fundamental consequentialist idea of looking to see what *difference* a given act makes. Being part of a group that together brings about bad results simply does not imply that my *own* act has bad results—or any results at all. My "share" of the total results has no particular connection at all to the actual results of my own individual act; it tells us nothing about whether or not my act actually makes a difference. Accordingly, I believe, appeal to such shares has no legitimate place in consequentialist thinking.

V

A final possible solution—our fifth—takes seriously the idea of looking to see what difference my individual act makes. It simply insists that results need not be perceptible to be real, and they need not be perceptible to count morally. But once we allow for the possibility of *imperceptible* harms (or benefits), we can insist that in the collective action cases it simply isn't true that my individual act makes no difference. On the contrary, it does make a difference. In the pollution case, for example, there is more toxin released as a result of my act—and while this may not leave any given individual perceptibly worse off (since one molecule more or less makes no perceptible difference), we can say that those who inhale a molecule of my toxin have been made imperceptibly worse off. Such an

4. In the kinds of cases that concern us, it is always true that $B > g$. Accordingly, I write the B in upper case, and the g in lower case, to help remind us of this fact.

5. For example, if five of us together save 25 lives, then according to this view even if only four of us were actually *needed* to save all those lives, each of us is "responsible" for saving 5 lives; and this would—unacceptably—be enough to justify my forgoing the chance to save the life of a *different* person who would otherwise die. (Examples like this can be found in Parfit, *Reasons and Persons*, Section 25.)

imperceptible harm will, obviously, be very small, but since I will have similarly harmed thousands, or millions, the cumulative amount of harm that I will have done will be very great—indeed, $1/n$th of all the harm done by the n polluters. Thus my act does make a difference, and the results would have been better had I not polluted. Had I not polluted, my myriad victims would each have been harmed (imperceptibly) less.[6]

This last approach unambiguously satisfies the constraints set by individualistic consequentialist analysis. The only question is whether or not we can accept the claim about the nature of good and bad results that it puts forward, that is, the claim that bad results (say) can be real and relevant to moral accounting, even if so small as to be imperceptible.

Now whether we find this plausible or not may well depend on whether we accept a mental state theory of well-being, such as hedonism. For if hedonism were true, then the only way that someone could be (directly) harmed would be by causing them pain, and thus to postulate the existence of imperceptible harms would be to postulate the existence of imperceptible pain—or of imperceptible differences in the level of one's pain. And this, obviously, might well give one pause. Thus, if one accepts hedonism, one might well resist the appeal to imperceptible harms as a solution to the collective action problem.

Of course, many reject hedonism, including many contemporary consequentialists. So I would not want to dismiss the possibility of a partial solution along these lines. Nonetheless, I do not think that such an appeal to imperceptible harms can constitute a complete solution to collective action problems. For even though there may be components of well-being that go beyond one's experiences—and thus can plausibly be thought to come in imperceptible amounts—it seems undeniable that one important component of well-being is indeed the presence of pleasure and the absence of pain. Imagine, then, that we face a collective action problem where the bad result in question is simply the pain we are together causing—but where each individual act makes no perceptible difference to anyone's pain. Here it is difficult to see how an appeal to imperceptible harm can help us, for such an appeal must again posit the existence of imperceptible pain, or imperceptible differences in the level of pain, and it must claim that these are to count as morally bad

6. The appeal to imperceptible harms is endorsed by Parfit in *Reasons and Persons*, Section 29.

results. But this, as we have seen, is difficult to believe. Even if there is more to well-being than hedonism posits, it is difficult to take the idea of an imperceptible pain, or an imperceptible increase in the level of pain, seriously, or to believe that it could be bad to impose such a thing. When the relevant bad outcome is pain, what matters morally are differences in how people *feel.*

Suppose, for example, that someone is wired to a torture machine with a thousand identical switches.[7] When none of the switches are flipped, no current runs through the machine, and so the victim is in no pain at all. If all thousand switches are flipped, then a sizable current runs through the machine and the victim is in tremendous pain (but no permanent damage is done to his body). But the flipping of any given switch increases the current only by a very small amount (well below the perceptually discriminable threshold for pain) so that the victim simply cannot tell whether one switch more or less has been flipped—regardless of how many other switches have already been flipped. (If one thousand switches isn't enough to guarantee that the difference between any two steps is imperceptible, then make it ten thousand switches, or a million!) Finally, imagine that a thousand different people each control a single switch and must decide whether to flip it or not.

Now suppose that 778 of us have flipped our switches. The victim is in a considerable amount of pain. But consequentialism cannot condemn my act unless my flipping my switch makes a difference. By hypothesis, however, whether or not I flip my switch makes no perceptible difference to our victim. He cannot tell the difference between the situation in which 778 switches are flipped and the situation in which a "mere" 777 switches are flipped. The two states feel the same.

Admittedly, if we appeal to the existence of imperceptible harm, we can disregard this fact and condemn my act on the grounds that throwing the extra switch does indeed harm the victim, albeit imperceptibly. But this move seems implausible here (regardless of its plausibility elsewhere), for the only harm that seems relevant to this case is the *pain* the victim is in, and it is difficult to see why it should be bad to increase pain in an imperceptible way. Indeed, it isn't even clear that it makes any *sense* to say that pain has been increased imperceptibly. On the contrary, it seems that the *pain* hasn't increased at all. To be sure, the current has

7. I here modify an example of Parfit's. See *Reasons and Persons*, Section 29.

been increased, and perhaps the victim's neurons are firing at a faster rate—but none of this seems morally relevant, given that the victim isn't in any greater pain.

Thus, even if an appeal to imperceptible harms helps us in some collective action cases, it doesn't seem to help us here. My throwing the extra switch doesn't seem to make a difference that matters morally, and so the consequentialist doesn't yet have grounds for condemning it. A complete solution to the collective action problem, along lines acceptable to the individualistic consequentialist, is still wanting.

VI

How, then, should the consequentialist respond to collective action problems? The first thing to do, I believe, is to try to get clearer about how cases of this sort could arise. How can it be that when enough acts of a given sort are done, the results are bad, and yet at the same time my own individual act makes no difference? Further reflection reveals, I think, that two basic sorts of cases seem possible.

In one kind of case, it isn't literally true that my individual act makes *no* difference at all. Rather, my act makes a real, but imperceptible, difference along some dimension. However, even though small enough changes along this dimension are imperceptible, enough such differences can add up to a sizable difference, and sizable differences can be perceived. If what matters morally is making a difference to perception, then perceptible differences to the underlying magnitude will be morally relevant, but imperceptible ones will not be. Thus, when enough of us perform the act in question, the results will be bad, but for all that, my individual act will make no morally relevant difference at all. To be sure, strictly speaking my act does make a *difference,* since it makes an imperceptible difference along the underlying dimension; but this difference is imperceptible, and so my act makes no morally *relevant* difference: the results would not be *better* had I acted differently.

In the torture case just considered, for example, pain is obviously morally relevant, and it depends upon the amount of current. But it seems plausible to claim that imperceptible differences in the amount of current have no moral relevance in themselves, even though large enough differences alter the amount of pain the victim feels. When I flip an extra switch, I make a difference to the current, but none to the pain,

so I make no difference that matters morally. Yet when enough of us do this, this makes a perceptible difference to the current, and thus to the victim's pain. Thus we together produce a bad result, even though the results would not have been better had I acted differently.

It is tempting to conjecture that all of our collective action cases are like this—where imperceptibly small changes along some underlying dimension aggregate into morally relevant (because perceptible) changes. But while some additional examples certainly seem to be of this sort (for example, the darkened skies pollution case), others, I think, are more readily understood as being instead of a second kind. Here, instead of insisting that every act makes some difference, but that no act makes a perceptible difference, it seems more accurate to suggest that most acts make no difference, but some single act makes a great deal of difference. (This account will eventually need to be refined, but it will do for now.) In effect, some single act works as a trigger—bringing about the morally relevant difference. But most acts make no difference at all. It might be, for example, that a certain number of such acts can occur without any harmful effect whatsoever, but then a threshold point is reached, and a single further act triggers the harmful result. If we don't know which act was the triggering act, then it seems that all we can say is that our own individual act is overwhelmingly likely not to have made a difference.

The fishing village example is perhaps best viewed as being of this sort. After all, it is not as though my catching an extra fish makes an imperceptible difference to the decline of the fish population's ability to replenish itself. Rather, up to a point, as each person takes one extra fish, this makes no difference at all—the fish will be fully capable of replenishing their population, despite the extra fish taken. But at a certain point—an unknown point—one too many fish is taken, the stock is no longer large enough, and the population crashes. Thus, when we all overfish, and the population crashes, we know that someone triggered the crash, though it is overwhelmingly likely that my own act of overfishing made no difference at all.

The example of buying chicken is likely of this sort (most purchases make no difference, but someone's purchase triggers an increase in the order placed next week), and the toxins in the smokestack case may be as well (up to a certain threshold it doesn't matter whether you have any molecules of the toxin in your body, but there is a threshold point, and the *n*th particle triggers the illness). I imagine that many of the cases that

we might initially be inclined to discuss in terms suggesting imperceptible differences might be better understood as triggering cases. But my purpose here is not to debate the details of any given case. It is rather to suggest that these two sorts of cases—imperceptible difference cases and triggering cases—are importantly different in terms of their underlying mechanisms, and thus require different analyses.

In point of fact, it seems to me likely that many cases are probably best understood (at least initially) as being mixtures of these two pure sorts. It might be, for example, that more than one act works as a trigger (to higher and higher levels of harm), and of those acts that do not act as triggers some make no difference at all, while others make imperceptible differences (which can aggregate to make a perceptible difference). If so, my two pure types don't actually exhaust the field. Still, if all cases can be viewed as involving one or the other or both of these two types of mechanisms, then adequate analyses of the two pure types may suffice to allow an adequate understanding of all of the collective action cases.

VII

Let us start by considering the (pure) triggering cases. Here, roughly, the idea is that it is indeed true for most acts that it makes no difference whether or not I do it, but for some act—the triggering act—it makes all the difference in the world. In the simplest cases like this (with just a single triggering action) if *n* of us perform the act, the results are bad, but this is because one of the *n* acts triggered those bad results. We don't know who performed this act, but we do know that for the rest of us, it made no difference whether or not we acted. Thus it is overwhelmingly likely that my own act made no difference.

How, then, can the consequentialist condemn my act? The key to the answer lies in the thought that it is only overwhelmingly likely that my act made no difference. It is unlikely, but possible, that it did make a difference—that my own act was the triggering act. But if it was, then of course it made a very significant difference indeed, for the triggering act brought about the various bad results.

What we have, then, is a familiar case of decision making under uncertainty. I cannot know for sure that my act brought about the bad results—indeed, I can know that most likely it did not: but even when I discount the overall bad results for the high likelihood that my act did

not bring them about, the net result of doing this remains negative. That is, my act has a negative expected utility. And that is why, from a consequentialist perspective, it should not be done.[8]

Somewhat more fully: If we suppose for simplicity that each individual act does the same amount of good, if any, then the expected utility of my act is $(1/n)(B + g/n) + ([n-1]/n)(0 + g/n)$, where B stands for the (total) additional bad results, and g stands for the (total) additional good results. This reduces to $g/n + B/n$. But in the kind of cases that concern us, the total (extra) overall results are negative. That is, B is greater than g, and so B/n is greater than g/n. Thus the expected utility of my act is negative. That is why the consequentialist condemns it.

(Strictly speaking, the calculation I just sketched isn't for the act's *total* expected utility, but rather for its *incremental* expected utility—the *difference* in expected utility if we perform the act in question rather than an alternative where the good results may be smaller but the chance of my triggering the extra bad is avoided. A negative number here indicates that the given act's *total* expected utility is lower than that for the alternative; and strictly speaking, *that's* why consequentialism condemns it. In any event, for simplicity I will continue to just talk about an act's having negative "expected utility," rather than negative "incremental" expected utility.)

It should be noted that this sort of account is only available for collective action problems that involve triggering. In triggering cases there is a small chance that the act makes a big (morally relevant) difference. And while the chance is only a small one, the difference it makes, if it does make a difference, is sufficiently great to guarantee that the expected utility of the given act is negative. That is the reason the consequentialist can condemn it.

8. For the purposes of the present paper I will simply take for granted the familiar assumption that consequentialism does indeed condemn acts with negative (incremental) expected utility. But we should not forget that many such acts (indeed, in the cases we are currently considering, most such acts) will actually have *good* results overall. So the question arises, what is the exact nature of the consequentialist condemnation of acts with negative expected utility, and what (if anything) justifies it? That's obviously an important question, but I won't try to answer it here. (Note, however, in this connection, that in the text I say only that consequentialism "condemns" such acts, holding that they "should not be done," rather than saying that it "prohibits" such acts, holding them to be "wrong" or "forbidden.")

(In more complicated triggering cases, of course, instead of there being a single triggering act which brings about *all* of the bad results, there may be a series of triggering acts, each of which brings about only a *portion* of the bad results. Admittedly, in such a case, even if you do one of the triggering acts the bad results you bring about are less than B. But there is, of course, a correspondingly greater chance that your act will *be* one of the triggering acts, so the expected utility of your act remains negative.)

In contrast, in (pure) cases of imperceptible difference there is *no* chance that a given act makes a morally relevant difference. The expected utility of my act remains, at worst, zero. Indeed, if $g>0$, the expected utility of my act will actually be positive: g/n. For this reason, cases of imperceptible difference are more problematic from the perspective of consequentialism. Triggering cases can be handled through a straightforward appeal to expected utility. Imperceptible difference cases, unfortunately, require a more complicated line of attack.

Many will find it clear enough how this sort of appeal to expected utility suffices to disarm the triggering cases. But some people find it difficult to see how the story is supposed to go. Accordingly, I am going to offer an extended discussion of one of our examples—the chicken buying example—so as to make the argument more explicit. Doing this will also allow us to move to a somewhat more realistic picture of how triggering actually works. But despite the extra details and complications that we are about to introduce, the basic idea will remain unchanged: in triggering cases the expected utility of my act is negative. That is why the consequentialist condemns it. Those already convinced of this point can skip ahead to Section XI.

VIII

Return, then, to my grocer, where the poultry shelf has a large number of already slaughtered chickens awaiting purchase. Intuitively, it makes no difference to the person running the butcher department whether or not I buy one of these chickens. The butcher is not paying close enough attention to know exactly how many chickens he has sold.

Things would be quite different if each time I bought a chicken, the butcher got on the phone to the chicken farm and told them to torture and then slaughter another chicken and send it over. Then, obviously enough, each act of purchasing a chicken would make a difference.

And given the assumption (which I am taking as granted) that the suffering that a chicken undergoes (being raised under the conditions of contemporary factory farming) is greater than the pleasure I get from eating the chicken, it would immediately follow that my act of purchasing a chicken does more harm than good. But this fantasy of a butcher completely attentive to my decision to purchase a chicken is indeed just that: a fantasy. In the real world, the butcher simply doesn't pay that kind of attention.

But that does not mean, of course, that he is completely *inattentive* to the purchases made either. He had better give some kind of attention to the number of chickens sold in a given day, say, or a given week. Otherwise he has no idea how many chickens to order for the next shipment. Presumably it works something like this: there are, perhaps, 25 chickens in a given crate of chickens. So the butcher looks to see when 25 chickens have been sold, so as to order 25 more. (Perhaps he starts the day with 30 chickens, and when he gets down to only 5 left, he orders another 25—so as never to run out. But he must throw away the excess chickens at the end of the day before they spoil, so he cannot simply start out with thousands of chickens and pay no attention at all to how many are sold.)

Here, then, it makes no difference to the butcher whether 7, 13, or 23 chickens have been sold. But when 25 have been sold this triggers the call to the chicken farm, and 25 more chickens are killed, and another 25 eggs are hatched to be raised and tortured. Thus, as a first approximation, we can say that only the 25th purchaser of a chicken makes a difference. It is *this* purchase that triggers the reaction from the butcher, this purchase that results in more chicken suffering. (Actually, we still have a simplification here, for presumably all that the 25th purchase really triggers is a call to the general supplier—who in turn only makes a call to the chicken farm when enough such calls from individual grocers have come in. And the chicken farm itself only increases the number of chickens it is raising—perhaps by a thousand—when a large enough number of crates of chickens have been shipped to the various suppliers. But although it is still artificially simple, our current story should nonetheless suffice for our purposes.)

Suppose that the butcher counter were wired so that a red light begins to flash (so as to notify the butcher) when the 25th chicken has been purchased. And imagine that just before this, when 24 chickens have been purchased, the light flashes yellow. The rest of the time, the light is green. As you approached the butcher counter, if the light were green, you could buy a chicken knowing that your act wouldn't make a difference. The butcher simply isn't paying attention to how many chickens get purchased while the light is green. But if the light were yellow as you approached, you would know: buy a chicken, and the order will go out to torture and slaughter another 25 chickens. You would know that buying this chicken would make a significant difference; the results would be far better if you didn't buy the chicken.

Of course there may be hundreds of purchasers of chickens each day, and the light may flash first yellow, and then red, several times a day. Each time it flashes red, the butcher orders another crate of chickens, and each time it is flashing yellow the customer would know that his act of buying a chicken would result in harm to another 25 chickens. One out of every 25 chicken purchasers triggers this large harm; the other 24 acts of buying chickens simply make no difference at all (again, we are still dealing with our first approximation of the truth).

(More complicatedly still, the butcher might actually begin the day with several crates worth of chickens on his counter. Instead of ordering another crate each time the red light flashes, he might simply keep track of how many times the light flashes red each day, and adjust his nightly order accordingly. But again, this doesn't affect the essential point.)

Presumably, your butcher doesn't actually have this sort of public light display, so when you approach the counter you won't actually know whether, on the one hand, you are about to trigger a new order (or prompt an increased nightly order), or whether, instead, your act will simply make no difference to the number of chickens ordered. All you will know—absent any special information to the contrary—is that there is a 1 in 25 chance that your purchase will in fact be a triggering purchase, and a 24 in 25 chance that your act simply won't make a difference.

This means, of course, that it is extremely likely that your act won't actually affect the suffering of any chickens. But on the other hand, if your act does turn out to be a triggering act then you affect the fate of not one but 25 chickens. Thus, you have a 1/25th chance of affecting the suffering of 25 chickens. In terms of chicken suffering, then, the expected disutility of your act is one chicken's worth of suffering. To be sure, this expected disutility will be offset somewhat by the pleasure you will get from eating the chicken. But we began by stipulating that the pleasure

you get from eating a chicken is less than the suffering that the given chicken undergoes from being raised for slaughter. And this means, obviously, that the net expected utility of your act remains negative.

That is why the consequentialist condemns your act. Given uncertainty about the actual results of an act, consequentialism tells us to guide ourselves by the expected results. If an act has negative expected results, I should not perform it. But this is the situation I find myself in as I walk by the butcher counter. I cannot know whether my act will have bad results or not, but I do know that the net expected results of my act are bad. So I should not buy a chicken.

The story I have been telling is still artificially simple, though for the most part not in ways that would significantly affect the argument. For example, there is no particular reason to think that the number of purchases relevant for triggering an increased order is really 25. It might be 50, or 100, or 144. Indeed, it may not even be a precise number. (The butcher may simply order another crate every time that he sees that *somewhere* around another 25 chickens have been bought.) But we do know that the butcher is not inattentive to the number of chickens being purchased, and that he adjusts his order to keep up with demand. So there is some number of purchases (more or less) that triggers the ordering of another supply of chickens. And since the butcher neither wants to fall behind demand nor end up with ever larger numbers of unsold rotting chickens, we know as well that the number of chickens he orders is more or less the same as the number of purchases required before a new order is triggered.

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 (more or less) 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.

IX

There is, however, one way in which I have simplified the discussion up until now that should certainly be corrected. To see it, let us suppose that the triggering number is indeed (precisely) 25. Imagine that I am the 25th purchaser: the light is flashing yellow as I approach the counter, and as I take a chicken from the counter it begins to flash red (before settling down, after a moment, to green). Because I take the 25th chicken, the results are bad. Things would have been better had I acted differently.

But a moment's reflection makes it clear that this is true not only with regard to my own purchase of a chicken, but also with regard to *each* of the 24 people who purchased a chicken before me. Had any one of them acted differently, the results would be different right now. My purchase would not be the 25th purchase, the red light would not be flashing, and it would not be the case that another 25 chickens had been condemned to suffering.

Thus it isn't really true to say—as I have been saying up to this point—that only the 25th purchaser makes a difference, and that the acts of those who came before simply make no difference at all. That was, as I put it, only a first approximation of the truth. The real truth is that every single one of the 25 of us who buys a chicken acts in such a way that the results would have been better had he acted differently.

At least, that is the case when exactly 25 of us purchase chickens today. When the total number of purchases is exactly 25, then had any one of the 25 acted differently, the results would have been better. And the same is true, of course, if the total number of purchases is an exact multiple of 25. For example, if exactly 150 of us purchased a chicken today, then each one of us can correctly say that had he acted differently, the results would have been better (only five more crates of chickens would have been ordered instead of six).

Suppose, however, that the total number of chicken purchasers is not an exact multiple of 25. Then the situation is quite different. Suppose, for example, that a total of 67 chickens are bought on a given day. Then each and every person can say that his act made no difference. For even if he had not bought a chicken, so that a total of 66 chickens would have been purchased instead of 67, the results would still have been the same. For two crates of chickens were going to be ordered, regardless of whether the day's total sales came to 66 or 67 chickens. Thus, if I am one of a

cohort of exactly 67 chicken purchasers, my purchasing a chicken makes no difference to the amount of chicken suffering.[9]

This is true even of the person who makes, say, the 50th purchase, thereby causing the light to flash red, which results in the ordering of a second crate of chickens. For given the fact that a total of 67 purchases of chickens took place today, even this person can correctly say that the results would have been the same had he acted differently. For even had he walked by the butcher counter, the only difference this would have made was that it would have been the next person who performed the triggering act, rather than him. But given that he was one of 67 purchasers, a second crate of chickens was going to be ordered, whether or not he himself purchased a chicken. Thus the results would have been just as bad, even if he hadn't bought a chicken. In terms of the total amount of chicken suffering, his own purchase simply made no difference.

The real question, therefore, is not whether I will be the 25th (or 50th, or 75th . . .) person to buy a chicken, but rather whether I will be part of a cohort of exactly 25 (or 50, or 75 . . .) chicken purchasers. If I am part of a cohort that makes up an exact multiple of 25, then every one of us can say that had he acted differently, the results would have been better. But if I am part of a cohort that does *not* make up an exact multiple of 25, then every one of us can say that even had he acted differently, the results would have been the same.

Typically, of course, I will have no idea at all how large the cohort of chicken purchasers is going to be today. But in the absence of any special information about the situation, I must assume that a cohort that is an exact multiple of 25 is no more likely or less likely than any of the other twenty-four possibilities (that is, that my cohort will have one more than an exact multiple of 25, that it will have two more, . . . , or that it will have 24 more).

Thus I have a 1 in 25 chance of being part of a cohort where I can correctly say that had I acted differently, the results would have been better, and a 24 in 25 chance of being part of a cohort where I can

9. Talk of a "cohort" is simply a convenient way of referring to the group of people buying chickens at the given store on the given day; it is not intended to suggest any sort of coordination on the part of the chicken purchasers (though it is, of course, compatible with that).

correctly say that had I acted differently, the results would have been the same. This means, of course, that it is most likely that my act of purchasing a chicken made no difference. But I cannot actually know whether this is the case. Instead, I can only know the expected results of my act. If I am part of a cohort that makes up an exact multiple of 25, then it will be true of me that had I acted differently 25 chickens would have been spared their suffering. And I have a 1/25th chance of being a member of such a cohort. Thus the expected disutility of my act is still equivalent to one chicken's suffering. Even when we offset this by the pleasure I get from eating the chicken, the net expected utility of my purchasing a chicken remains negative. And so consequentialism still condemns my act of purchasing a chicken.

That is why it wasn't dangerously misleading to pretend, as we did initially, that only the 25th purchaser (or 50th, or 75th, and so forth) made a difference. Strictly speaking, to focus on where your purchase falls in the overall sequence of purchases is a mistake, since the morally important question actually concerns the overall *size* of your cohort, rather than your place in the butcher's line. Nonetheless, thinking about the situation in these terms wasn't a bad first approximation, for it made it easy to see how the expected utility of your act could be negative, even if most such acts make no difference whatsoever.[10]

X

Purchasing a chicken is condemned by the consequentialist on the grounds that the expected utility of such an act is negative. I may not know what the actual triggering number, T, is, but I do know that I have a 1 in T chance (more or less) of being part of a cohort that triggers an increased order, and that if I am part of such a cohort then another T chickens (more or less) will suffer. The net expected utility of my act of purchasing a chicken is negative, and so the consequentialist condemns it.

Of course, this argument assumes that I have no special information about my chances of being in a cohort of the relevant size. And we can imagine special cases in which this isn't so. Suppose, for example, that

10. Presumably there are some triggering cases where it is actually the initial analysis that is more apt. Suppose that each person objectively has a $1/n$ chance of triggering some bad result, regardless of what others do. Here the size of the cohort would be irrelevant, since only the person who *actually* triggers the bad result makes a difference.

the New Haven Friends of Chicken Consumption phone me up, proving to me that they have reliable information about the relevant triggering number at my local supermarket, and precise information about the number of chicken purchases that are going to be made today. Suppose I know for a fact that increased orders are triggered at multiples of 25, and that other than me there will be exactly 66 chickens purchased today. If I buy a chicken as well, then the total number of purchases will be precisely 67—and there is no danger that more, or fewer, people will come in than I have been told about.

In such a case I know that my act makes no difference, and results will not be any better if I refrain from buying a chicken. In such a case, therefore, the expected disutility of my purchasing a chicken is actually zero, and when we remember to throw in the pleasure I may get from eating the chicken, the net expected utility of my act will in fact be positive. Accordingly, the consequentialist cannot now condemn my act. (At least, not unless other factors are brought in, such as the effect my purchase may have in encouraging others—including the Friends of Chicken Consumption—to eat chickens. These effects are, of course, often significant, and may be decisive. But they are not our concern here, so I leave them aside.)

Some may find this implication disappointing. They want to condemn buying a chicken even when one *knows* that doing this does not in any way increase the suffering of chickens. But such a thought is not likely to be one that appeals to consequentialists, who are concerned, after all, with the *results* of our actions. If we know that an act has no bad results, there is no consequentialist objection to it, and this is precisely what we are imagining to be the case when we bring in the omniscient Friends of Chicken Consumption. In particular, then, were I to get such a call from the Friends of Chicken Consumption, consequentialists would not condemn my act, because I would find myself in one of the rare cases where the expected utility of buying a chicken is not negative.

But be that as it may, under *normal* circumstances, in any event, the phone call simply does not come, and so I have no idea at all what size my cohort will be in comparison to the triggering number. And in such cases, as we have now seen, the expected utility of buying a chicken is indeed negative. Consequentialists will, accordingly, condemn those who buy them.

XI

I have discussed the example of purchasing a chicken at considerable length, because I take it to be a fairly representative case of the situation we often find ourselves in with regard to collective action problems. Whether or not all collective action cases involve triggering in this way, many of them certainly do—perhaps most of them. But if my discussion of this sort of case is correct, then the consequentialist can handle such cases using the familiar appeal to expected utility. Admittedly, in such cases, I may not be able to know whether or not, if I act, I will be part of a cohort of the relevant size for triggering the bad results. But no matter. I can still know that the *expected* utility of my act is negative. And that will be enough to allow the consequentialist to condemn my act.

Triggering cases, therefore, do not pose a serious threat to the consequentialist. But the situation is significantly different for those (pure) cases that turn on imperceptible differences. For here there is *no* chance that my act will make a morally relevant difference—not even a small chance, as with the triggering cases.

Strictly speaking, what I have just said is an overstatement. As we have already seen, imperceptible differences won't pose a particular problem for the consequentialist so long as they constitute real harms, even if imperceptible ones. For so long as my act harms my victim (albeit imperceptibly), I will still be able to say that my act makes a difference, in that had I acted differently, my victim (or victims) would have been better off. But as we have also noted, in at least some of the cases that concern us it simply doesn't seem plausible to claim that the difference my act makes constitutes a harm (even though enough such differences do constitute a harm).

In the case of the torture machine, for example, it is very difficult to believe that the pain could be worse if it is not perceptibly worse (what is bad about pain is its qualitative perceptual aspect), and it is even harder to believe that an imperceptible difference in the amount of current running through the machine, or an imperceptible difference in the rate at which my neurons are firing, could count as a harm in and of itself. Similarly, in the visual pollution case, a single extra particle in the visual field of any given observer might make no difference at all to their perceptual experience (even though enough such particles do make a perceptible difference), so insofar as we are concerned with the harm that

pollution causes to our experience of the darkened sky, it is difficult to believe that the imperceptible difference constitutes a genuine harm. Thus, here too the appeal to imperceptible harms seems implausible.

This, therefore, is the sort of collective action case that remains a threat to consequentialism. If my act makes only an imperceptible difference, and that difference does not itself constitute a harm, then even had I acted differently the results would have been no better. In such a case, (individualistic) consequentialism seems incapable of condemning my act, even though when enough of us act in this way the results are very bad indeed. As we have seen, even consequentialists will concede that this implication of their theory is problematic.

How then should we respond to such cases of imperceptible difference? (For simplicity, let us hereafter restrict our attention to such cases of "harmless" imperceptible differences.) I believe the correct thing to do is this: we should deny their existence.

That, at any rate, is what I intend to do. I am going to claim that there simply *are* no cases of the kind that we are now concerned with. It is *never* the case that a large enough number of acts make a morally relevant difference, but each individual act makes no perceptible difference at all.

In effect, I want to concede that were there cases of this sort, they would indeed pose a serious challenge to the adequacy of consequentialism, but insist nonetheless that this fact needn't concern the consequentialist at all, for cases of this sort simply do not exist. And in saying this, I should emphasize, I am not merely making the modest (empirical) claim that as it happens cases like this never actually arise. For were that all I was prepared to argue, then it might reasonably be suggested that, whether or not there are any actual cases of imperceptible differences, the mere possibility of such cases suffices to pose the threat to consequentialism. But in fact I want to make the bolder (conceptual) claim that there could not possibly be cases of imperceptible difference. Thus, the mere possibility of such cases poses no threat at all, for they are not so much as possible.

In saying this, I concede, I seem to be making an obviously false claim. Not only do cases of imperceptible differences seem logically possible, it seems clear that such cases actually exist! So my position appears to fly in the face of plain fact. And at any rate, whether or not there are such cases certainly seems to be an empirical question, yet my argument will be comfortably a priori. So my position seems like it must be a nonstarter.

For obvious reasons, I will want to return to these concerns later. But first, let me simply try to make out the positive case for my position.

XII

Let us consider the case of the harmless torturers more closely. Each time an additional switch is thrown, the current increases very slightly. If enough switches are thrown the victim is in incredible pain, but each additional switch increases the current so minimally, that the difference this makes to the victim's body cannot be perceived by the victim himself. While his neurons may be firing at a slightly higher rate, this difference falls below the victim's perceptual threshold. He cannot tell that anything different is happening to him. He does not feel in any greater pain. He simply cannot tell the difference between the state he is in when, say, 778 switches have been thrown, and when a "mere" 777 switches have been thrown. And something similar is true, by hypothesis, no matter how many switches have been thrown. That is, no matter how many switches have been thrown the victim cannot ever tell the difference between the state he is in and the state he would be in had one less switch been thrown.

Thus, there is simply no perceptible difference between any two "adjacent" states (in the sequence from 0 switches thrown, to 1,000 switches thrown). Obviously, there is an easily perceivable difference between state 0 (no switches thrown, no pain at all) and state 1,000 (immense suffering). And in fact there are perceivable differences between closer states than this, since, for example, one can perceive the difference between, say, state 100 (mild discomfort), and state 500 (considerable discomfort). But there is no perceivable difference between any two adjacent states.

That, at least, is the claim. Obviously, what I need to argue is that we cannot actually have a case like this. And equally obviously, my objection cannot simply be that a mere 1,000 steps is too few to make the move from no pain at all to torture, traveling all the way via imperceptible steps. For if that were the problem, we could simply increase the number of switches. More particularly, then, what I need to claim is that there could not really be a series of states like this at all, where one end of the sequence differs from the other in perceptible ways, but the differences between any two adjacent states are imperceptible.

Here then is the argument. It is argument by reductio. By hypothesis, when the person is in state 0 they are in no pain. If we ask them whether they are in pain, they will answer "no." In state 1,000 they are in excruciating pain. If we ask them whether they are in pain, they will answer "yes." Suppose then we consider state 1. Since this is adjacent to state 0, the difference between state 0 and state 1 must be imperceptible. Hence, if we ask someone in state 1 whether they are in pain, they must give the same answer as they gave when the same question is posed with regard to state 0, that is, they must answer "no." (If their answer in state 1 differed from their answer in state 0 this would presumably indicate a difference in their perception of the two states, contrary to hypothesis.) Now consider state 2, which is of course adjacent to state 1. Since by hypothesis the two adjacent states are imperceptibly different, the answer to the question "are you in pain?" must be the same. But the answer to this question with regard to state 1 is "no," hence the answer with regard to state 2 must be "no" as well.

This argument obviously reiterates, and we are led to the conclusion that when the question is asked whether you are in pain with regard to state 1,000, the answer must again be "no." (Otherwise, this state can be perceptually distinguished from the state adjacent to it, state 999.)

But of course we know this to be false, for it is given that in state 1,000 the person is in tremendous pain, and can easily distinguish his situation from state 0. When we ask the person in state 1,000 whether he is in pain, by hypothesis he answers "yes." So we now have a contradiction. Thus the case as described cannot exist. QED

Once stated, the point seems obvious.[11] It simply cannot be that every state feels like the one before it, for by hypothesis state 0 feels like no pain, while state 1,000 feels like pain. Hence at least one state must feel different from the one that came before. At some point the answer given to the question "are you in pain?" must differ from the answer given immediately before—otherwise the victim would still be answering "no" at state 1,000 (just as they answered "no" at state 0), something we know to be false. Thus not every state feels like the one before it. At least one

11. Unsurprisingly, then, a similar idea has occurred to others as well. See, for example, Section IV of Alastair Norcross, "Comparing Harms: Headaches and Human Lives," *Philosophy & Public Affairs* 26 (1997): 135–67. Another example (brought to my attention after I had completed this essay) is Frank Arntzenius and David McCarthy, "Self Torture and Group Beneficence," *Erkenntnis* 47 (1997): 129–44.

state is perceptibly different from the one before it. (In fact, of course, many states will be perceptibly different. But the existence of one perceptible difference is enough to show the impossibility of the case.)

(A word of clarification. It might well be that the longer the current is running, the greater the pain, even when the current stays constant. Thus it could be that if we imagine moving, over time, from state 0 to state 1 to state 2, and so forth, two adjacent states will indeed feel different, but simply because of the greater time the pain has been endured. Strictly, then, what we need to do is to take a person in a single, constant state, and ask the counterfactual question: *were* the person in a given alternative state, how would he answer the question "are you in pain?" But having noted this point, I won't insist upon the more careful formulation of the argument.[12])

If this argument is sound, then it works quite generally. We needn't restrict ourselves to concluding that as we move from no pain to excruciating pain at least one state must perceptibly differ from its neighbor. Rather, we will be able to conclude that, no matter *how* small the difference is between the first state in such a sequence and the last state in such a sequence, as long as there is a perceptible difference between these two states, it cannot be the case that there is a series (however long or short) of intervening states, such that no state is perceptibly different from the state adjacent to it. For by hypothesis, if we ask of someone in the first such state whether they are in the first state (as opposed to the last state), they will perceive that they are—for it is a perceptible difference—and will answer "yes." And if we ask of someone in the last such state whether they are in the first state (as opposed to the last state), they will perceive that they are not—for it is a perceptible difference—and they will answer "no." And from this we can conclude that for at least one state between the first state and the last state (and perhaps more than one state) there is a perceptible difference between that state and the one before it. For if each state were imperceptibly different from the state before it, each time the question is asked whether the person is in the first state (as opposed to the last state) the answer would have to be "yes"—and this would have to remain true, therefore, even for the *last* state, contrary to hypothesis.

12. I owe this point to Caspar Hare.

In short, for any such sequence, for any two nonadjacent states that are perceptibly different, there must be at least one state that is perceptibly different from its neighbor. And this means, quite generally, that there simply cannot be cases of imperceptible differences of the sort we were looking for. The problematic case simply cannot exist.[13]

XIII

A few points of clarification about this argument may be in order. First of all, in claiming that at some point in the sequence from state 0 to state 1,000 the answer to the question "are you in pain?" must change, I certainly do not mean to suggest that the answer must change as soon as we move from state 0 to state 1! It could certainly be the case that there *are* adjacent states that are genuinely imperceptibly different from one another. My claim is simply that there cannot be a sequence of such states, *all* of which are imperceptibly different from the ones adjacent, where nonetheless there is a perceptible difference between the first state in the sequence and the last state in the sequence. With the torture machine, the answer to "are you in pain?" obviously need not stop being "no" as soon as we leave state 0. But it will have to stop being "no" at some point.

Second, in claiming that at least one state must be perceptibly different from the state before it, I also do not mean to suggest that at the first such state (that is, the first state that is perceptibly different from the one before it in the sequence) the answer suddenly switches from an unqualified "no pain" (at the previous state) to an unqualified "pain" (at the new state). There is simply no reason to expect the perceptible difference to be anything that large. If I am right, there must be a perceptible difference *somewhere* in the sequence, but it need not elicit anything like that kind of huge shift in the way the two adjacent states are characterized.

13. If my argument is sound, does it also provide a response to *sorites* style arguments? Not at all. The argument does of course show that at some point in a sorites sequence, the move from one step to the next must involve a *perceptible* difference (since the end points in the sequence are perceptibly different). But this doesn't threaten sorites arguments, since such arguments merely insist that adding a grain of sand, say, doesn't change a nonheap into a heap; they needn't claim—and if I am right, they shouldn't claim—that such additions can't even make a perceptible *difference*.

What we might expect, instead, is a slow hedging of the characterizations. Initially, perhaps, the person will confidently report that he is in no pain at all. And at some point he will say that he is in no pain, but he will report this with less confidence, and at still further points he will report uncertainty as to whether he is or is not in pain, or he might report that he is perhaps in some extremely mild discomfort (though nothing that could be called pain), and so forth and so on. (Similarly, it might be that at a given state hesitant reports of mild discomfort are mixed with denials of being in any pain at all; but the denials grow less frequent, and the reports of discomfort more frequent, at later states.) The difference between the ways that the two adjacent states are characterized will presumably be quite slight, befitting the fact that the difference in the level of pain, although perceptible, is extremely slight. But this much seems clear: there will be a first state, for example, at which the answer is no longer the confident, unhesitant reply "no pain" of state 0, a first state where some hedging, qualifying, or hesitation is introduced. Thus this state will perceptibly differ from the state before it.

XIV

One natural objection to the argument I have just given is this. From the fact that inevitably there must be a state where the answer (to the question "are you in pain?") differs from the answer given in the state immediately before, it certainly follows that these states must themselves be different in various ways. But this doesn't really show that the differences are themselves *perceivable.* All I have really shown is that we can have good reason—here, based on the torture victim's reports—for *inferring* the existence of various differences in at least one pair of neighboring states. But no one ever denied the fact that the neighboring states differ in various ways, nor was it ever denied that we might well be able to make reasonable inferences concerning the nature and existence of these differences. The claim, rather, was that any differences between adjacent states were all *imperceptible* ones. And nothing I have said gives us any reason to think that the differences are indeed perceivable ones after all. So I haven't really demonstrated the impossibility of imperceptible difference cases.

I believe, however, that it is a mistake to claim that all that I have shown is one more way of inferring the existence of differences

between certain adjacent states. On the contrary, I think it appropriate to say that the differences in question are indeed perceivable. More precisely, the differences in the amount of *pain* are perceivable (other differences, of course, may be only inferable). After all, it cannot be denied—given the earlier argument—that the victim's *reports* must sometimes differ from one another (in neighboring cases), and so it is important to bear in mind that these 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. Given that there is a difference in the victim's spontaneous reports concerning how much pain he is in, I take it there is a perceptible difference in the amount of pain. Accordingly, I think that I have indeed demonstrated the impossibility of imperceptible difference cases.

Of course, it must be conceded that this difference in pain might not be noticed if the victim limits himself to direct pairwise comparisons between the two adjacent states. This is especially true if the victim limits himself to a single such direct comparison (or to a small number of such comparisons). After all, even if there is a difference between some pair of neighboring states with regard to how much pain they involve, this difference will typically be extremely small, and might easily be overlooked. There is no particular reason to assume that any single direct comparison (or a small sample of such) must enable the victim to immediately detect that difference. Perceivable differences can be genuine even if we are not infallible at detecting them, especially when we are limited to the evidence of a few, brief, pairwise comparisons.

What's more, even if it should turn out that the difference in pain cannot be detected at all when we limit ourselves to direct pairwise comparisons (no matter how many, or how sustained), this would still not justify our concluding that the differences here are not perceivable ones. For it might well be that perceivable differences between two neighboring states are only noticed when we compare each of the two states to some third, baseline state. Thus it might be that when *directly* comparing states 66 and 67, for example, the victim does not detect any difference in how they feel, but he does detect such a difference when he compares each of these two to some earlier state (such as state 0):

comparing the two neighboring states to the earlier state may enable him to detect the difference that he otherwise overlooks. (State 66 may feel like state 0, while state 67 does not.) If so, this too constitutes a perceivable difference in the two adjacent states, albeit one that would not be noticed if we were restricted to direct comparisons between the two. (Analogously, there may be a perceivable difference between two complex geometrical figures, but we might not notice it until we visually compare the two to some third figure.)

In effect, I have argued that there must be at least one pair of neighboring states where there is a perceivable difference in the amount of pain, whether or not it is a "directly" perceivable difference (a difference that can be detected even if we limit ourselves to direct pairwise comparisons of the neighboring states). But I recognize that some may remain uncomfortable saying that the relevant differences here are perceivable ones if they are not in fact *directly* perceptible. Perhaps, however, some who feel this unease will be prepared to accept an alternative description of the situation: since the victim's immediate and spontaneous reports differ for at least one pair of adjacent states, it must be true that there is a difference in how the victim *feels* in those two states, regardless of whether these differences are "perceivable." (That is, what it feels like to be in the first of the neighboring states differs from what it feels like to be in the second of the two states.) Thus, even if there are no perceptible differences here, there are differences in how the victim *feels*.[14]

Of course, if we do adopt this alternative way of describing the situation, then strictly speaking it won't be true that I have shown the impossibility of imperceptible difference cases. (The torture case will be such a case, and doubtless there will be others.) But this concession will no longer be worrisome. With hindsight, we can say that what was genuinely troubling about the appeal to imperceptible harms in cases like the one we are focusing on—cases where the bad outcome is bad by virtue of involving pain—was the suggestion that an act might leave someone worse off, even though it makes no difference to how they feel. But even if we adopt the alternative description of the situation, the genuinely troubling suggestion is still avoided: imperceptible

14. I owe this alternative description to Caspar Hare.

difference cases will be possible; but for all that, such cases will still involve differences in how people *feel*.

As it happens, I don't myself think that this alternative way of describing the situation is preferable. I think it perfectly appropriate to say that the differences in question are indeed perceivable ones (even if it should turn out that they are not "directly" perceivable). But so long as it is conceded that the differences are indeed differences in what it feels like (that is, differences in the amount of pain that is felt), then that seems sufficient for my purposes, regardless of whether we go on to say that the differences are "perceptible."

I conclude, therefore, that the sort of case that would be genuinely troubling for the consequentialist—a case where enough acts would make a significant difference in the amount of pain that the victim feels, but no individual act makes any difference at all—that sort of case simply cannot exist.

Something similar is true, of course, for other qualitative aspects of experience as well. For example, suppose we try to understand the case where pollution darkens the sky as an imperceptible difference case—holding that while each act of releasing pollutants makes no perceptible difference to anyone's visual experience, enough such acts together significantly impair one's visual experience of the sky. Reasoning similar to what we have offered for the torture machine would show here too that this example cannot actually have the structure we would then be taking it to have. After all, as the number of pollutants increases, sooner or later our reports of our visual experiences would have to vary (since, by hypothesis, when enough pollutants are released, we report gray skies, while we report blue skies when none are released). So it cannot be true, I think, of each such act of releasing pollutants that it makes no difference at all to our visual experiences: on the contrary, for at least some such acts, releasing the extra pollutants makes a *perceptible* difference. And even if there are some who prefer to say that the differences remain imperceptible—since they may not be "directly" perceivable—it will still be true that some such acts will make a difference to how the sky *looks*.

In short, there cannot be cases where it is true of each act that it makes no difference at all to the relevant qualitative aspect of experience, yet enough such acts together do make a difference. Such cases would be troubling for the consequentialist. But they simply cannot exist.

XV

Let me return then to the worry (voiced at the end of Section XI) that I have joined the distinguished, but lamentable, philosophical tradition of offering a priori arguments on matters that are undeniably empirical. Surely it is an empirical question whether there can be cases of the sort that I have been discussing. How, then, can I attempt to disprove their very possibility on the basis of this sort of merely conceptual argument?

My answer, unsurprisingly, is that it is *not* in fact an empirical question whether there can be cases of the kind we have been discussing. On the contrary, I have shown that such cases simply cannot exist. It does not take empirical science to show that if we react to a first state in a series one way, and to a later state in that series another way, at some point in moving from one state to the next in that series our reactions must change. That is a purely conceptual observation, not an empirical one. And it is not an empirical remark to point out that if our reactions are immediate and spontaneous reports of our own experience (and how it compares to other experiences), then differences in the content of those reports indicate differences in the experiences themselves. That too is a conceptual observation (here, about the concept of an observation report). But these observations suffice to rule out the very possibility of the kinds of cases that we have been considering. So the relevant kind of case is impossible, as a matter of *conceptual* necessity.

What *is* an empirical question, of course, is why we might take there to be such cases. And a large part of the answer to this, presumably, will turn on the question—also empirical—as to why we might so readily overlook the differences between the relevant neighboring states, why the differences often cannot be detected through simple, direct, pairwise comparisons.

I have already indicated some of what I take to be the answer to that last question as well. Even if there are differences between a pair of adjacent states, the differences—by hypothesis—may be quite small, indeed *extremely* small. If one isn't paying very careful attention—and paying attention for a sufficiently long time—one simply might not notice the differences at all. For example, it might be that the difference in pain for a pair of neighboring states is slight enough, so that it only manifests itself as a very small increase in the frequency with which the

victim reports that he is in some mild discomfort. Perhaps instead of saying that he is in such discomfort ten times out of a hundred, he now says this eleven times out of a hundred. A difference like this will be subtle and easy to overlook.

Of course, suggestions like these, plausible though they may be, are only placeholders for the real account. It is indeed a task for empirical psychology to provide an adequate account of why we so readily overlook differences between adjacent states that nonetheless differ in terms of how they feel. But whatever the best explanation here may be, the fundamental point remains: there simply cannot be any cases at all where each individual act makes no difference to the qualitative aspects of experience, yet enough such acts taken together do make a difference. Such cases simply cannot exist.

XVI

I began the critical examination of collective action problems (in Section VI) by suggesting that there seemed to be two kinds of pure cases—triggering cases, on the one hand, and imperceptible difference cases, on the other. (Other cases, I conceded, might well be mixtures of the two pure types.) However, I have now denied the very possibility of imperceptible difference cases (at least in the troubling form, where the bad outcome is to be understood in terms of a qualitative aspect of experience, yet the individual acts make no difference to what is felt). The conclusion, of course, is that there is only one sort of collective action problem after all: triggering cases.

This seems to me, in fact, the correct conclusion. Collective action cases are all triggering cases, though they differ, of course, in their details. (For example, they may differ with regard to whether there are many triggering acts, each making only a very small difference to the overall outcome, or only a few triggering acts, each making a large difference.)

But triggering cases, I have claimed, are readily handled by familiar consequentialist machinery. In the kind of triggering cases that concerned us, although it may not be very *likely* that my act will make a difference, there is a great enough chance that it will, so that—given the difference that it *might* make—my act will have negative expected utility. That is why the consequentialist condemns it.

I conclude, therefore, that there is no need to go beyond familiar consequentialist theory to adequately deal with collective action problems of the sort that have concerned us here.[15] We are already in a position to answer the question I posed in the title of this essay.

Do I make a difference? I *might*.

15. Nonetheless, two qualifications should be kept in mind. First, as noted earlier (in Section IV), there are other, rather different problems that might also be legitimately called "collective action problems"—and I obviously haven't tried to address those here. Second, despite its familiarity, the consequentialist's appeal to expected utility might itself be questioned (see note 8). But that too is a topic for a different occasion.