

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

The Morality of Collective Harm

Permalink

<https://escholarship.org/uc/item/9s49g22t>

Author

Nefsky, Julia

Publication Date

2012

Peer reviewed|Thesis/dissertation

The Morality of Collective Harm

By

Julia Nefsky

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Philosophy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Niko Kolodny, Co-chair
Professor R. Jay Wallace, Co-chair
Professor Christopher Kutz

Fall 2012

The Morality of Collective Harm

Copyright 2012
by
Julia Nefsky

Abstract

The Morality of Collective Harm

by

Julia Nefsky

Doctor of Philosophy in Philosophy

University of California, Berkeley

Professor Niko Kolodny, Co-chair

Professor R. Jay Wallace, Co-chair

Collective Harm Cases pervade the modern world. In these cases, people collectively cause harm, or fail to prevent it, but no individual seems to make a difference. For example, while climate change is caused by millions of people acting in certain ways (driving in cars, flying in planes, using coal-generated electricity, etc.), it seems that no *individual* such act will make a difference. Similarly, while the large-scale consumer patterns of wealthy nations can affect the lives of people in poorer nations, it is hard to believe that any individual purchase will make a difference for the worse. In cases of this sort, if it is true that no individual act will make a difference, it is unclear how we can say that anyone ought, or even has *reason*, to act otherwise. But if we cannot, this would be very troubling. Morality would be powerless in a wide array of cases in which it should have force. This is the Problem of Collective Harm.

I argue that the key to solving the Problem of Collective Harm is to reject a standard, intuitive assumption that underlies it. The assumption is that if an act cannot *make a difference* with respect to an outcome, then it cannot play a morally significant, non-superfluous causal role in bringing that outcome about. Or, to put it another way, we must reject the assumption that *helping* to bring about an outcome requires making a difference to it. This assumption is highly natural, but I argue that it is mistaken and that once we see that it is mistaken, the central puzzle of the Problem of Collective Harm is resolved.

Chapter 1 is devoted to setting out the problem and how I will approach it. Chapters 2 and 3 discuss a range of proposed solutions that have been put forth by others. The upshot of the critique of these views will be that we cannot solve the Problem of Collective Harm *unless* we can reject the assumption that helping requires making a difference. In Chapter 4, I turn to my positive

proposal. There I give an account of what is required for an act to help to bring about an outcome, where this does not require being able to make a difference. I use this account to explain our reasons for action in Collective Harm Cases, and I give an error theory that explains why the standard assumption seems right, even though it is not. In Chapter 5, I highlight some noteworthy features of my proposal and some important questions about it that still need to be addressed.

For Alex

Contents

Acknowledgments	iv
Chapter 1: Introduction	1
1.1 The Problem of Collective Harm	1
1.2 What I Will Argue	3
1.3 Views that Bite the Bullet	6
1.3.1 The Group Did It	6
1.3.2 The Government versus the Individual	8
1.4 Methodology	10
Chapter 2: Rejecting the Inference	14
2.1 Introduction	14
2.2 Membership	14
2.3 Participation	16
2.4 Fairness	21
2.5 The Common Problem	24
2.6 Can't We Just Apply a Non-Consequentialist Theory and Call it a Day?..	29
2.7 The Common Problem and the Way Forward	31
2.8 The Participation Approach Revisited	33
Chapter 3: Denying the Description	37
3.1 Introduction	37
3.2 Imperceptible Difference Cases versus Triggering Cases	39
3.3 Triggering Cases: the Expected Utility Solution	41
3.4 The Appeal to Imperceptible Harms and Benefits	44
3.5 Non-Phenomenal Imperceptible Difference Cases	46
3.6 Revising the Distinction: Triggering versus Non-Triggering Cases	48
3.7 Kagan's Harmless Torturers Argument and the Reports Premise	51
3.8 Perceptible Difference versus Worsening of Pain	54
3.9 The Role of the Reports Premise	56
3.10 Reports and Vague Boundaries	58
3.11 Quasi-Triggering Cases	60
3.12 Summary	65
3.13 Denying the Description: the State of the Project	65
Chapter 4: Helping Without Making a Difference	68

4.1 Introduction	68
4.2 The Standard Assumption	68
4.3 Part of the Cause versus Helping	71
4.4 Helping Does Not Require Making a Difference	74
4.5 Conditions for Helping	77
4.6 Reasons to Refrain	81
4.7 Triggering versus Non-Triggering Cases	83
4.8 Why it <i>Seems</i> As Though Helping Requires Making a Difference	85
4.9 Notions of Possibility	88
 Chapter 5: Concluding Remarks	 91
5.1 Separability	91
5.2 Other Reasons	91
5.3 Extendability	92
5.4 Individual Over Time Cases	94
5.5 Strength of the Reason	96
5.6 Moral Obligation	97
 Bibliography	 100

Acknowledgements

I am extremely grateful to my advisors, Niko Kolodny and R. Jay Wallace. Their insightful and challenging comments on drafts of my work were invaluable. So many conversations with them, both in person and through written correspondence, were crucial to the progression and refinement of my thinking. I owe them tremendous thanks for their incredible guidance, support and encouragement.

I would also like to give special thanks to John MacFarlane for much help throughout my time in graduate school, for reading and commenting on a draft of Chapter 3, and for several helpful conversations about the project.

Thank you to Christopher Kutz for serving as the external member of the committee, for several excellent discussions on the topic during the qualifying exam phase of my studies, and for his engaging, richly developed work in the area, from which I learned a great deal.

For helpful discussions about various aspects of the topic and, in some cases, for reading and commenting on earlier chapter drafts, I would also like to thank: Brian Berkey, Markus Kohl, Diana Raffman, Josh Sheptow, Sergio Tenenbaum, Seth Yalcin, all the participants of the Fall 2008 and Fall 2010 Dissertation Seminars, and audiences at Berkeley, University of Waterloo, and University of Toronto. Chapter 3 is a revised version of my paper, “Consequentialism and the Problem of Collective Harm: A Reply to Kagan”, which was published in *Philosophy & Public Affairs*; I would like to thank two anonymous referees from the journal for their comments. I am also grateful for the support of the Social Sciences and Humanities Research Council of Canada.

I am immensely grateful to my parents, Phyllis and Gil, and to my brother, Colman, for their encouragement and unwavering support.

Finally, I am deeply indebted to my husband, Alex. He has read through more or less every chapter draft and indulged me in countless conversations on the topic. He has been a constant source of support and happiness, and always knows just when to pull me away from my desk for a walk in the hills.

Chapter 1

Introduction

1.1 The Problem of Collective Harm

Let's begin with an example from Derek Parfit's *Reasons and Persons*, which he calls "Drops of Water". Imagine that there are ten thousand men lying out in the desert, suffering from intensely painful thirst. We are a group of ten thousand people near the desert, and each of us has a pint of water. If we pour our pints into a water cart, it will be driven out into the desert and the water will be evenly distributed amongst the men. Each would then be able to drink a full pint of water, and this would alleviate his suffering.¹

On the one hand, it seems clear that what we ought to do is pour our pints into the cart, so that the men's suffering can be relieved. The problem is, though, that – while together these acts would do a lot of good – it does not seem that any *individual* such act will make a difference. If you pour your pint into the cart, this will only enable each man to drink an extra ten thousandth of a pint of water. This is no more than a single drop, and a single drop more or less is too miniscule an amount to make any difference to how they feel. But if this is right, it is unclear why any of us has reason to add our pints. Yes, these men are suffering, but if it will not make a difference whether or not I add my pint – if things will not be any better for them if I add it than if I don't – then what reason do I have to do so?

This is just one example from a wide variety of cases that pose a problem for moral thought: cases in which people can *collectively* cause some morally significant outcome – such as a harmful or beneficial outcome – but no *individual* act seems to make a difference. The problem in these cases is that it seems that each person can argue, "it makes no difference whether or not I do *X*, so I have no reason to do it." People give this sort of argument frequently in everyday moral deliberation, and while we might sometimes try to dismiss it as a poor excuse, it is hard to say where it goes wrong. So, in these cases, even if through our voluntary actions, we end up collectively causing or

¹ Slightly adjusted from Parfit, *Reasons and Persons*, p. 76.

failing to prevent serious harm, it's unclear how we can say that anyone ought – or even, had *reason* – to have acted otherwise. I will call this “the Problem of Collective Harm,” and the cases “Collective Harm Cases.”

Here are some real-world examples.

Global warming is caused in part by millions of people acting in certain ways – driving in cars, traveling in planes, using coal-generated electricity... But, for any individual such act, it doesn't seem to make a difference. People might say, ‘global warming poses a serious threat; so we all ought to conserve electricity, reduce the amount we drive, and so on.’ But each individual can reply: ‘sure, I'm worried about global warming, but will it really make any difference at all to the problem if, for example, I take the bus today instead of driving? And if it's not going to make any difference – if things will be just as bad either way – then what reason do I have to do so?’

Another example – a classic – is that of voting in a large national election. It is overwhelmingly likely that no single vote will make a difference: the winner will be the same, give or take any individual vote. So, what reason do I have to vote?

Here's another example. There is currently a food crisis in West Africa. While a large number of modest donations to Oxfam's famine relief fund could significantly change what Oxfam is able to do to help, what difference does a single such donation – say \$50 – make? It is tempting to think that what Oxfam does with this money is that it buys \$50 worth of food and gives it out to people. But, as Garrett Cullity writes,

On reflection, this looks wrong. Relief agencies do not buy food in such small amounts. Instead, they make a large-scale calculation of the size of the overall need, the amount of money they are likely to be able to raise from various sources to pay for it, and the extent to which it makes sense to draw on contingency funds... Guided by these large-scale considerations, they then arrange the shipment of a large quantity of food.²

If this is correct, then it seems that a donation of \$50 isn't going to make a difference to what Oxfam actually does. They are making decisions in a way that is just not sensitive to a difference of \$50.

There are many more examples. Think of times you might have heard someone reason along these lines. ‘It won't make a difference whether the cup of coffee I buy this morning is fair trade or not. So, why should I choose a blend that is not my favourite?’ Or, ‘even if exposure to a substantial amount

² Cullity, “Pooled Beneficence,” p. 2.

of pesticide would be bad for someone's health, will using pesticide on *my garden* really make any difference to anyone's health? It sure will make a difference to my slug problem!

While cases of this sort have always existed, in the modern world they have become pervasive. Perhaps in the past Collective Harm Cases could have been ignored as peripheral, or odd exceptions. Now, though, a large number of our day-to-day decisions fall within the ambit of one such case or another, and a substantial portion of the major problems in the world – environmental, social, economic – have at least in part the structure of a Collective Harm Case.³ If we cannot explain why individuals have reason for action in these cases, this would be highly troubling from both a theoretical and a practical perspective.

I think we underestimate the force of the “it makes no difference” argument and the extent to which it has an impact on our choices. Sometimes we explicitly appeal to this argument in conscious deliberation and conversation, but at other times our choices are – no doubt – implicitly affected by the impression that acting in the relevant way won't make a difference. People often lament the unwillingness of others to change their way of life and make even small sacrifices in the face of the world's most serious problems. How can people be so stuck in their ways, selfish or short-sighted that they will not change their lifestyles, even in small, simple ways, in the face of climate change, or global poverty? But, as tempting as it might be to focus our blame on the ills of human nature, it seems to me that if we want to understand and address the motivational problem here, a central component is recognizing the force of the ‘it makes no difference’ line of thought and trying to seriously determine where it goes wrong. If it seems that the problem of climate change will be *exactly the same* whether or not I reduce the amount I drive, then it's no wonder if I am not motivated to make this lifestyle change. Stubbornness, selfishness, shortsightedness, may have little to do with it.

1.2 What I Will Argue

Proposed solutions to the Problem of Collective Harm can be divided into two categories. In the first category there are views that accept that an individual act won't make a difference in at least some Collective Harm Cases but argue that there are other sorts of reasons for action, reasons that don't have to do with whether or not you can make a difference. These views, in other words, *reject the inference* from ‘it won't make a difference’ to ‘there's no

³ For an excellent discussion of the nature and significance of this shift see Samuel Scheffler, “Individual Responsibility in a Global Age.”

reason to do it'. In the second category there are views that, instead, try to deny the description of the cases as ones in which one won't make a difference. People who take such an approach argue that even though it *appears* otherwise, individual acts do, or at least might, make a morally significant difference in Collective Harm Cases and that this is what gives us reason for action. So, views in the first category – 'rejecting-the-inference' views – approach the problem by trying to explain how, even if one can't make a difference by doing so, there is still reason for action, while views in the second category – 'denying-the-description' views – approach the problem by resisting the claim that your act isn't going to make a difference.

In Chapter 2, I will discuss several currently available rejecting-the-inference proposals. What we will find is that, while these proposals do identify considerations *other than* the difference you make, it's not clear that those considerations are reasons for action *unless* your individual act will make a difference. In other words, it's not clear that they pick out considerations that do not *depend* on being able to make a difference. More specifically, I argue that as long as your act would be causally *superfluous* – or, as long as it won't itself play a significant role in bringing about the outcome in question – then the explanations these views are offering as to why you have reason to do it can't actually get off the ground. The Problem of Collective Harm arises because, even if we know that unless enough people act in the relevant way, bad consequences will result, it doesn't seem it will make any difference whether or not you act in this way yourself. So, acting in this way seems to be *merely superfluous* – it doesn't seem it would do anything useful with respect to the outcome in question. Currently available rejecting-the-inference proposals respond by saying 'that's okay, there are other sorts of reasons for action.' But what I argue is that these proposals cannot actually make sense as long as it's true that your individual act is causally superfluous.

The moral of Chapter 2 will be that we cannot adequately solve the Problem of Collective Harm unless we can show that an individual act of the relevant sort can play a *significant*, or *non-superfluous*, role in bringing about the outcome in question. This suggests that we should look for a solution in the *second* category. It suggests that if we want to solve the problem, we need to deny the claim that an individual act won't make a difference in Collective Harm Cases. In Chapter 3, I turn to such proposals. Now, I don't doubt that in some particular cases, this will be the right way to go: we are surely sometimes making a mistake when we think that an individual act won't make a difference. But the question in Chapter 3 is whether we can *in general* say that this impression is mistaken. Can we deny that there are Collective Harm Cases in which individual acts of the relevant sort really won't make a difference? I argue in Chapter 3 that we are not in a position to do so. We cannot say – as

denying-the-description approaches attempt to – that individual acts will, or might, make a morally significant difference in all Collective Harm Cases. Thus, we do not have, and cannot expect to have, a general solution to the problem of the denying-the-description sort.

By the end of Chapter 3, then, we will have concluded that (a) we cannot solve the Problem of Collective Harm unless we can show that individual acts of the relevant sort could play a *significant, non-superfluous role* in bringing about the outcome in question, and yet (b) we are not in a position to deny that there are cases in which each act *won't make a difference*. It might seem that we have hit a standstill at this point. But my proposal is that, far from a standstill, what we have hit is the space in which a solution to the problem can be found. I turn to my proposal in Chapter 4.

My proposal begins from the observation that underlying the Problem of Collective Harm is a standard, intuitive assumption. The assumption is that if an act cannot *make a difference* with respect to an outcome, then it cannot play a morally or prudentially significant causal role in bringing that outcome about. Or, to put it another way, *helping* to bring about an outcome requires *making a difference* to it. This assumption is highly intuitive and is quite ingrained in the way we think and talk about our actions. If an act won't make a difference, this means that things will be *the same* whether or not it is performed, and we take this to mean that it would be causally superfluous – that it won't itself do anything significant with respect to bringing about the outcome in question. While this assumption has much intuitive appeal, my proposal is that it is mistaken and that the central task of solving the Problem of Collective Harm is to explain how it is mistaken. The conclusions reached in Chapters 2 and 3 – (a) and (b) above – entail that we cannot solve the problem unless we can reject the assumption. We cannot solve the problem unless we can show that an individual act can play a significant causal role even if it can't make a difference. But furthermore, if we can reject the assumption, this – I propose – would solve the problem. If by acting in the relevant way, you could help to bring about the good outcome, or help prevent the bad one, this would give you reason to do so, and thus we would have an answer to the Problem of Collective Harm.

I develop this proposal in Chapter 4 by giving an account of what it is to play a significant causal role in bringing about an outcome (that is, to *help*) where this does not require being able to make a difference, and by using this account to explain our reasons for action in Collective Harm Cases. I also argue that, with this account, we can see why the standard assumption – that helping requires making a difference – *seemed* to be true, even though it isn't. The basic idea of this error theory is that in typical Non-Collective Harm Cases, if an act cannot make a difference this is *because* the conditions for

helping do not hold. In Collective Harm Cases, by contrast, it is often that your act doesn't make a difference because of other factors and thus the conditions for helping can still obtain.

1.3 Views that Bite the Bullet

When I say that proposed solutions either reject the inference from 'it won't make a difference' to 'there's no reason to do it' *or* deny the claim that one won't make a difference, I am not counting as "solutions" views that bite the bullet. Some philosophers, though, *have* essentially bitten the bullet; that is, there are some who have essentially agreed that there are Collective Harm Cases in which it doesn't make a difference whether or not one acts in the relevant way and agreed that, when this is so, there is no moral reason to do so. In this section I want to consider two such views. Each attempts to explain why biting the bullet is the right way to go and isn't as problematic as it looks, and I want to briefly say why I don't think they are correct.

1.3.1 *The Group Did It*

In "Group Morality", Frank Jackson argues that the key to making sense of Collective Harm Cases is to distinguish between *group actions* and *individual actions* and to see that a group can act wrongly without any individual acting wrongly.

Jackson thinks that we should accept what he calls "the Difference Principle": that the "the morality of an action depends on the difference it makes – it depends ... on the relationship between what would be the case were the act performed and what would be the case were the act not performed."⁴ Jackson also does not deny that there are cases in which each individual act won't make a morally significant difference. So, according to Jackson, in at least some Collective Harm Cases, no individual has moral reason for acting in the relevant way (or, depending on the case, refraining from acting). But, his claim is that this conclusion – that is, biting the bullet – is not actually problematic, because we can still make sense of the intuition that "something wrong is *done*."⁵ We can make sense of this, he says, by "enlarging our conception of what kinds of actions can be wrong (and right)"⁶ to include *group actions*.

⁴ Jackson, "Group Morality," p. 94.

⁵ *Ibid.*, p. 100.

⁶ *Ibid.*, p. 101.

To illustrate, let's introduce a new Collective Harm Case – a case which I will use frequently in the coming chapters. This is a version of another case from Parfit, called the “Harmless Torturers”.⁷ Imagine that there is an electric shock machine, with settings s_0 to s_{1000} , hooked up to a victim and that there are a thousand torturers, each of whom turns the dial up a single setting. At s_0 there is no electric current at all. At s_1 there is a non-zero electric current, but it is too small to be perceived. In general, the settings increase the voltage by increments so tiny that the difference between any two adjacent settings is too small to be perceived. But after many increases the victim feels pain, and by s_{1000} , the victim is in excruciating pain. Since the difference in voltage between any two adjacent settings is too tiny to make a difference to how the victim feels, it seems that no torturer makes a difference for the worse. Take any individual away, and the victim would be in the exact same amount of pain. So, the Problem of Collective Harm arises: while serious harm results from enough people turning the dial, each torturer can say ‘had I acted differently, things would be *just as bad*, and so I did not do anything wrong.’

According to Jackson, this reasoning is correct. It is *true* that no torturer ought to act otherwise: none is making a morally significant difference and so none has done anything wrong. But Jackson suggests that this should not seem so problematic, because we can still say that the *group action* is wrong. For Jackson, a group action is “any old aggregation or mereological sum of individual actions.” (E.g. “My last eye-blink together with Nero’s burning of Rome is a group action; a highly heterogeneous one of no particular interest to anyone, but a group action nevertheless.”⁸) And his claim is that “the moral standing of a group action can be partially or totally at variance with the standings of its constituents.”⁹ In the Harmless Torturers case, the group action consisting of the sum of all the torturers’ dial-turns *does* make a difference for the worse. The suffering would not have happened, had these acts of turning the dial *collectively* not occurred. So, even though no individual act of dial-turning is wrong – even though no individual ought to have acted otherwise – the *group action* is wrong.¹⁰

It’s not clear, though, why Jackson thinks that this point about group actions helps. What does the claim that the *group did wrong* amount to? If it does

⁷ Parfit, *Reasons and Persons*, p. 80. This varies from Parfit’s version just in that there is only one victim rather than a thousand victims. This variation follows Shelly Kagan’s presentation of the example. I’m using this version because I will be discussing Kagan’s arguments extensively in Chapter 3.

⁸ Jackson, “Group Morality”, p. 93.

⁹ *Ibid.*, p. 101.

¹⁰ For someone else who takes a similar position see Tannsjo, “The Morality of Collective Action.”

not say that any individual ought, or even had reason, to have acted otherwise, then it doesn't seem to be a normative claim at all. If it does not have any *normative implications* for the individual agents involved, then saying that the group has done wrong doesn't seem to amount to anything more than a descriptive claim. It seems to be no different, really, than simply saying *something bad happened* (or, something bad happened as a result of human action).¹¹ But, if this is right, it doesn't add anything at all to our understanding of the morality of the situation.

In any case, even if we suppose that saying that the group did wrong – where this has no normative implications for individuals – does something toward appeasing our intuition that “something wrong is done”, this doesn't help with what is most puzzling and unsettling about Collective Harm Cases. The puzzle is that, if serious harm results from enough people voluntarily acting in a certain way, we should be able to explain why the individuals involved have at least some moral reason *not* to act in this way. Yet, as far as Jackson's proposal is concerned, we cannot explain this. If Jackson were right, then since it is *individuals* who can be guided by morality and not mereological sums of individuals, morality would be powerless in a wide array of cases in which it should have force.¹²

1.3.2 *The Government versus the Individual*

The second ‘biting the bullet’ approach that I want to consider is taken by Walter Sinnott-Armstrong in his article “It's Not *My* Fault: Global Warming and Individual Moral Obligations.” Sinnott-Armstrong is focused only on the global warming case, and the question he asks is: given the facts about global warming, is there a moral obligation not to take a leisurely Sunday drive in a gas-guzzling SUV? His conclusion is that there is no such obligation. He comes to this view by searching for a moral principle that could generate the obligation. Not finding any that work, he concludes that one is not doing anything wrong when one does things like taking a joy-ride in a gas-guzzling SUV. But he goes on to say that this does not mean that no one is doing anything wrong when it comes to the problem of global warming. It is, he says, *governments* – rather than individuals qua individuals – who have an

¹¹ Jackson says that it would be a mistake for a case like Harmless Torturers to be viewed as morally on par with a case of “natural misfortune, like a flood.” (Jackson, “Group Morality”, p. 100.) But it is hard to see how his view really avoids this.

¹² Note that a different proposal from Jackson's would be to say that the group did wrong and that this explains why the individuals in it ought to have acted otherwise. This alternative proposal does not bite the bullet, and we will discuss it in Chapter 2.

obligation to take action. Governments *can* make a difference for the better, and they ought to do so. He writes:

Global warming is such a large problem that it is not individuals who cause it or who need to fix it. Instead, governments need to fix it, and quickly. Finding and implementing a real solution is the task of governments. Environmentalists should focus their efforts on those who are not doing their job rather than on those who take Sunday afternoon drives just for fun.¹³

And:

It is better to enjoy your Sunday drive while working to change the law so as to make it illegal for you to enjoy your Sunday drive.¹⁴

I agree that in the case of global warming, and in many – though, not all – real-world Collective Harm Cases, governments, and other institutions, *can* make a difference through changes in policy, projects and funding, and *ought* to take action in these sorts of ways. I also agree that we ought to push them to do so.¹⁵ This is an important point to keep in mind. Many real-world Collective Harm Cases are only *partially* Collective Harm Cases in the sense that there are some individuals (e.g. individuals working in powerful positions in government) who clearly might be able to make a morally significant difference. What this means is that the Problem of Collective Harm is far from the only important question to ask about the morality of Collective Harm Cases. Thinking about what should happen at the institutional level is a crucial part of making sense of the morality of many – though certainly not all – Collective Harm Cases, and the answers to questions about what institutions should do will have normative implications for what individuals ought, or have reason, to do (both individuals who are part of those institutions and individuals on the outside who are in a position to influence them.)

¹³ Sinnott-Armstrong. “It’s Not *My* Fault: Global Warming and Individual Moral Obligations”, p. 3.12.

¹⁴ *Ibid.*, p. 312.

¹⁵ Though it is worth noting that, even here, instances of the Problem of Collective Harm can arise. For many forms of social activism – rallies, petitions, boycotts – we can wonder what difference most particular individuals make. While a large number of acts of the relevant sort (signatures, bodies marching, etc.) are often needed for these forms of activism to be effective, it often doesn’t seem that any *individual* such act will make a difference. It doesn’t seem that the outcome would be different with one more or less signature on the petition, for instance.

But while there are, indeed, other levels to the moral story of many real-world Collective Harm Cases, addressing the Problem of Collective Harm remains a crucially important part of understanding these cases. Sinnott-Armstrong's conclusion that we should work to change the law so as to make it illegal to go on a Sunday drive just for fun, *even though* there is nothing at all wrong with going on a Sunday drive just for fun, cannot be correct. Even if it is true that we should work to change the law in that way, what doesn't seem true is that what we would be doing is working to make it illegal to do something that we have *no independent moral reason* to refrain from doing. Whether or not the government makes it illegal to do so, we should be able to explain why individuals have moral reason *not* to do things that will lead to serious, avoidable harm.

For a different example, consider the case of a national election. It is true that some large institutions will be in a position to make a difference. Through their ad campaigns, PACs and Super PACs can have a major influence on the votes of a large portion of the population. Similarly, various news organizations can have a major influence on the way certain demographics vote. This likely means that these organizations – or, in other words, many of the individuals involved in them – have certain moral obligations, and there are interesting and important questions to ask about these obligations. But we shouldn't take this to mean that it is acceptable to conclude that most individuals have no moral reason to vote in the national election. The fact that there are people and organizations that can make a difference to the results of the election, and so who clearly do have moral responsibilities related to the election, does not allow us to dismiss the Problem of Collective Harm that arises with respect to why an individual should vote.

The point is that, even if in many cases there are important things that ought to happen at the institutional level (e.g. making it illegal to drive a gas-guzzling SUV just for fun), if we are to make sense of the morality of Collective Harm Cases, we need to make sense of what is going on all the way down to the individual level. And we cannot make sense of what is going on at the individual level – what individuals qua individuals ought, or have reason, to do – in a satisfactory way, *unless* we can resolve the Problem of Collective Harm.

1.4 Methodology

Before going on to Chapter 2, I want to say a couple things about my methodology. The first is that my overall approach in this dissertation is one of looking at the Problem of Collective Harm *directly*, rather from within the

perspective of a particular moral theory. In Chapter 2 we will largely be thinking about proposals that qualify as non-consequentialist,¹⁶ and in Chapter 3 we will be focused on an attempt at a consequentialist solution to the problem. But my overall line of argument and the proposal that I develop in Chapter 4 neither presuppose nor aim to defend any one global moral theory. My primary concern is with the Problem of Collective Harm *as it arises* in everyday practical and moral deliberation, rather than the problem as it is for any specific moral theory.

Turning to the second methodological point, many of the philosophers who have put forth the proposals that we will look at in Chapters 2 and 3 have been focused on whether there is a moral *obligation* to act in the relevant way in Collective Harm Cases.¹⁷ My proposal in Chapter 4, and the way I tend to present things elsewhere (particularly in Chapter 2), puts the focus, instead, on the question of whether there is *reason* to act in the relevant way. One might wonder whether this is adequate. Don't we need to be able to explain why one *not just* has some reason to act in the relevant way, but is *morally obligated* to do so? Surely – you might think – it is not just that one has some reason not to turn up the dial in the Harmless Torturers case, but that one is obligated not to turn it. We need to explain – the thought goes – why it would be wrong to turn the dial.

While the Problem of Collective Harm is often phrased as one about moral obligations, I am intentionally not approaching it in this way. For one thing, the fact that I am not coming at the problem from within the perspective of a specific, developed global moral theory has made it more natural to talk about reasons rather than moral obligations. But, more importantly, I think that starting with the question of whether or not we are morally obligated to act in the relevant ways is methodologically problematic. When one says, “but it won't make any difference”, more than just saying “it doesn't seem there is an *obligation* to act in this way”, one is saying “there doesn't seem to be *any point at all* in acting in this way”. If it seems that acting in the relevant way won't make a difference, it is hard to see how there could be *any* reason to do so. Acting in this way looks, in this light, like a mere waste of effort. There is a lot of room between ‘there is *no point at all* in *X*-ing’ (or, ‘it would be a mere waste to *X*’) and ‘one is morally *obligated* to *X*’. And, in my

¹⁶ Our discussion will centre around three proposals, two of which are clearly non-consequentialist, and one of which is presented by Parfit as a consequentialist proposal – though there is disagreement about whether consequentialists could actually adopt it. But in addition to the fact that two of the three proposals are clearly of a non-consequentialist variety, we will also talk in Chapter 2 about why it does not seem that one could get out of the problem by *simply* applying standard non-consequentialist moral theories.

¹⁷ This is true, also, of Sinnott-Armstrong and Jackson.

view, it is *because* people have looked straight away for a moral obligation, without attending to this more basic issue, that many have really missed what is hard about these cases, and as a result, have not adequately addressed the problem. If we think that by donating \$15 to Oxfam instead of going to the movies tonight you could make a difference for the better to some impoverished person (or people), we can still debate the question of whether or not you are morally obligated to donate the money to Oxfam. We can debate whether it is impermissible for you to go to the movies, if you could make even a small difference for the better to someone out there in the world who needs help. But if it doesn't seem that making the donation will make *any difference at all* to anyone, not only is it not clear whether you are obligated to donate it; it's hard to see how there could be *any* morally relevant reason to donate it. This seems to me to be the more basic and distinctive challenge that the Problem of Collective Harm presents us with. And putting the problem in terms of moral obligations has, I think, encouraged people to skip over, and even – in some instances – make unwarranted, implicit assumptions about this central, more basic part of the puzzle.¹⁸

¹⁸ I think this is part of what is behind the central issue that we will discuss with the proposals in Chapter 2. It also, I think, explains what is going on with Walter Sinnott-Armstrong's article. Sinnott-Armstrong denies that we are under a moral obligation to refrain from doing things like driving a gas-guzzling SUV just for fun, and says that our obligations are, instead, to work to change what the government does. This is not itself an unreasonable view about our *obligations*. But the problem is that it does not provide an adequate response to the "it makes no difference" argument. To the extent that Sinnott-Armstrong actually sees himself to be responding to the "it makes no difference" line of thought, I think the root of the problem is that he is skipping over what is difficult about it, and is – rather – assuming that there is at least good reason to refrain from driving the gas-guzzler. With that assumption, his proposal about our moral obligations seems fairly plausible. The assumption that there is at least some reason to refrain actually shows up quite clearly in his writing. For instance, he writes:

Some environmentalists keep their hands clean by withdrawing into a simple life where they use very little fossil fuels. That is great. I encourage it. But some of these escapees then think that they have done their duty, so they rarely come down out of the hills to work for political candidates who could and would change government policies. This attitude helps nobody. We should not think that we can do enough simply by buying fuel-efficient cars, insulating our houses, and setting up a windmill to make our own electricity. That is all wonderful, but it does little or nothing to stop global warming, nor does this focus fulfill our *real* moral obligations, which are to get governments to do their job to prevent the disaster of excessive global warming. (Sinnott-Armstrong, p. 312, my underline).

He says nothing to explain why he thinks "it is great" and "wonderful" when individuals do these things. And yet, in my view at least, the real puzzle that the "it makes no difference"

A closely related (but not identical) reason for focusing on *reasons* rather than obligations is that I don't think it is obvious that there should be something as strong as a moral obligation in many of these cases. I think it's clear that we *should* be able to explain why there is reason to take your bicycle and not the car, given the way in which widespread car pollution is leading to immense environmental damage. If we cannot make sense of this, this is obviously troubling – both from a theoretical and practical perspective. But I don't think it is similarly clear that we should be able to say that you are *morally obligated* to refrain from taking your car – that you would be acting wrongly if you drive on a given occasion. Now, in some cases, like Harmless Torturers, it might seem clearer that we should be able to get a moral obligation. But, since it's not clear that we should always be looking for a moral obligation, it is better to at least begin with the more modest aim of trying to make sense of how there is some morally relevant consideration in favor of acting in the relevant ways in these cases – whether or not it amounts to, or can support, a full-blown obligation.

argument gives is that it's not clear how there could be any point at all. If – as Sinnott-Armstrong suggests – it “helps nobody” if I greatly reduce my fossil fuel consumption, what reason do I have to do so?

Chapter 2

Rejecting the Inference

2.1 Introduction

In this chapter I will discuss currently available rejecting-the-inference proposals: proposals that attempt to establish that there is moral reason for action in Collective Harm Cases even when doing so won't make a difference. The chapter will centre around three such views: 'Membership', 'Participation', and 'Fairness'. I will discuss these views individually in sections 2.2 – 2.4. I will then go on, in 2.5, to argue that there is a common, underlying problem with all three: while they do identify considerations *other than* the difference you make, it's not clear that these considerations can make sense as providing reasons for action as long as it's true that your act won't make a difference. This is because as long as it seems that acting in the relevant way would be causally superfluous with respect to the outcome in question, the explanations they are offering do not seem to work. Bringing out this common problem will clarify why Collective Harm Cases pose a problem not just for consequentialists but for non-consequentialists as well, and it will help focus our attention on what the central issue really is.

2.2 Membership

In *Reasons and Persons*, Derek Parfit proposes that we can solve the Problem of Collective Harm by appealing to “what groups together do.”¹⁹ Even if you do not yourself harm or benefit anyone, you may be a member of a *group* that does, and this can explain why you have acted wrongly or rightly. In Drops of Water, for instance, while adding your pint to the cart does not make a difference to any of the suffering men, your act is a member of a set of acts that together make a big difference for the better (our acts *taken together* prevent a lot of harm), and for this reason it is the right thing to do. Doing so makes you part of a group that benefits others, even though you do not benefit anyone yourself. Similarly, in Harmless Torturers, while turning up the dial

¹⁹ Parfit, *Reasons and Persons*, p. 82.

only one notch makes no difference to the harm suffered by the victim, it is one of a set of acts that *together* cause him a lot of pain and is therefore wrong. We together make a difference for the worse, and so each of us has acted wrongly. Parfit captures this view with the following principle, which he calls “C7”:

(C7) Even if an act harms no one, this act may be wrong because it is one of a *set* of acts that *together* harm other people. Similarly, even if some act benefits no one, it can be what someone ought to do, because it is one of a set of acts that together benefit other people.²⁰

A question that arises with this proposal is: what determines which acts are in the set? Why can't each individual say, 'my act makes no difference, and so it's *not* part of the set of acts that together make a difference for the worse?' It's unclear, in other words, why I should count as part of the group that harms if what I do *makes no difference* to that harm.

Parfit gives a principle for determining who is in the group, but his principle only reinforces this concern, rather than addressing it. The principle is:

(C8) When some group together harm or benefit other people, this group is the *smallest* group of whom it is true that, if they had all acted differently, the other people would not have been harmed, or benefited.²¹

C8 successfully excludes acts that we would want to count as totally irrelevant. For example, without C8, one could reply to C7 as follows. Take the set of acts consisting of the thousand torturers' dial turns *plus* Fred Astaire's dancing the rumba.²² This set of acts makes a difference for the worse: had they not been performed (had the torturers not turned up the dials and Fred not danced the rumba), the victim would not have suffered. But it makes no sense to conclude

²⁰ Ibid., p.70. Parfit intends this to be a consequentialist proposal. He says that, while it is tempting for consequentialists to think that the only effects that matter when assessing the morality of an act are the effects of that individual act, this is a mistake. The consequences of a *set* of acts of which your act is a member are relevant. For a discussion of whether Parfit's proposal could be accepted by consequentialists see Jackson, "Which Effects" and Eggleston's "Should Consequentialists Make Parfit's Second Mistake: A Reply to Jackson." In this chapter, though, I am interested in evaluating the proposal independently of whether it would be attractive to a consequentialist or not.

²¹ Parfit, *Reasons and Persons*, p. 71.

²² The example of Fred Astaire is from Parfit. See Ibid., p. 72.

from this that, therefore, Fred Astaire ought not to have danced the rumba. C8 deals well with this sort of objection. Fred Astaire is not part of the smallest group that is such that if everyone in it had acted differently, the victim would not have been harmed. Leave Fred Astaire dancing the rumba and change only the torturers' acts and the harm would have been entirely avoided. But the problem is that this is also true of each of the torturers. We are assuming that none of the torturers makes a difference for the worse. The difference in voltage made by any of their individual acts is too tiny to itself make a difference in pain. Things will be the same for the victim give or take any single act, and this implies that *no torturer* is part of the smallest group that is such that had they all acted differently, the harm would have been avoided. According to C8, then, no particular individual is part of the group that together harms the victim. C8 thus only reinforces the thought that if I make no difference in harm, it cannot be asserted that I am part of the group that harms – at least, not if we want to derive moral implications directly from group membership.²³

Still, while it's not clear that Parfit's Membership proposal can work,²⁴ its central thought might seem to be on the right track. Its central thought is that even if you do not *by yourself* make a difference, if *you together with others* do, this can explain why you have acted wrongly or rightly. The next proposal – the Participation approach – can be thought of as an attempt to substantiate this idea.²⁵ It employs a more robust conception of *acting with others* – of participating in a collective action – and it attempts to explain how *merely* being an act of participation can have normative implications.

2.3 Participation

In his book *Complicity*, Christopher Kutz develops an account of what it is to participate in a collective action and uses this to explain how individuals can be held accountable for collectively caused harms. On Kutz's account you *participate* in a collective action when you act with a “participatory intention”,

²³ On Jackson's proposal, on the other hand, since saying that the group did wrong has no normative implications for individuals, saying that Fred Astaire is part of the group is perfectly fine.

²⁴ For other critiques of Parfit's Membership approach, see: Jackson, “Which Effects”, Gruzalski, “Parfit's Impact on Utilitarianism,” Otsuka, “The Paradox of Group Beneficence”, and Eggleston, “Does Participation Matter? An Inconsistency in Parfit's Moral Mathematics.” For some replies, see the section on Gruzalski in Parfit, “Comments.”

²⁵ When substantiated in this way, though, it is clearly not consequentialist.

an intention to do your part in what you conceive as a collective project.²⁶ “Jointly acting groups,” Kutz writes, “consist of individuals who intend to contribute to a collective end.”²⁷ Suppose we are building a sandcastle. Your act of filling the bucket with water is part of our collective action of sandcastle building because it is performed with the intention of contributing to our collective end, a sandcastle. You fill the bucket so that we can mix water with sand, which will allow us to sculpt the castle. If you had instead filled the bucket with the intention of pouring it on my head, this wouldn’t have been part of our collective action. Collective actions, Kutz explains, are “individualistic with respect to agency”²⁸ – they are entirely “explicable in terms of the intentionality of individuals.”²⁹ But the content of the individual intentions is “irreducibly collective.”³⁰ What makes your act part of our joint action is that you perform it with the intention of doing your part in our collective project.

With this understanding of collective action in place, Kutz argues that (1) you are accountable for actions that are *ascribable* to you and for the consequences (whether intended or not) of those actions, and that (2) when you participate in a collective action, that action is ascribable to you. When you act with a participatory intention and so take part in a group action, you are one of a number of people who can say “we did it”. It is, in this sense, *your* action, and so you can be held accountable for it and its consequences.³¹ Kutz claims that this holds *regardless* of whether or not what you do individually makes any difference with respect to the relevant consequences. To illustrate, Kutz gives examples such as the following. Suppose that you and I are having a picnic (we are both intentionally participating in the project of picnicking together.) In the process, you place the blanket down on a bed of flowers and ruin it. Even though it is your individual act and not mine that causes this damage, on Kutz’s view I can be held partially accountable because I am an intentional participant in the group action that has this damage as a consequence.³² Your individual act of putting the blanket down is performed as part of our collective action of picnicking together. So, even though none of my individual acts contributes to causing the damage, the damage results from a group

²⁶ Kutz. *Complicity*, Chapter 3. There is a further condition, namely that your conception of the joint project sufficiently overlaps with that of the other members of the group. (Ibid., pp. 94 – 95.)

²⁷ Ibid., p. 89.

²⁸ Ibid., p. 112.

²⁹ Ibid., p. 71.

³⁰ Ibid., p. 67.

³¹ Ibid., pp. 137 – 139.

³² Ibid., p. 154.

action that is ascribable to me, and for this reason I am in some sense accountable.

Now, Kutz's theory is about accountability. Asserting that someone can be held accountable for a harm that has already occurred is not equivalent to saying that this person did something that she *ought not to have done*. Indeed, the conception of accountability that comes into Kutz's theory is specifically not desert-based.³³ On Kutz's conception of accountability, one can be accountable for a harm merely by *association*, even though one did not do anything wrong oneself.³⁴ So, a natural question to ask is: is Kutz's theory supposed to say anything about whether one ought, or has moral reason, to act in the relevant ways in Collective Harm Cases? Or is it only talking about who can be held accountable, in some form or other, for collectively caused harms after the fact? In other words, is Kutz's account *supposed* to provide a solution to the problem that we are concerned with, or is it really concerned with a different – though related – question?

Kutz makes clear that he wants his theory to explain why we have individual moral reasons to *avoid* collectively causing harms to begin with, and not just why individuals can be held in some sense accountable for collective harms once they occur. In explaining his project Kutz writes:

A theory of individual accountability must provide individual agents with *reasons to avoid* ... even marginal participation in harmful acts. In the absence of such a theory, there is little social hope of controlling the essentially collective harms that threaten individual interests.³⁵

It's not clear, though, whether the Participation account can do this in Collective Harm Cases – cases in which each individual act appears to make no morally significant difference. For instance, the Participation approach says that in Harmless Torturers, if I turn up the dial on the torture machine, I am participating in a collective action of torturing the victim. Since I am participating, the collective action is ascribable to me and so there is at least some sense in which I can be held accountable for it. But what is it, according to the Participation approach, that differentiates my act of dial-turning in Harmless Torturers from my acts of picnicking in the picnic example? Perhaps it is true that, in a certain sense, I am accountable for the damage caused to

³³ Ibid., p. 17 – 18. Accountability, for Kutz, concerns “warranted response”, where this is *not* only a matter of responses owed to an agent “because of the rights and wrongs they have done”. (Ibid., p. 18.)

³⁴ See for instance, Ibid., pp. 43 – 44.

³⁵ Ibid., p.123.

the flowers in the picnic example. But what is clear is that I did not do anything I ought not to have done. Everything *I* did was perfectly fine. Perhaps I ought to express apology or help compensate the gardener because of the way in which I am *associated* to the damage through our collective action. That much seems quite plausible. But it is not as though we can say (or would want to say) that I ought not to have bought the baguette for the picnic, or ought not to have gone on the picnic in the first place. As Kutz explains:

I am not at fault with regard to the gardener, but it seems I do nonetheless owe some form of response ... Because I have made our picnic mine by my intentional involvement, I have also made its consequences mine. My accountability is therefore intermediate, between what one owes for faultless harms (for *I* was not at fault) and what one owes for faulty harms (for *you*, and *hence we*, were). Your fault sticks to me but not with its full force.³⁶

Are things similar in the Harmless Torturers case? My individual act of turning the dial makes no difference to the suffering the victim undergoes – things would have been the same for him had I not done it. So, it's not clear that the Participation approach would say that I am at fault. Perhaps I can be held accountable, in some sense, because of the way in which I am associated (through my participation) with what *was* causally responsible for the harm.³⁷ But if what I do makes no difference with respect to the harm, it doesn't seem that the Participation account gives us any grounds for thinking that I did something I ought, or had moral reason, not to have done.

Perhaps the reply would be: the picnic case is described as one in which it is “reasonable for me to trust your discretion.”³⁸ So, I have no reason to *expect* the bad outcome. This is why we cannot say that I ought to have acted otherwise. Had I foreseen the bad consequences, I *would* have had reason to act otherwise – to insist on placing the blanket down myself, or to chose a park with fewer camouflaged flowerbeds. In Harmless Torturers, on the other hand, the collective project *is* a harmful one, and so I can expect that, while my individual act won't play any significant role, the collective action will result in suffering.

But it's not clear that this reply explains why I have reason to refrain from turning the dial. What expecting the damage in the picnic example

³⁶ Ibid., p.154.

³⁷ Though, in this case, there may be no individuals in the group who are at fault and so whose fault can “stick to me”.

³⁸ Ibid., p. 154.

seems to give me reason to do is to take steps to *prevent* the damage. It gives me reason to do things that will likely make a difference for the better, like placing the blanket down myself, or choosing a different park. It doesn't seem to give me reason to do things that won't make any difference. So, similarly in Harmless Torturers, my awareness of the potential harm might give me reason to do something to prevent or mitigate the harm, if I can. But why would it give me reason to do something that won't make any difference one way or another?

Perhaps the idea is that foreseeing the harm gives me reason to refrain from participating *because* participating could make me accountable. Even if by turning the dial I would have acted faultlessly, if I turn it I will be associated with the harm through my participation. The harm would result from a collective action that is, in a sense, *mine*; this would make me to some extent accountable, and this risk of accountability itself gives me reason to refrain. It's not clear, though, that a concern to avoid this kind of faultless, associative accountability can provide the sort of reason we are looking for in response to the Problem of Collective Harm. I will not press this issue here, though. I will return to this point, somewhat indirectly, in section 2.8, but let's now turn to a different issue.

Another issue that arises is that there are, as Kutz points out, a large number of Collective Harm Cases that simply do not involve collective action. These are cases of what Kutz calls *unstructured collective harms*: cases in which harm results from many people acting in a certain way, without there being any joint project that they are all intentionally participating in.³⁹ Much environmental damage, and resulting harm to humans and animals, is unstructured, rather than the result of collective action. Drivers, for instance, are not intending to do their part in some collective project when they drive. Similarly, people who buy conventional rather than fair trade goods are not intending to do their part in some joint project, nor are people who use pesticide, or who refrain from donating to famine relief efforts, and the list goes on.⁴⁰ So, it seems that there are a large number of Collective Harm Cases in which the Participation approach does not apply, including – as Kutz points out – “many of the most serious collective harms.”⁴¹

One proposal is that if the Participation approach works in the case of harmful collective actions, there may be a straightforward way of extending it

³⁹ Ibid., p. 166.

⁴⁰ Indeed, in most large-scale Collective Harm Cases there is a total disconnect between people's intentions in action and the consequences that result from enough people acting in this way.

⁴¹ Ibid., p. 166.

to cases of unstructured harm. Typically, when there is a risk of unstructured collective harm, there is also an opportunity to participate in a collective action of *preventing* that harm. In Drops of Water, for instance, if none of us add our pints and so the men continue to suffer, we are not intentionally participating in some collective project of harming the men. We are not engaged in any collective action at all. So, this would be a case of unstructured collective harm. But the other side of the coin is that if we *do* add our pints, we would be – it seems the Participation approach could say – engaged in a joint project of alleviating their suffering. Similarly, while drivers are not engaged in a collective project when they drive, we could participate in a collective action aimed at mitigating environmental damage *if* we each reduce the amount we drive with the aim of doing our part in such a project. If the Participation approach is supposed to give you moral reason to *refrain from participating* in harmful collective actions even if you make no difference, it seems it should also be able to explain why you have moral reason to *participate* in beneficial collective actions even if you make no difference.⁴² One problem that arises, though, is that unless people already have the joint project, there wouldn't be any collective action to participate in. So, as long as we do not already see ourselves as having a joint project of reducing environmental damage or saving the suffering men, it's not clear whether the Participation approach could have anything to say.

2.4 Fairness

Let's turn to the third proposal. Some philosophers – including Colin Strang and Garrett Cullity – have argued that we can solve the Problem of Collective Harm by appealing to considerations of fairness.⁴³ This approach first claims that in Collective Harm Cases there is a *collective* obligation to prevent (or to avoid bringing about) the relevant bad outcome. Unless enough people act in a certain way, bad consequences will result and so there is an obligation on the *group as a whole* to ensure that this does not happen. But if we have a collective obligation – the argument goes – it is *unfair* for some but not others to do the work of satisfying it. If I stand by while others do their part in fulfilling the obligation, this is not fair to the others in the group and is

⁴² Kutz discusses unstructured cases in his chapter “Problematic Accountability” (Chapter 6). Contained in his discussion is some talk of “positive” accountability and joint effort to prevent harm. But much of his discussion in Chapter 6 is not relevant here, because it takes us away from the appeal to *participation in a collective action* and instead concerns how we can “foster a sense of accountability” in the absence of participatory intentions.

⁴³ See Strang, “What if Everyone Did That?”, and Cullity, “Pooled Beneficence”.

therefore wrong. Thus, regardless of whether or not one could make a difference by doing so, there is reason to act; “it is a matter of pulling one’s weight in what we all ought to be doing.”⁴⁴

Suppose, for instance, that each of the potential torturers would receive a benefit for turning up the dial – say, fifty dollars. Each knows that he won’t make a perceptible difference to anyone by turning it. So, why should he refrain? According to the Fairness approach, without making any assumptions about the torturers’ *individual* obligations, we can say that there is an obligation on the *group as a whole* not to cause the victim to suffer. Most people in the group must refrain from turning the dial if this collective obligation is to be met, and if most people refrain while you go ahead and get the fifty dollars, this is unfair. So, the fair, and therefore right, thing to do is to refrain.

On at least Cullity’s version of the Fairness approach, the point is *not* that from a collective obligation we can automatically derive individual obligations. Cullity seems to agree with Jackson that saying that there is a collective obligation does not itself say anything about what the individuals in the collective ought to do. He writes:

The wrongness of a collective action does not always distribute onto the actions of individual group members...from the fact that we are collectively acting wrongly, it does not immediately follow that any individual member of the group is acting wrongly.⁴⁵

The idea is, though, that when there is a collective obligation *and* people in the group are trying to fulfill it, it can be unfair for me not to do so as well. If others are doing their part in what we as a group ought to be doing, then I ought to do my part too.

One problem that arises, then, is that the Fairness approach doesn’t seem able to explain why anyone ought to act differently in the *first place*. If no one is doing anything to try to satisfy the collective obligation, then I am not being unfair by refraining from acting myself. Indeed, Cullity makes this exact point, but he doesn’t seem to see it as an objection. He writes:

I have not been arguing that individual imperatives can be derived from every collective imperative... [For example,] perhaps it is rational for a group to be behaving peacefully, but rational for each group member to be contributing to a conflict; and perhaps as a group it would be wrong for us not to make a gesture of reconciliation but if we

⁴⁴ Cullity, “Pooled Beneficence”, p. 17.

⁴⁵ *Ibid.*, pp.13-14.

fail to do so no group member acts wrongly. In both of these cases, it is significant that the collective imperative is not being met. If so, no one is relying on others to satisfy that imperative while excepting himself from doing so. Before the collective action is started, none of us is arrogating any special privileges in refusing to get it started. We are collectively acting wrongly, but no individual is taking advantage of others' propensity to contribute to doing what we ought to be doing.⁴⁶

This is a *problem* for the Fairness approach. This means that as long as all the torturers continue turning up the dials, or as long as no one has given their pint to the water cart, no individual ought to act otherwise. But the Problem of Collective Harm is that it's not clear why *any* individual has reason to give her pint to the cart, or to refrain from turning the dial, or to reduce the amount she drives, and so on. Each individual such act seems to make no difference to the morally relevant outcome, and thus it is hard to see why anyone has reason to do it. If the Fairness approach only kicks in once people have already started acting in the relevant way, it can't really solve the problem.

Furthermore, unless the collective obligation itself has at least some normative implications for individuals, it is hard to see why considerations of fairness should tell me to act when others start doing so. You might do all sorts of things, but typically I have no reason coming from fairness to act as you do. If you bang your head against the wall, it is not as though it is only fair for me to do so as well. If the collective obligation does not itself give me at least *some* reason to act in the relevant way, then why would I have to do so out of fairness?

A solution to both issues would be to take it for granted that the presence of a collective obligation itself gives individuals at least some reason to act. But, if the notion of collective obligation needed for the Fairness approach is one which itself says that individuals have reason to act in the relevant way, there is a risk of circularity. As we just said, the problem in Collective Harm Cases is that it's not clear why any individual has reason to act in the relevant way. The Fairness approach attempts to address this by *first* claiming that we at least know that there is a *collective* obligation and then arguing from there to individual obligations through an appeal to fairness. If in saying that there is a collective obligation we are implicitly saying that individuals have moral reason to act in the relevant way, we may not be

⁴⁶ Ibid., p.18, my emphasis.

justified in saying that there is such a collective obligation in the first place – doing so would seem to beg the question.⁴⁷

There is a closely related concern that is similar to the one raised with the Membership approach. The problem with the Membership approach was that if we want to derive normative implications *directly* from group membership, it does not seem we are justified in saying that I am part of the group if I *make no difference*. Similarly, with the Fairness approach, if we want to derive normative implications for individuals directly from the existence of a collective obligation, then it's not clear we are justified in saying that I am part of the collective. If I cannot make any difference with respect to \mathcal{V} , why am I part of the group on whom the obligation to bring about \mathcal{V} rests?

2.5 The Common Problem

Jonathan Glover dismisses the Fairness approach, claiming that it invokes a notion of fairness that we do not – or, at least, should not – care about. He considers a case in which a car needs to be pushed up a hill; only six people are necessary for the job, but there are eight altogether. Even though six could manage the task on their own, it would be unfair for two to sit by

⁴⁷ Cullity gives a lengthy argument to support his claim about the collective obligation, but none of what he says helps with the issues that we are raising here. Cullity is focused on a particular kind of Collective Harm Case: pools of beneficence, like the Oxfam case and Drops of Water. His argument that we have a collective obligation to contribute to a pool of beneficence goes more or less as follows. If one could help individually, it would be wrong not to. But if we *all* contribute to a pool this would be more effective than if we *all* help individually. Therefore, if the alternatives available to us *as a group* are (1) helping individually, (2) contributing to a pool, or (3) not doing anything, then contributing to the pool is the best (in the sense of ‘most effective’, ‘does the most good’), and therefore *we as a group* ought to help through contributing to a pool. But then, he argues, remove the alternative of helping individually, and consider only the alternative of helping through the pool and not doing anything, and we still have a collective obligation. (Ibid., pp. 6 – 14) The move in Cullity’s argument that we are questioning is the first move that invokes a collective obligation: the move from ‘it would be more effective if we all contribute to a pool than if we all help individually’ to ‘we as a group are under a collective obligation to contribute to a pool’. The concern we are raising has to do with Cullity’s use of the notion of a *collective obligation*. The argument he is giving *presupposes* that he can say that the group is *collectively* obligated to do what would be *collectively* most effective, even if each individual would make no difference. But it is not clear that he can presuppose this. Note that, given that he is making this presupposition, his argument is unnecessarily complicated. The actual “work” in the argument for the collective obligation seems to be done by the move from ‘the consequences would be better if we all contribute to the pool’ to ‘we are collectively obligated to contribute through the pool’. The rest (the move from three options to two options, and the lengthy justification he gives of this move) does not seem to me to add anything.

while others do all the work. This sounds much like the point that the Fairness approach tries to make in Collective Harm Cases. In the voting case, for instance, even though we do not *all* need to vote, it would be unfair for some but not others to do so. But, Glover replies:

We are free to accept the argument from injustice in the car pushing case without accepting it in the voting case. If I do not push the car, the others will have to push a bit harder. Many of us are against the kind of injustice that involves giving benefits to some at the cost of additional hardships to others. But no-one has to vote harder because I do not vote. It seems a dog-in-the-manger version of justice that objects to one person benefiting because others are left unchanged.⁴⁸

Glover is right that there is some disanalogy between the car-pushing case and the voting case. But it misunderstands the Fairness approach to say that it invokes a “dog-in-the-manger version of justice” or that it “objects to one person benefiting because others are left unchanged.” Take the Drops of Water case. Unlike a dog in the manger situation, each of us would benefit from keeping our pints – we would each have it as a refreshing drink.⁴⁹ The point the Fairness reply makes is that it would be unfair for some to keep their pints, receiving the benefits that *each of us* would have enjoyed, while others give theirs up in order to achieve what we *as a group* are obligated to achieve. The type of unfairness involved in keeping my pint is not that doing so benefits only myself and not others – that it benefits one person while leaving others unchanged. It is, rather, that if I keep my water, while others give theirs up, I am not – as Cullity writes – “pulling my weight in what we all ought to be doing.”⁵⁰ I am “relying on others to satisfy ... [the collective] imperative, while excepting myself from doing so.”⁵¹ It seems reasonable to propose that this general sort of unfairness is itself morally objectionable. And someone who finds this view compelling will not be moved by Glover’s reply that we should only care about unfairness *if* it is matter of one person gaining a benefit by making things worse for others. Such a reply simply contradicts the proposal on offer, rather than identifying a problem with it.

So I don’t think Glover’s objection is effective as it stands. But I do think there is an important issue in the vicinity. The Fairness reply says: if we

⁴⁸ Glover, “It Makes No Difference Whether or Not I Do It,” p. 182.

⁴⁹ That is, this is not a situation of one person preventing others from having something that she herself has no use for.

⁵⁰ Cullity, “Pooled Beneficence”, p. 17.

⁵¹ *Ibid.*, p. 17.

collectively ought to bring about some outcome, it is unfair to let others do the work of satisfying this obligation without doing so yourself. You ought to *pull your weight* in what we collectively ought to do, regardless of whether in doing so you will make any difference. Glover's reply is: we shouldn't care about the unfairness in not pulling one's weight if it won't make any difference to the others involved. But it seems to me that the problem is more basic. If acting in the relevant way won't make any difference, then it does not seem that that it *pulls any weight at all*. It doesn't seem to count as doing the work of satisfying the collective obligation.

Consider this revised version of the car-pushing example. We need to get the car up the hill, but one of us – call him 'Small' – has a rare physical condition: while he is otherwise healthy, he has the muscles of a one-year-old. He is, let's say, an adult trapped in a one-year-old's body. Because of this, it makes no difference whether or not Small pushes. If he pushes with all his might, this will exert a tiny amount of force on the car, but this tiny amount of force is *too tiny* to make any difference at all with respect to the task at hand. The car won't progress up the hill any faster if he pushes than if he doesn't, nor will it be any easier for the others involved. And this is so, regardless of how many others push: whether or not enough other people push to get the car up the hill, adding or removing Small from the equation simply won't change things in any relevant respect. In this situation, does it make sense to say that Small ought to push because this is his fair share of our collective burden? His act of pushing wouldn't help us achieve the goal. It would be purely superfluous. It is true that there is a burden on the collective to get the car up the hill. But if his pushing won't help discharge this burden, how could it be his 'fair share'? It doesn't seem to be any real share at all. Instead it seems like a mere waste of his efforts. Perhaps Small should do *something* out of fairness. If there is some useful task he could perform (e.g. fetching water for the others), he might have a duty of fairness to do it. But it doesn't make much sense to say that what he ought to do is something merely superfluous and unhelpful.

In Collective Harm Cases *each* individual act seems to be like Small's act of car-pushing. It isn't going to make any difference that matters, and so seems to be merely superfluous. But what we are seeing is that the Fairness approach can't get off the ground if this is so. The Fairness approach appeals to (a) the presence of a collective obligation to bring about some outcome, and (b) the idea that it is unfair to let others take on the task of satisfying this obligation while excepting yourself from doing so. But as long as my act *won't* help satisfy the collective obligation – as long as it would be merely superfluous with respect to what we are collectively obligated to do – (a) and (b) cannot explain why I should do it. Even if there is a sense in which we are collectively

obligated to relieve the men's suffering in Drops of Water, if adding my pint won't help to relieve their suffering, this collective obligation wouldn't give me reason to add my pint. And bringing in the point about fairness does not fill the gap. How could adding my pint be my *fair share*? It seems to be no real share at all.⁵²

The problem we are running up against with the Fairness approach arises just as much with the other two views. All three views try to solve the Problem of Collective Harm by identifying relevant reasons for action that don't have to do with whether or not your individual act can make a difference. These views are, thus, focused on showing that even if your act cannot *itself* play any significant causal role in bringing about the outcome, there are *other sorts* of reasons for action. But the problem is that the considerations they pick out can't seem to get off the ground as providing reasons for action *as long as* your act won't play any significant causal role. As long as your act would be causally superfluous, the reasons they are offering don't really make sense.

We have essentially already seen this with the Membership approach. The Membership approach says that even if you don't harm anyone *yourself*, if you together with others do, this explains why you have acted wrongly. But it's not just that you don't cause the harm in question *all by yourself* in Collective Harm Cases; it's that you *make no difference*. Your act doesn't seem to play any sort of significant role in bringing about the harm. Instead it seems to be entirely extraneous with respect to it. So, why should you count as part of the group to begin with? Why is your act part of the set of acts that harms, if it does not contribute in any non-superfluous way to bringing about the harm? The point, in other words, is that as long as your act doesn't play any significant role in what the group does, we cannot simply *point* to what the group does to explain why you have acted wrongly (or rightly). And yet this is just what the Membership approach tries to do.

The Participation approach does not do any better. We raised some concerns about whether mere participation (without making any difference) can explain why you ought, or have reason, to refrain from acting. But there is a more basic issue: if your act isn't going to make any difference, it's not clear it makes sense to call it an act of participation to begin with. In Harmless Torturers, for instance, the Participation approach says: 'even if each torturer

⁵² Cullity touches on this objection, and offers a reply. (See *Ibid.*, pp. 16 – 17.) I think that part of what he says there is on the right track, but there is a considerable amount of confusion in his reply. The issues there are quite similar to issues that we will see with denying-the-description approaches in Chapter 3. Cullity fails to distinguish properly between the *underlying dimension* and the *morally significant outcome*, and between *perceptible* and *morally significant*.

makes no difference to the harm, the torturers would be engaged in a collective action of harming the victim. By turning the dial, you are participating in this collective action, and when you *participate* that action is – in a sense – yours, and so you are accountable for it.’ But, on closer inspection, it doesn’t seem this story can get off the ground. For any given torturer, how can we say that she is participating in this collective action? To participate in a collective action is to intentionally *do your part* in a shared project – in this case, a project of torturing the victim. But for any given “torturer”, her act of turning up the dial a single notch makes no difference to the suffering of the victim. It doesn’t seem to play any non-superfluous role in causing the harm. So, how can we say that in turning it she is intending to *do her part* in this joint project? Her intention in turning up the dial could just be to acquire fifty dollars at *no cost to anyone*. Indeed, if she decided to turn the dial after realizing that it won’t make any difference, this probably *is* her intention. Kutz explains that the intentions that make an act participatory need not be “explicit in deliberation”; they can be “functionally implicit” in one’s behavior.⁵³ But this does not help here. We can’t say that, while an intention to participate in a collective action of torturing the victim may not be explicit in her deliberation, it is functionally implicit in her behavior. Her behavior, after all, is behavior that *makes no difference* to the victim.

Out of all the examples of Collective Harm Cases we have seen so far, most are not cases of structured collective harm – most are cases in which harm results from a “confluence of individual behaviors,”⁵⁴ rather than from a collective action. Harmless Torturers, at least, had seemed to be one example of a structured Collective Harm Case. But what we are finding is that even in that case, we cannot get the Participation account going because its not clear that we can count any particular individual act as part of the collective action. As long as it’s true that one’s individual act would play no significant role in bringing about the outcome, it doesn’t seem we can say that it is an act of participation in the project of bringing about that outcome.

We have said that to handle cases like Drops of Water or environmental damage – unstructured cases – it seems that the Participation account could talk about participating in a collective action that *prevents* harm, rather than just about accountability for a collective action that causes harm. But, of course, the same problem arises here. Again, participating in a collective action is to intentionally *do your part* in a shared project. In Drops of Water it seems it would be a project of alleviating the men’s suffering; in the case of environmental damage it seems it would be a project of mitigating

⁵³ Kutz, *Complicity*, p. 82.

⁵⁴ *Ibid.*, p. 166.

environmental damage. But how can doing something that would be merely *superfluous* with respect to that project be ‘doing your part’? Kutz writes, “my part is defined as the task I ought to perform if we are to be successful in realizing our shared goal.”⁵⁵ A task that *makes no difference* is surely no such task. Indeed, if we could have said that adding my pint is the task that I *ought* to perform if we are to relieve the men’s suffering, there would not have been a problem there to begin with. The Problem of Collective Harm arises because it seems that adding my pint won’t do *anything* useful when it comes to alleviating the men’s suffering.

Rejecting the Inference approaches try to solve the Problem of Collective Harm by accepting that your individual act won’t make any difference, and explaining why nonetheless there is still reason to do it (or, depending on the case, reason to refrain). But the problem we are running into is that the considerations they identify don’t really seem to apply *unless* your act might play a significant causal role with respect to the outcome in question. And yet not playing any such role seems to be just what we are accepting when we accept that your act won’t make a difference.

2.6 Can’t We Just Apply a Non-Consequentialist Theory and Call it a Day?

It is sometimes thought that it is only consequentialists who will have trouble with the Problem of Collective Harm.⁵⁶ After all, non-consequentialists do not think that all that matters morally is the difference you make in outcome. So – the suggestion goes – their theories will do just fine. The common problem we are running up against with the three proposals under discussion illustrates why this view about non-consequentialist theories is mistaken.

It is important to notice that we cannot solve the Problem of Collective Harm by giving a reason for action that has nothing at all to do with the morally significant outcome in question. I may have all sorts of reasons to refrain from driving: to get more exercise, to save money, to reduce the risk of accident, to avoid stressful traffic jams. But, of course, none of these can provide a solution to the Problem of Collective Harm that arises with respect to car driving and environmental damage. We need a reason for action that

⁵⁵ Ibid., p. 81.

⁵⁶ This is a reaction I have sometimes received from philosophers in conversation about the Problem of Collective Harm.

connects in some appropriate way to the fact that widespread car use causes environmental damage.

The problem for non-consequentialists is that, even though they do not think that *all* that matters morally is the difference you make in outcome, it's not clear how there could be any reason to *X*, which connects appropriately to outcome *Y*, if *X*-ing cannot make any difference with respect to *Y*. If acting in the relevant way would be merely superfluous with respect to the outcome in question, it's not clear that we can get *any* relevant story going as to why you have reason to do it. Suppose, for instance, that you subscribe to a theory of virtue ethics. In particular, suppose you think that the right thing to do is what the virtuous person would do. Well, as long as adding your pint won't help the suffering men, it does not seem we can say that the virtuous person would add her pint. The virtuous person, after all, does not act wastefully, and we haven't explained why adding your pint isn't just wasting it. Unless we can explain how acting in this way would help to relieve the men's suffering, it doesn't seem virtue ethics can give a good answer to the Problem of Collective Harm.

Or, suppose you are a Kantian. In particular, suppose you think Kant's Formula of Humanity gets things right. You agree that you should "act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means."⁵⁷ Can this explain why one ought to refrain from turning the dial up a single notch in Harmless Torturers? If turning the dial won't make a difference for the worse to anyone, it's not clear why it would count as treating someone as a mere means. If you are not contributing in any non-superfluous sense to bringing about the victim's suffering by turning the dial, then how can we say that in doing so you are treating him as a mere means? You don't really seem to be treating him in any way at all by doing that particular act.⁵⁸

⁵⁷ Kant, *Groundwork of the Metaphysic of Morals*, 4: 429.

⁵⁸ Now, one might think that Kant's first formulation of the Categorical Imperative, the Formula of the Universal Law of Nature, does better. According to this formulation, to determine whether an act is permissible one must ask whether it is possible to will the maxim as a universal law: can we conceive without contradiction a world in which everyone acts on the same maxim. To check for this, we must look for, first, a "contradiction in conception" and, second – if there is no contradiction in conception – a "contradiction of the will". If there is a contradiction in conception, one has a perfect duty not to act in this way. If there is only a contradiction of the will, one has an imperfect duty not to act in this way. And if there is neither, acting in this way is permissible. Now, there is a difficult question concerning how we are supposed to individuate maxims. But it seems clear that, however we individuate the maxims in Collective Harm Cases, we are not going to get a contradiction in conception. Whether the maxim is, for instance, 'do not reduce the amount you drive' or 'drive when doing so is convenient' or 'drive your car when you want to and when it won't make a difference for the worse', there is no contradiction in imagining a world in which everyone

2.7 The Common Problem and the Way Forward

In Collective Harm Cases, one's individual act (or omission) doesn't seem to make any difference. It appears that things will be *the same* in all relevant respects, whether or not it is performed. And what this seems to mean is that acting in this way would be causally superfluous, that it would not play any significant causal role with respect to the outcome in question. This is *why* it is hard to see how there could be any point in acting. While one knows that unless enough people act in this way bad consequences will result, one is faced with the apparent *uselessness* of acting in this way oneself.

The views we have been considering in this chapter approach the problem by trying to show that, even if acting in the relevant way can't make a difference, there is still reason to do so. They do this by turning the focus away from the concern about the apparent causal superfluity of your individual act, and onto other sorts of considerations: group membership, participation in a collective action, fairness in the context of a collective obligation. But the problem we are running into is that the considerations these views identify can't get a grip as long as your act would be merely superfluous. Unless your act could play a significant causal role with respect to the outcome in question, the explanations these views are offering as to why you have reason to do it can't get off the ground.

What I think this reveals is that to solve the Problem of Collective Harm, we need to show that one's individual act *isn't* purely superfluous. We need to show that we are mistaken when it seems to us that an individual act of the relevant sort can't play a significant causal role. Of course, it's not clear that we can do this. But it seems to me that unless we can, there won't be a satisfactory way of solving the problem.

If we can show that an individual act of the relevant sort isn't merely superfluous, and rather could genuinely help to bring about the relevant outcome, this would address most of the problems we have raised with the

acts on this maxim. One could still act on the maxim in a world in which it is universalized. Whether or not there is a contradiction of the will is a more complicated question, and this is because there is much that is under debate about how to understand this test and, specifically, what sort of contradiction we are looking for. Also, whether or not there is such a contradiction in Collective Harm Cases might very well depend on how we individuate the maxims. I don't want to get into these interpretative questions here. The point is, though, that how the universalization test handles Collective Harm Cases comes down to whether or not the relevant sorts of acts (or omissions) pass the *contradiction of the will* test and it is not clear whether or not they do. I would like to thank Markus Kohl for a helpful discussion about the Kantian universalization test.

views seen so far. But the point isn't just that this would allow views like Fairness, Membership and Participation to get off the ground. Rather than just being a matter of filling in the hole that we've identified with these views, showing that an individual act can play a *significant* part in bringing about the outcome is – it seems to me – the *central task* of solving the Problem of Collective Harm. I don't think we can solve the problem in a satisfactory way without showing this. But, moreover, if we *can* explain how it is that your act is genuinely helping to bring about the outcome, this would largely address the Problem of Collective Harm – or, at least, the problem as it arises in moral deliberation (rather than the problem as it is for any particular moral theory). What leads us to fail to see how we could have reason for action in Collective Harm Cases *is* the impression that doing so would be useless – that it wouldn't help to bring about (or, depending on the case, wouldn't help to prevent) the outcome in question. If we could show that this were a mistake, this central issue would be addressed.

This pushes us toward seeking a solution in the second category, a denying-the-description approach. Such views try to show that it isn't actually true that one's individual act won't make a difference in Collective Harm Cases. They claim that, contrary to appearances, individual acts do, or at least might, make a morally significant difference in all Collective Harm Cases, and this is why we have reason for action. This line of response is largely taken by consequentialists, but what our discussion in this chapter shows is that non-consequentialists also have a stake in whether or not such a view can succeed.

In Chapter 3, I turn to discuss attempted solutions of this sort. Unfortunately, we will find that taking such an approach is not promising. We are not in a position to deny that there are Collective Harm Cases in which an individual act just won't make a difference. So where can we go from there? On the one hand, it doesn't seem we can get any adequate solution going *unless* we can obtain that one's act plays a significant, non-superfluous role. But on the other hand, we will find that we are not in a position to deny that there are cases in which individual acts can't make a difference. The way forward, I will propose, is to reject the assumption that not being able to make a difference implies that your act would be causally merely superfluous. The key to solving the Problem of Collective Harm is to make sense of how it can be that an act can *help* to bring about an outcome in a morally significant sense, even when it cannot make a difference. I will develop this line of response in Chapter 4.

2.8 The Participation Approach Revisited

Before moving on to consider views in the second category, I want to revisit the Participation Approach and point out that in a small class of Collective Harm Cases this approach escapes the central problem that we raised with it. I think it is instructive to see that even in such cases, the Participation approach cannot provide a satisfactory solution to the Problem of Collective Harm that arises – at least not unless it is able to incorporate an account of why acting in the relevant way is not causally merely superfluous.

We saw that many Collective Harm Cases do not involve participation in a collective action (unstructured cases), and we saw that even in Collective Harm Cases that might appear to involve collective action – like Harmless Torturers – we cannot count an individual act as genuinely participatory unless we can say that it plays a non-superfluous causal role in bringing about the outcome. But there is an exception to this: Collective Harm Cases in which there is a collective action aimed at some *other end* besides the morally relevant outcome in question but where the harm results from this collective action as an *unintended consequence*. In such a case, regardless of whether or not your individual act plays any role at all in bringing about the harm, you count as doing your part in the shared project because the shared project is aimed at some other end. Does the Participation approach provide an adequate solution to the Problem of Collective Harm that arises in such cases, even if it cannot provide a solution to the *general* Problem of Collective Harm?

There are many examples of *Non-Collective Harm* Cases that are like this. The picnic example was one such case. Our collective end was a picnic, but an unintended consequence was ruined flowers.⁵⁹ But it is actually quite hard to find convincing real-world examples of *Collective Harm Cases* that fit this description. We would need a case in which (i) harm results from genuine collective action; (ii) this collective action is aimed at some *other end* besides the harm in question; but (iii) no *individual act* seems to make any difference with respect to the harm in question. None of the examples we have seen so far fit this description. But we can adjust one of our examples to create such a case.

Recall the pesticide example from Chapter 1. This is an unstructured case: there is no big joint project that pesticide-users are all participating in. But here is a collective action version of it: suppose that there is a monopoly in our city on professional gardening. One massive company, The Harmless Gardeners, does every gardening job. This company has a somewhat unusual

⁵⁹ The picnic case isn't a Collective Harm Case, as *your* act of putting down the blanket on the flowers made a difference for the worse.

business model: each gardener is more or less free to garden however he or she finds conducive to success. The company has as its aim to beautify the city and make profits, and it finds that giving its gardeners a large amount of freedom is key to succeeding in this mission.⁶⁰ So, the company provides its gardeners with the support they need to pursue the collective aim (making profits and beautifying the city) largely as they individually see fit. Now, it just so happens that, while some gardeners go totally pesticide-free, most choose to use pesticide at least on occasion. There are thousands of gardeners and hundreds of thousands of properties, and with all these applications of pesticide, enough leaches into the water supply to cause serious illness in a small portion of the population. Still, while the *overall* use of pesticide causes serious illness, it seems that any individual act of pesticide use won't make any difference for the worse. Each single application of pesticide leaches only a miniscule amount into the water supply, and such a miniscule amount doesn't seem to be enough to make a difference to anyone's health. So, the Problem of Collective Harm arises. Any gardener, on any given occasion can ask: "will it make a difference for the worse whether or not I use pesticide here? And if it won't make a difference, why should I refrain?"⁶¹

The Participation approach replies as follows. Whenever a gardener uses the pesticide, she does so with the intention of doing *her part* in a collective project, a collective project of making the company a profit and beautifying the city. Thus, each act of pesticide use is part of a large collective action, where this collective action has serious harm as an unintended (but foreseeable) consequence. If you participate in this collective action, that action is ascribable to you, and so you are accountable for it and its consequences. This, the claim goes, explains why you have reason to refrain from using the pesticide.

Does this reply work? It does not fall prey to the issue raised with examples like Harmless Torturers. The gardener's act of spraying pesticide

⁶⁰ I'm making this stipulation to avoid complications that take us away from the problem at hand. If, for instance, the gardeners are trained to garden in a very specific way that involves using pesticide and are required to garden in this way, there is a sense in which each gardener cannot opt out, or, at least, a sense in which it is not *their* decisions that matter. On the other hand, one might have quite the opposite sense that the gardeners are *more* responsible than in the Harmless Gardeners case because they have clear reason to object to the company's policies and therefore clear reason not to go along with those policies. Either way, whichever of these impressions seems right, having pesticide use a matter of company policy – I think – clouds our intuitions about the problem we are concerned with.

⁶¹ If you want, for simplicity, you can suppose that no single gardener's total use makes a difference: for any gardener, you could take him or her out of the equation entirely and the harm would be just as bad.

counts as participating in the collective action because it is done with the aim of promoting the collective end. This conception is not compromised when we learn that it can't make a difference to the *harm*, because there is a clearly identifiable collective end and it is *not* the harm. So, here is a case in which it's not true that the Participation approach cannot get off the ground unless we can explain how an individual act plays a significant role in bringing about the morally relevant outcome in question.

But, even though we can say that a single act of pesticide use is an act of participating in this collective action, and even if this does explain why there is at least some reason to refrain, this cannot provide a satisfactory solution to the Problem of Collective Harm that arises in this case. *Any* act of gardening by the gardeners – whether it is planting flowers, trimming a hedge, pulling a weed – counts as participating in this collective action. If what explains why you ought not to have applied the pesticide is *just* that in doing so you are participating in a collective action that harms, then this exact same point applies to every other act of gardening for the company. That is, the Participation account is unable to differentiate between an act of pesticide use and any other act of gardening for the company: *neither* makes any difference that matters but *both* count as acts of participation in a collective action that harms.⁶²

So, the Participation account – as it stands – cannot differentiate morally between using pesticide and planting flowers. It is also not a good sign that, on the other hand, it *does* differentiate between using pesticide while working for the company and using pesticide while working independently. It is unable to explain why a gardener who leaves the company but continues to garden on her own as an independent worker still has reason to refrain from using the pesticide on any given occasion. The Problem of Collective Harm arises in the Harmless Gardeners case because widespread pesticide use results in serious harm and yet no individual application of pesticide seems to make a difference. The fact that the gardeners are working for the same company might add an additional morally relevant dimension that affects the full story about what individuals ought, or have reason, to do. But the central puzzle is the *same* whether or not they are working for a single company. We need an explanation of why an individual act of pesticide use is morally problematic, even though it appears to make no difference. If the account only works when

⁶² This is not to say that the gardeners – *including* the pesticide-free gardeners – don't have reason to refrain from participating in the collective action because of the fact that the collective action causes unintended but foreseeable harm. The point is that, if it cannot distinguish between the pesticide-using gardeners and the pesticide-free gardeners, this is not addressing what is centrally at issue in the *Problem of Collective Harm* that arises.

there is some other joint project (that is, one aimed at some *other* end) that the relevant agents are participating in, this can hardly provide a satisfactory explanation, even in a case in which such a project exists.

Chapter 3

Denying the Description⁶³

3.1 Introduction

In this chapter, I turn to the question of whether we can reply to the Problem of Collective Harm by denying the claim that one's individual act won't make a difference. Perhaps we are simply making a mistake when it seems to us that acting in the relevant way won't make a difference, and so perhaps we can get out of the problem that way. We will discuss this question by focusing on Shelly Kagan's attempt at such a solution. While the focus will be on Kagan's arguments, the key points that I will make in response apply to other currently available attempts at taking this sort of approach.⁶⁴ Kagan's discussion is a good one to focus on because it is one of the most comprehensive attempts at treating the problem by denying the description (that is, it explicitly attempts to handle *all* Collective Harm Cases in this way),⁶⁵ and because it nicely covers the key ways in which such an approach has been taken. In particular, Kagan considers – but only partially adopts – the most common way of taking this sort of approach: appealing to the existence of imperceptible harms and benefits, and their aggregation. But since he is attempting to deal with all Collective Harm Cases and since Kagan does not think that the appeal to imperceptible harms and benefits will work in all cases, his discussion goes well beyond that proposal.

⁶³ A version of this chapter was published in *Philosophy & Public Affairs*. See Nefsky, J., "Consequentialism and the Problem of Collective Harm: A Reply to Kagan." *Philosophy & Public Affairs*, 39 (2011): 364 – 395.

⁶⁴ Here are some examples: Glover, "It Makes No Difference Whether or Not I Do It." Parfit, "What We Together Do." Otsuka, "The Paradox of Group Beneficence." Frank Arntzenius and David McCarthy, "Self Torture and Group Beneficence." Regan, "Perceiving Imperceptible Harms: With Other Thoughts on Transitivity, Cumulative Effects, and Consequentialism."

⁶⁵ For instance, Glover and Otsuka each take a denying-the-description approach in some Collective Harm Cases. But Glover takes a different approach in other Collective Harm Cases (e.g. he does not take such an approach in the voting case). And Otsuka is not discussing all Collective Harm Cases; he is focused only on examples involving pools of beneficence.

Now, Kagan – like most others who have tried to take a denying-the-description approach – is interested in the problem that Collective Harm Cases pose for *consequentialism*. Consequentialism, as he explains, seems to fail by its own lights in these cases. He writes:

Consequentialism appears to fail even in its own favored terrain, where we are concerned with consequences and nothing but consequences...[T]he acts in question need to be condemned because of the results that eventuate from everyone’s performing them. Such a situation ought to be grist for the consequentialist’s mill. Yet ... it seems as though the consequentialist simply isn’t in a position to condemn the relevant acts – given the fact that for any given individual, it simply makes no difference whether or not the individual’s particular act is performed.⁶⁶

Kagan’s aim is to show that – contrary to initial appearances – consequentialism can handle these cases perfectly well. His overall conclusion is that in all Collective Harm Cases, acting in the relevant way either *will* or at least *might* make a morally significant difference, and that therefore we can simply appeal to a standard consequentialist tool to handle them: the notion of expected utility. Kagan’s argument is structured as follows. He breaks down Collective Harm Cases into two basic types, *triggering cases* and *imperceptible difference cases*,⁶⁷ and he argues that, since there is at least some chance of making a difference in triggering cases, they can be handled nicely using expected utility. Thus the real problem, if there is one, is with imperceptible difference cases. However, Kagan says, there is a subclass of imperceptible difference cases in which we can appeal to the existence of imperceptible harms and, in that way, say that each act *does* make a difference. So, it is only the remaining subclass that would be truly problematic for consequentialists. Kagan then goes on to argue that there cannot possibly exist any cases of this problematic sort. Cases that seem to fit the problematic subclass must really just be triggering cases in disguise. Thus, in all Collective Harm Cases there is at least a chance of making a difference, and there is no problem for consequentialists after all.

In what follows, I will call into question Kagan’s analysis of triggering cases and the extent to which the appeal to expected utility can provide a satisfactory solution in such cases, but the focus will be on his discussion of

⁶⁶ Kagan, “Do I Make a Difference?”, p. 107.

⁶⁷ Kagan notes that there may be some cases that are a mixture of these two “pure” varieties (Ibid., p. 119).

imperceptible difference cases. I will argue that his treatment of imperceptible difference cases is problematic on several levels, and thus that he doesn't succeed in showing that they do not pose a problem for consequentialists. The overall conclusion of the chapter is that the project of responding to the Problem of Collective Harm by denying the description – by arguing that we are making a mistake when we think that an individual act won't make a difference – is not a promising one. While in some specific cases such an approach might work, we do not have and cannot expect to have a general solution to the problem along these lines.

3.2 Imperceptible Difference Cases versus Triggering Cases

In this section I summarize Kagan's way of dividing Collective Harm Cases. Note, though, that I will later argue that this division is misleading and should be revised.

In Kagan's *imperceptible difference cases* each act pushes things further along in an imperceptibly tiny way. Enough such changes add up to a noticeable, morally relevant difference, but any single change of this sort is – it seems – itself too small to make a difference to what matters morally. Kagan explains:

[In these cases] it isn't literally true that my act makes *no* difference at all...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.⁶⁸

Kagan gives the Harmless Torturers case as an example. In Harmless Torturers, if enough people turn up the dial, the victim will be in excruciating pain. But each single turn of the dial increases the voltage by an amount that is too tiny to be perceived. Thus, while each torturer does make a difference to the voltage (the underlying dimension), this difference is *imperceptible* and so not a morally relevant difference in outcome.

In contrast to cases in which each act makes an imperceptible difference, in *triggering cases* each act most likely makes “no difference at all.”⁶⁹ In these cases, some act triggers the relevant outcome: a threshold number *n* of

⁶⁸ Ibid., p. 117.

⁶⁹ Ibid., p. 118.

acts of a certain type is needed for a particular morally relevant outcome, and so when the n th act is performed, this triggers the change.⁷⁰ But it is most likely that no individual will make a difference because it is most likely that the threshold will not be hit exactly.⁷¹

Kagan gives the following example. Even if we agree that factory farming chicken causes more harm than good overall, it seems that, for any individual, it will not make a difference whether or not he buys a chicken in a given day. It seems unlikely that the production of chickens by a large factory farm is affected by a single purchase. So, this is a Collective Harm Case, and Kagan suggests that it is a triggering case. It is not that each act makes an imperceptible, and so not morally relevant, difference; rather, most likely one's act makes no difference, but there is a chance that it will make a substantial difference for the worse.⁷² Kagan explains:

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....Here then it makes no difference... 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 and raised and tortured.⁷³

⁷⁰ Note that the threshold need not be knowable to us. Also, it could be that in a given case the location of the threshold shifts depending on various factors. Finally, as we will see with the example below, there can be many thresholds in a given triggering case, each corresponding to a different morally relevant outcome.

⁷¹ Ibid., pp. 125–126. Kagan at first describes triggering cases a bit differently, saying that in these cases “most acts make no difference, but some single act makes a great deal of difference.” (Ibid., p. 118) But he revises this later: it is *not* that most acts make no difference but the triggering act makes a great deal of difference. If the threshold is reached and exceeded, then had the act that in fact triggered the outcome not been performed, the outcome would still have come about. The idea is rather that each act has a chance at making a difference, because there is a chance that a threshold will be hit *exactly*. When a threshold is hit exactly then, had anyone not acted, the outcome would have been different. In a footnote, Kagan says that there may be some triggering cases that have a different structure than that of thresholds: “Suppose that each person objectively has a $1/n$ chance of triggering some bad result, regardless of what others do. Here size of the cohort would be irrelevant, since only the person who *actually* triggers the bad result makes a difference.” (Ibid., p. 127, n.10) It seems to me, though, that such a case would not be a *Collective Harm Case*, and so is simply not of the kind that is under consideration in his paper.

⁷² We will see below that it is misleading to say that these are the two options.

⁷³ Ibid., p. 122. Note that the idea is that at the start of each day the count goes back to zero.

Kagan is describing the chicken purchasing case as a triggering case with multiple thresholds, each corresponding to a different morally relevant outcome. The thresholds are 25 chicken purchases, 50 chicken purchases, and so on – corresponding to 25 more chickens raised, 50 more chickens raised, and so on. If the number of chickens bought is some exact multiple of 25 then each person who bought one that day made a substantial difference for the worse: for each person, had he not purchased a chicken, the relevant multiple of 25 threshold would not have been reached and another crate would not have been ordered and – the claim goes – another 25 chickens would not have been tortured.⁷⁴ But if the number of chickens ordered in a given day is *not* an exact multiple of 25 then it seems no one made a difference: for each person, had he not purchased a chicken, the same number would still have been raised and tortured.

3.3 Triggering Cases: The Expected Utility Solution

Kagan argues that triggering cases can be straightforwardly handled using a standard consequentialist idea: that one ought to do whichever available act has the greatest expected utility.⁷⁵ Rather than saying one ought to do what will in fact bring about the best outcome, this view says that one ought to do what will have the best *expected* outcome.⁷⁶ In triggering cases it is often unlikely but still possible that you will make a substantial difference. In such cases, Kagan says, consequentialism can condemn your act because its expected utility is negative.⁷⁷ To illustrate, consider voting in a national election – another example of a triggering case. If it were a hundred percent certain that your vote would not make a difference to the outcome (that the winner would be same with or without your vote) and that it would be a bit costly to you, then the expected utility of voting would be lower than that of refraining. However, it is not certain that your act won't make a difference. There is a tiny chance that there will be a tie or a one-vote-win, in which case

⁷⁴ This assumes that no one else would have bought a chicken *had* you refrained. For example, it could be that had you not purchased yours, someone else would have reached the front of the line before closing time and bought one.

⁷⁵ *Ibid.*, p. 124. One could say “greatest expected goodness” or “best expected outcome,” instead of “utility,” if one wanted to allow for a wider view than the utilitarian’s about the goodness and badness of consequences.

⁷⁶ An act’s expected outcome is its possible good effects multiplied by the probability that they will be produced *minus* its possible bad effects multiplied by the probability that they will be produced.

⁷⁷ This response to the triggering variety of Collective Harm Cases is not new. For instance, Parfit gives this reply in *Reasons and Persons*, pp. 73-75.

each vote, including your own, would make a great deal of difference. So, the consequentialist can say that even if your vote does not in fact make a difference, the expected utility of not voting may be less than that of voting. A tiny probability of very bad consequences (e.g. Dr. Evil becomes president) may yield a low expected goodness, even when refraining from voting has a high probability of fairly good consequences (e.g., you have time to take your child to the park).

Now, Kagan seems to think that in all triggering Collective Harm Cases the relevant acts are *guaranteed* to have a negative expected utility. For instance, he writes:

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.⁷⁸

This cannot be right; there is no guarantee that the expected utility will come out negative in every triggering case. Whether it does or not depends on the probabilities and on the goodness and badness of the relevant consequences. In the chicken case in particular, Kagan is right that the numbers *as he has them*, entail a negative expected utility for purchasing a single chicken. However, the numbers clearly need not be as he imagines them to be. And it is only because he has made the numbers match up – for some N , N purchases causes N additional chickens to be hatched and raised *who wouldn't otherwise have been* – that he gets the results he wants: automatic negative expected utility.⁷⁹ There is no reason to think that the numbers match up in this way. Kagan tries to justify the matching by saying, “since the butcher neither wants to fall behind demand nor end up with ever larger numbers of unsold rotting chickens, we know ... that the number of chickens he orders is more or less the

⁷⁸ Kagan, “Do I Make a Difference?”, p. 120, my emphasis. Here is another similar quote: “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.” (Ibid., p. 129.)

⁷⁹ If another N chickens are raised who wouldn't otherwise have been every time N chickens are sold, then your chance of making a difference to whether N chickens are raised is $1/N$, which means the expected disutility of your act is the equivalent of the suffering *one chicken* undergoes in being raised. By hypothesis, the suffering one chicken undergoes in being raised for slaughter outweighs the benefits gained from eating one chicken, and so this entails that the expected utility is negative.

same as the number of purchases required before a new order is triggered.”⁸⁰ But, first, it is not clear that the only way a store could both keep up with demand and avoid vastly exceeding it, is to have some number N such that when (and only when) N items are purchased, N more are ordered. Some stores may have more complex ordering strategies, in which decisions depend on a variety of factors – including upcoming marketing plans, the availability and cost of other similar products, news about one’s competitors, and a wide range of statistics about past sales. Moreover, Kagan needs that the *factory farm* will increase or decrease future production by N chickens depending on whether or not one particular butcher orders another N chickens. This seems highly unlikely. The factory farm most likely chugs along producing as much as it can, given the space and resources it has. If one particular butcher orders fewer chickens than usual, this might result in the distributor searching for a new client, rather than in a decrease in production. The factors that go into deciding whether to increase or decrease future production, and by how much, might include not only facts about how the distributors have done in their sales, but also the physical limitations of the space, the financial burdens of expanding and so on.

In short, what is at issue is how my single purchase might *affect* chicken production. We are asking about the chances my act has of *making a difference* to the number of chickens raised and killed. Whether my single purchase makes such a difference depends, ultimately, on whether it makes a difference to decisions to increase (or decrease) production by the factory farm. It seems to me that we have no reason to think that my purchase must have a $1/N$ chance of making a difference to whether the farm increases production by N chickens, for some N .

So, “maximize expected utility” does not automatically dictate that one ought not do the relevant act in triggering cases. This may not be considered problematic by the consequentialist. The consequentialist could say: “we need to be able to say something about how the badness of the consequences in these cases affects the morality of the individual acts involved; we do *not* need to obtain the intuitive verdicts. If the expected utility of buying a chicken, or of refraining from voting, comes out positive, consequentialism does not condemn your act and that is perfectly fine by me.” Of course, it’s not clear that the consequentialist *should* be satisfied with biting the bullet in this way.

⁸⁰ Ibid., p. 124.

3.4 The Appeal to Imperceptible Harms and Benefits

In imperceptible difference cases – unlike triggering cases – there is, it seems, no chance that one’s act will make a morally relevant difference. Yes, your act will make a difference along the underlying dimension (e.g. to the voltage of the electric current), but this difference is imperceptible and so, it seems, *cannot* leave anyone worse off. So expected utility does not help. Consider, for a new example, Kagan’s pollution case: Suppose that hundreds of thousands of factories release a toxin into the air and that if one inhales enough of this toxin, one will get a disease. But, while many people do inhale enough to get this disease, the toxin disperses so widely upon release that no one inhales more than a single molecule from any one factory. A single molecule is not enough to make a difference to one’s health: whatever one’s current state of health, a single molecule more will not be enough to give you the disease, and, if you already have the disease, a single molecule more will not be enough to make it worse. So, it seems no factory has even a chance at making a difference for the worse to anyone.⁸¹

But is this true? Kagan considers a denying-the-description reply that has been proposed by others in response to imperceptible difference cases: that the solution to the problem in such cases lies in accepting that there can be *imperceptible harms and benefits*. This reply is given by Parfit as an alternative to his appeal to group membership. In *Reasons and Persons*, he remains open about which view is the correct way to go, but later, in an unpublished paper, “What We Together Do”, Parfit abandons the appeal to group membership and endorses this other line of response. Jonathan Glover and Michael Otsuka also each appeal to imperceptible harms and benefits in their attempts to address the Problem of Collective Harm.⁸² As Kagan explains, this solution

simply insists that results need not be perceptible ... to count morally. But once we allow for the possibility of imperceptible harms (or benefits), we can insist that ... it simply isn’t true that my individual act makes no [morally relevant] difference... In the pollution case, for example, there is more toxin released as the 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

⁸¹ Ibid, pp. 108 – 109.

⁸² See Parfit, *Reasons and Persons*, section 28, and Parfit, “What We Together Do.” See also Jonathan Glover, “It Makes No Difference Whether or Not I Do It,” and Michael Otsuka, “The Paradox of Group Beneficence.”

imperceptibly worse off. Such an 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.⁸³

One obvious issue with this reply is that there aren't always thousands of victims in Collective Harm Cases. Sometimes there are just a few victims, or even only one, as is the case in our version of *Harmless Torturers*. Is causing an imperceptibly tiny harm to *one* victim enough to render one's act wrong? What if one obtains a perceptible benefit from doing it (e.g., each torturer gets a free massage)? In that case, we would have an imperceptibly tiny harm to one individual being weighed against a perceptible benefit to one individual. To the extent that Kagan adopts the imperceptible harm proposal, Kagan needs to address the question of whether it is effective in cases with small numbers of victims.⁸⁴

To what extent does Kagan adopt this proposal? Kagan thinks that it may work for some imperceptible differences cases, but that it cannot work for all:

[I]mperceptible differences won't pose a particular problem for the consequentialist as long as they constitute real harms, even if imperceptible ones... But ... 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 ... 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 harm in and of itself.⁸⁵

Kagan's point is that the appeal to imperceptible differences can only work if the imperceptible difference in question constitutes a *real* harm. The consequentialist cannot simply insist that there is a harm; there needs to actually be something harmful about it. In some cases it makes sense to say

⁸³ Kagan, "Do I Make a Difference?", pp. 114 – 115.

⁸⁴ Perhaps a reply to this would be similar to the "biting the bullet" reply I mentioned with respect to the expected utility solution for triggering cases: we do not need to get the intuitive verdict.

⁸⁵ *Ibid.*, p. 129.

that the harm can come in imperceptibly small amounts, and so – Kagan thinks – in those cases it makes sense to say each individual causes an imperceptible harm. But in cases in which what is bad about the outcome is *only* qualitative aspects of experience, Kagan does not think it makes sense to talk about imperceptible harms, and so in such cases the appeal to imperceptible harms cannot work.

The result is that Kagan takes the hard cases – the ones that remain a real challenge for the consequentialist – to be the imperceptible difference cases in which all that matters morally are qualitative aspects of experience. I will call these cases the “phenomenal” cases, and the imperceptible difference cases that Kagan dismisses as not a challenge the “non-phenomenal” cases. That is, by “non-phenomenal cases” I mean cases in which it is not *only* qualitative aspects of experience that matter morally.

3.5 Non-Phenomenal Imperceptible Difference Cases

It seems to me that Kagan too quickly dismisses the non-phenomenal cases. What we are given in imperceptible difference cases is that each act makes an imperceptible difference along the *underlying dimension*. Even if one has a case in which there could be imperceptible differences in harm, this does not mean that the difference I make along the underlying dimension is itself a difference in harm. Take the pollution case. Suppose we think that there is something bad about illness beyond and apart from the pain and suffering involved. In that case, there may be differences in health that, while imperceptible, are genuine differences for the worse (e.g. a difference in the size of a cancerous growth). If this is so, then this counts as a case in which differences in harm may come in imperceptibly small quantities. But it does not follow from this that a single factory is making any such difference. Each of these factories is, by assumption, making a difference to the number of toxic molecules inhaled: some people inhale a single molecule more as a result of a factory’s activities. But all we have said is that changes for the worse in *health* can be morally relevant even if only imperceptible, and we have been given no reason to think that inhaling a single molecule makes any difference in health, imperceptible or perceptible.⁸⁶

Let’s look at another example. Consider a consequentialist view that treats *fairness* as a morally relevant aspect of outcomes: being unfair is something bad about an outcome, and something that is *in itself* bad. Now, imagine that there

⁸⁶ And just as Kagan did not think it plausible that the increase in voltage is harmful in and of itself, it does not seem plausible that the increase in the number of molecules inhaled is harmful in and of itself.

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. Now, suppose the committee does the following: they initially divide the water into two approximately even amounts, but then one by one each committee member takes a single drop of water from the supply meant for A and moves it to the supply meant for B.⁸⁷ A single drop is simply not enough to make a difference to whether or not the distribution is fair. If the distribution is fair, moving one drop will not make it unfair (or even, slightly less fair). So, it seems, no single act of drop moving makes a difference for the worse. But eventually, after many thousands of such acts (this is a very large committee), there is a substantial amount more in B's supply than in A's, and so the distribution is unfair. So, we have an imperceptible difference case: while each act does make an imperceptible difference along the underlying dimension (the quantity of water that goes to each community), no individual act makes a difference to this morally relevant aspect of the outcome of whether the distribution is fair. Now, the point I am making about Kagan's dismissal of non-phenomenal cases is that it does not solve the problem to simply point out that there can be imperceptible differences in fairness that count as differences for the worse. While this may be true (perhaps a situation could be unnoticeably unfair, and perhaps this would still be a bad thing), this does not mean that a single drop of water *makes* such a difference. A single drop of water may not make any difference at all in the fairness of the distribution – neither an imperceptible nor a perceptible one. Fairness, on our proposal, is simply not sensitive to such tiny amounts, even though there could be some *other* difference in fairness that is real even though it is unnoticeable to those involved.

The point is that even if we grant that in a certain case there could be imperceptible differences that constitute genuine harms, this does not mean that just *any* imperceptible difference along the underlying dimension constitutes a genuine harm. So it seems Kagan simply does not show that in the non-phenomenal cases each act makes a difference in harm. I do not think, therefore, that we can rule these cases out as unproblematic.

⁸⁷ Assume that there is no coordination between them. Each is acting independently. We can also suppose there is some small motivation to move a single drop of water.

3.6 Revising the Distinction: Triggering versus Non-Triggering Cases

What might have led Kagan to overlook this point? The answer, I now want to suggest, is that the way he originally divides up the Collective Harm Cases is misleading. Specifically, the “imperceptible difference” side of the distinction does not get things right: it does not capture what is distinctive about non-triggering cases and it misplaces the issue by focusing on imperceptibility.⁸⁸

First, contrary to what Kagan’s distinction suggests, in *all* Collective Harm Cases each relevant act makes a difference along the underlying dimension. This is not something distinctive of “imperceptible difference” cases. Just as in *Harmless Torturers* each act of turning the dial makes a difference to the voltage (the underlying dimension), but not, it seems, to the pain (the morally relevant outcome), in the chicken case each act of buying a chicken makes a difference to the number removed from the butcher’s supply (the underlying dimension), but not, it seems, to the suffering of any chickens (the morally relevant outcome). And in the case of voting, each act makes a difference to the number of votes (the underlying dimension), but not, it seems, to who wins the election (the morally relevant outcome). So, it is misleading to say – as Kagan does – that in triggering cases each act most likely makes no difference, while in imperceptible difference cases each act makes a real but imperceptible difference along the underlying dimension.⁸⁹ It suggests that there is something important, or of special interest, in the fact that one makes at least *some* difference in the non-triggering cases, when really this is just an uninformative point that is true of all Collective Harm Cases.⁹⁰

⁸⁸ While I will put my discussion in this section in terms of Kagan’s distinction, the issues that I will raise show up in many other places in the literature.

⁸⁹ When he says this, Kagan is shifting the way he is talking about difference-making depending on whether he is talking about triggering or imperceptible difference cases. When Kagan says that in triggering cases each person most likely makes no difference at all, he is talking about a difference in *morally relevant outcome*. When he says that in imperceptible difference cases it is not true that each act makes no difference, and that instead in such cases each act makes an imperceptible difference, he is just talking about making a difference to *something* – the relevant underlying dimension – and not about the morally relevant outcome. This shift results in a poor characterization of what is distinctive about the two varieties of Collective Harm Cases.

⁹⁰ It is uninformative because *of course* a difference to the relevant underlying dimension is made by each of the acts under consideration: that is why these acts are being considered in the first place.

A second problem is that the characterization of the non-triggering variety of cases as ones in which one makes no morally relevant difference *because* the difference one makes is imperceptible is overly restrictive. Asserting that the barrier must be imperceptibility, no doubt, gives the impression that the hard non-triggering cases are going to be those in which the morally relevant aspects are purely experiential. Those are the ones in which imperceptibility *would* stand in the way. But the mistake is thinking that it is imperceptibility that must be the problem. As Kagan points out when he says that some harms may come in imperceptible quantities, making a morally relevant difference and making a perceptible difference are not the same thing. The problem in the water distribution case, for instance, is that a single drop is simply not enough to make any difference in *fairness*. If this is right, the water distribution case is not a triggering case, but this has nothing in particular to do with perceptibility.

Related to this last point, defining the non-triggering side in terms of imperceptibility excludes a potential variety of cases: those which do not have a triggering structure but in which one makes a *perceptible* difference along the underlying dimension. Just as not all imperceptible differences are morally relevant (even if some are), not all perceptible differences are morally relevant.⁹¹ For example, at one point Kagan discusses a case of overfishing: if every angler in the village takes one more fish than she is allotted, this will result in the fish population losing its ability to replenish itself; but, for each angler, taking a single fish more is not itself enough to make a difference. Kagan suggests that this is a triggering case.⁹² But I think that had Kagan not cashed out the non-triggering side of the distinction in terms of imperceptibility, it would have been more clearly open that this might be a non-triggering case. What would it take for it to be a non-triggering case? Instead of the removal of some particular fish *triggering* the problem, it could be that there is no precise triggering point – no sharp boundary between the population having a healthy ability to replenish and not having it. If this were so, it would not be a triggering case, but this would have nothing to do with imperceptible differences: one more fish removed from the lake is a *perceptible*

⁹¹ Even in *Harmless Torturers*, making the problem out to be that you do not make a perceptible difference is inaccurate. The problem is that you do not make a difference in *pain* – one specific kind of perceptible difference, which is the morally relevant outcome in this case – and not that you do not make *any* sort of perceptible difference. The difference an act makes could be perceptible in various other ways. Perhaps there is a meter that displays the voltage, so that the victim sees a difference between adjacent settings. A difference that one sees is a perceptible difference. It is just not a difference in harm, since – we can suppose – seeing the different numbers has no effect on the victim’s pain state.

⁹² *Ibid.*, p. 118.

difference along the underlying dimension (the angler, after all, perceives the difference when she counts her catch) and it would be *no difference* (neither perceptible nor imperceptible) along the morally relevant dimension in question.

Rather than contrasting triggering cases with “imperceptible difference cases”, Kagan should simply divide the cases into triggering and non-triggering cases.⁹³ What is distinctive about non-triggering cases is that no single act serves as a trigger. So, the structure is not that of a tipping point: there is no precise point at which a limit is hit and the next act triggers a change in morally relevant outcome. Instead, the boundaries between one morally relevant outcome and another are vague, and so the difference between $n-1$ and n acts of the relevant type can never, no matter what n is, make the difference between one morally relevant outcome and another.

The relationship between the harm and individual acts of the relevant type in non-triggering cases is the same as the relationship between a vague property and the underlying changing phenomena in familiar sorites cases – for example, the relationship between being bald and the number of hairs you have. *Presumably*,⁹⁴ if you have enough hairs to count as not bald, then you have more than enough hairs – taking one away won’t make you bald (nor, will it make you partially bald or “balding”). And if you count as bald, gaining one more hair won’t push you over the edge to being not bald (or less bald). Similarly, in the non-triggering *Harmless Torturers* case, for example, if enough people have turned the dial for the victim to be in a certain amount of pain then more than enough have. And if not enough have turned the dial, my turning it a single setting will not push him over the edge.

Putting Kagan’s project in terms of this revised distinction, his goal – after having argued that consequentialism can handle triggering cases – is to show that there cannot actually be any non-triggering cases. If Kagan were

⁹³ This is consistent with there being some cases that are a mix of the two pure varieties. A case could have a triggering structure with respect to one morally relevant aspect of the outcome, but a non-triggering structure with respect to a different morally relevant aspect.

⁹⁴ I say “presumably” because this is the intuitive, pre-theoretic view, and many – though not all – accounts of vagueness take this to be evident. Some theories of vagueness deny that there can exist vague boundaries of this sort. In particular, epistemicism denies that there can be any such vague boundaries. (Note, though, that it is not accurate to say that epistemicism denies the existence of the phenomenon of *vagueness*, or of *vague predicates*. According to the epistemicist, while there *is* a sharp boundary between being bald and not being bald, this boundary is *unknowable* to us, and it is this *epistemic* condition that makes “bald” a vague predicate.) Epistemicists would – as Kagan wants to – deny the possibility of non-triggering cases. But epistemicism is highly controversial, and we should begin our evaluation of Kagan’s arguments – as he begins his arguments – from the intuitive, pre-theoretic perspective in which non-triggering cases *seem* to be a real possibility.

successful in showing that in the non-phenomenal cases each individual does an imperceptible amount of harm, this would mean that those were triggering cases: triggering cases in which *every* act triggers a change, but just a very small one, in morally relevant outcome. But, as I have said, I do not think he succeeds in showing this. Thus, it remains open that those cases are genuine non-triggering cases. Let's now turn to Kagan's argument that there cannot exist any non-triggering cases of the purely phenomenal variety.

3.7 Kagan's Harmless Torturers Argument and the Reports Premise

Kagan's argument that there cannot be any phenomenal non-triggering cases is given in terms of *Harmless Torturers* (from now on "HT"). The idea is that the same line of argument could be run for any purported phenomenal case. The argument is a reductio on the claim that there could be a perceptible difference in pain sensation between far apart settings but no perceptible difference for any two adjacent settings. Here it is, written out step-by-step:

1. Suppose (for reductio) that for each n , states s_n and s_{n+1} feel exactly the same to the victim. There is no perceptible difference in pain sensation between any two adjacent settings.
2. By hypothesis, the victim is in no pain at s_0 and if asked, the victim will say he is in no pain.
3. By 1, since s_0 and s_1 are adjacent states, there is no perceptible difference between them.
4. (Premise) If there is no perceptible difference in sensation between two states, then the victim will have the same answer to "are you in pain?" at one of those states as he does at the other. If he has a different answer, there must be a perceptible difference between them.
5. Therefore, by 3 and 4, his answer to "are you in pain?" must be the same at s_1 as it is at s_0 .
6. So, by 2 and 5, he must answer "no" when asked "are you in pain?" at s_1 .
7. Since s_2 is adjacent to s_1 , by 1, there is no perceptible difference between them.
8. Therefore, by 4 and 7, his answer to "are you in pain?" must be the same at s_2 as it is at s_1 .
9. Therefore, by 6 and 8, his answer at s_2 must be "no" as well.
- ⋮
3002. When asked "are you in pain?" at s_{1000} , the victim's answer must be "no".

3003. But in s_{1000} he is in excruciating pain, and, if asked, he would say he is in excruciating pain.

3004. Contradiction! (Between 3002 and 3003).

3005. Thus, I must be false.⁹⁵

As you can see, the argument relies on the truth of the premise – in (4) – that a difference in the victim’s pain reports implies a perceptible difference between the two states (call this the “Reports Premise”). But is this premise true?

Surely the victim’s reports need not be perfectly accurate or precise. So, why should we accept that differences in the victim’s pain reports are perfect indicators of differences in his sensations? Why should we take the contradiction to imply that the assumption in (1) is false, rather than the assumption in (4)? The only thing Kagan says in the way of support for the Reports Premise is that the victim’s reports are immediate and spontaneous. Kagan writes:

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 perceptible difference in the amount of pain.⁹⁶

Similarly, he later asserts, “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.”⁹⁷ I do not see why we should think that immediate and spontaneous reports will always be accurate in this way. Consider times in which you have been asked to report on experiences that may or may not contain subtle differences. When the eye doctor has asked me to say whether one of two images is clearer than the other, I answer in an immediate and spontaneous fashion. But it has never seemed to me that my answers are perfect indicators of whether or not the two images are equally clear. I would not be surprised if, even though I said that the first image seemed clearer than the second, the two were actually indistinguishable to me. I often think it

⁹⁵ Kagan gives this argument (in paragraph form) on page 132.

⁹⁶ Ibid., p. 136.

⁹⁷ Ibid., p. 139.

would be better if somehow the eye doctor could directly attend to what I am seeing, rather than relying on my ability to report what I am seeing accurately.

It seems to me that, contrary to what Kagan suggests, immediate and spontaneous reports of one's perceptual states may sometimes be only roughly right and may, at other times, be quite inaccurate. Yes, you are simply reporting how things are to you, but – no matter how immediately and spontaneously you do so – it can be difficult to put one's finger precisely on what one is experiencing. So, especially when it comes to questions regarding subtle differences in experiential states, *or the lack thereof*, I do not see how such reports can be taken as infallible indicators.⁹⁸

Kagan himself makes the point that we are not infallible with respect to comparisons of our own sensations. He says that, even though he has shown that there must be a difference in pain between some adjacent states,

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 ... 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.⁹⁹

On what basis can Kagan both (i) accept that pairwise comparisons made by the victim may be mistaken, and (ii) think that a difference in answer to the question “are you in pain?” *cannot* be due to fallibility or inaccuracy?¹⁰⁰

⁹⁸ Kagan says that his claim about differences in reports being perfect indicators is not an “empirical remark” but rather a purely “conceptual observation” about “the concept of an observational report” (Ibid., p. 139). But this seems implausible. I see no reason to think it follows simply from the concept of an observational report that *if* such reports are immediate and spontaneous, then a difference in them always correctly indicates a difference in the experiences themselves. Kagan says nothing to explain his claim that it does.

⁹⁹ Ibid., p. 136.

¹⁰⁰ Kagan seems to think that reports of two states feeling the *same* are not reliable but reports of two states feeling *differently* are. But, if this is his claim, why should we accept this asymmetry? Justification would surely be needed. As I said, there seem to be plausible counterexamples to the suggestion that reports of differences are always accurate – such as the example given above of reporting to the eye doctor that one image is clearer than another when in fact there were no relevant differences between the two visual experiences.

3.8 Perceptible Difference versus Worsening of Pain

Let's put aside the issue with Kagan's Reports Premise for a moment and turn to a different question. The direct conclusion of the argument is that there must be perceptible differences between some adjacent settings. Is this enough for Kagan's purposes? Specifically, does this necessarily mean that there is a morally relevant difference in pain between some adjacent settings? Kagan needs it to be the case that for some adjacent settings the victim not only feels *differently*, but *feels worse* at one of the settings. Even if Kagan's argument is successful in showing that it cannot be that each setting s_i *feels the same as* the one adjacent to it s_{i+1} , does this mean that the victim is in a *worse* pain state at some s_{i+1} than he is at s_i ? Not necessarily. In a recent article on phenomenal sorites series, Rosanna Keefe explains that, *even if* one could prove that *looks the same as* is transitive (which she disputes), this is not enough to show that there can be no sorites series involving *looks red*. She considers a series of patches (a_i) created by starting with a large bucket of red paint and mixing a single drop of yellow paint into the bucket each time before painting the next patch. She writes:

We might say of a pair 'I don't know whether they look (exactly) the same, but they are so similar that it is certainly the case that if one looks red, so does the other.' We [can] ... accept the successive conditionals corresponding to instances of the premise [if a_i looks red then a_{i+1} looks red] without first confirming that the relevant pairs look the same... Denying the non-transitivity of *looks the same as* will thus not help to solve this compelling sorites paradox on 'looks red'.¹⁰¹

Similarly, in the HT series in which the voltage is only increased by a minuscule amount each setting, it could be that, *even if* Kagan is right that we cannot assert that s_i feels the same as s_{i+1} , for all i , it still may be that they are, nonetheless, so similar that if s_i feels like no pain then so does s_{i+1} .¹⁰² This would be enough to make it a non-triggering case. So, the idea is that a perceptible difference between sensations is not necessarily equivalent to one sensation being *a worse state of pain* than the other. Kagan needs the latter, but all his argument explicitly tries to show is the former. I will, for the sake of argument going forward, assume Kagan is right that these are equivalent. So, I will follow Kagan and move freely between claims about perceptible

¹⁰¹ Rosanna Keefe, "Phenomenal Sorites Paradoxes and Looking the Same," pp. 331-332.

¹⁰² This relates to the point made earlier that not all perceptible differences are differences in morally relevant aspects.

differences and claims about differences for the worse in pain state. But the point is that, for all he has said, these may not be equivalent, and therefore there is a gap in Kagan's argument that still needs to be filled.

The only place Kagan mentions a connection between phenomenal non-triggering cases and sorites series is when he says in a footnote that his argument is not "meant to provide a response to sorites style arguments." He writes:

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.¹⁰³

This is misleading. The purported non-triggering series he is trying to dismiss as impossible *is* a sorites series – just one of the *phenomenal* variety.¹⁰⁴ The claim Kagan is trying to refute is that a single minuscule increase in voltage cannot make the difference between *non-pain* and *pain*, and in general between *one pain state* and *a worse one*. This is just like the claim that a single grain of sand cannot make the difference between a *non-heap* and a *heap*. It falls into the subclass of sorites cases known as "phenomenal sorites cases", which include – for example – series from *does not sound loud* to *sounds loud*, from *looks red* to *looks orange*, from *feels heavy* to *does not feel heavy*. Arguing, as Kagan is trying to, that there must be some precise point at which the victim goes from feeling no pain to feeling some pain (or, from one pain state to a worse one) amounts to arguing for a certain solution to the phenomenal sorites series: namely, denying that this kind of sorites series can actually exist. This is not itself an objection to Kagan's argument.¹⁰⁵ But it is important that Kagan recognize that he is entering into the debate about phenomenal sorites series, and that the relationship between the sorites paradox and his argument is not as weak

¹⁰³ Kagan, "Do I Make a Difference?" p. 134, n. 13.

¹⁰⁴ In general, as I said above, non-triggering cases involve vague boundaries.

¹⁰⁵ Note that Kagan is in good company in trying to argue for this conclusion. There are other philosophers who try to argue that there is a solution available to purported phenomenal sorites series that is not available in non-phenomenal cases, where this solution amounts to denying their possibility. For instance, Delia Graff Fara argues for such a solution in her article "Phenomenal Continua and the Sorites". Keefe's article, on the other hand, argues against such solutions and asserts that "we should not expect a different solution [in phenomenal cases] from that of other sorites." (Keefe, "Phenomenal Sorites Paradoxes and Looking the Same," p. 331.)

as he makes it out to be. In 3.9, we will see that it is because he is dealing with a purported sorites series (a purported case of vague boundaries) that the Reports Premise plays an indispensable role in his argument. In 3.10 that we will see that, even were the Reports Premise true, it may not actually be capable of playing the role he needs it to play.

3.9 The Role of the Reports Premise

Returning, then, to the objection that Kagan's Reports Premise does not seem to be true, it is worth asking whether Kagan could do away with it. Maybe Kagan's use of reports *unnecessarily* weakens his argument. What would Kagan's argument look like if he removed reference to reports, and instead ran the argument directly in terms of the victim's pain sensations? It would go something like this:

1. Suppose (for a reductio) that there is no perceptible difference between any two adjacent settings. So, for any i , if one is in no pain at s_i , then one is in no pain at s_{i+1} .
2. By hypothesis, no pain is felt at s_0 .
3. Therefore, by 1 and 2, no pain is felt at s_1 .
4. Therefore, by 1 and 3, no pain is felt at s_2 .
- ⋮
1002. Therefore, by 1 and 1001, no pain is felt at s_{1000} .
1003. But by hypothesis, a huge amount of pain is felt at s_{1000} .
1004. Contradiction!
1005. Thus, 1 must be false.¹⁰⁶

Let's call this argument "Version 2." The problem with Version 2 is that it amounts to giving a sorites argument as though it were a simple reductio proof that there cannot be vague boundaries. So, Version 2 is no better than arguing as follows:

¹⁰⁶ Kagan does at moments sound as though he is suggesting this line of argument. For example, after giving his argument he writes: "Once stated, the point seems obvious. 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." (Kagan, "What Difference Do I Make?", p. 132.) This sounds like Version 2.

1. Suppose that no single hair makes the difference between being bald and not being bald. So, if we let a_n be the state of having exactly n hairs, then, for any i , if one is bald at a_i then one is bald at a_{i+1} .
2. One is bald at a_0 .
3. Therefore, by 1 and 2, one is bald at a_1 .
4. Therefore, by 1 and 3, one is bald at a_2 .
- ⋮
- 100,002. Therefore, by 1 and 100,001, one is bald at $a_{100,000}$.
- 100,003. But one is not bald at $a_{100,000}$.
- 100,004. Contradiction!
- 100,005. Therefore 1 must be false.

If one thought that in giving this argument one had established its conclusion one would be missing the fact that the sorites argument presents a paradox that needs to be worked out and for which several different types of solutions may be attempted, not a simple reductio of one of the premises.¹⁰⁷ It may turn out that the right answer to the sorites paradox is to reject the claim that there is no precise number of hairs at which you first qualify as not bald (i.e. reject the assumption in 1). That is what epistemicists argue. But this is by no means obvious or uncontroversial, and it cannot be *established* by simply giving the sorites argument.

A different way of putting this problem with Version 2 is that it would be *question begging*. An argument following the same sort of reasoning could be perfectly acceptable (though probably rather uninformative) *if* it were working with a phenomenon for which we know there are no vague boundaries. But HT seems to be a non-triggering case: it seems that for any two adjacent states, neither is going to feel worse for the victim. This, as I have said, *is* a claim about vague boundaries. It claims that there are no precise points in the series at which the victim goes from one pain state to a more severe one. If there were a precise point at which the victim first felt pain, say, this would – of course – be a specific point in the series at which the victim felt worse than

¹⁰⁷ In a paradoxical argument, highly intuitive, apparently true premises are shown to lead to contradiction through a series of seemingly valid steps. A reductio argument, on other hand, is an argument which has the form of supposing the opposite of its conclusion in order to generate a contradiction; by showing that the supposition of p leads to contradiction, one proves that $\sim p$ must be the case. A paradoxical argument sets up a problem for which there is no immediate or clear solution; a reductio argument attempts to prove a conclusion through a particular type of strategy. One cannot simply treat a paradoxical argument as a reductio of one of its premises, at least not without *first* arguing that other potential solutions to the paradox are untenable and that rejecting the premise is the best option of those available.

he did at the setting immediately prior. But HT seems not to have any such sharp boundaries. Kagan's goal is to show that this is not actually an accurate understanding of the nature of *HT*. So, were Kagan to give the argument in Version 2 as though it were a reductio proof he would be begging the question. His argument would simply take for granted that there cannot be any vague boundaries between the pain states, because if one were leaving it open that the boundaries between one pain state and the next could be vague, it would be clear that giving such an argument as one's proof would not be fair game.

This is why it is important that Kagan runs the argument in terms of reports. His use of reports is not an unnecessary complication to what would otherwise be a strong argument. Rather, by appealing to reports, Kagan seems to be avoiding presupposing exactly what he wants to prove: that there cannot be vague boundaries between the victim's pain states.¹⁰⁸

Now, if we thought that *reports* could have vague boundaries, Kagan's strategy would not work: he would be making the same mistake that Version 2 makes, but with respect to reports and in a slightly more convoluted way. But it seems we can all agree that there are no vague boundaries between one report and another. This is why running the argument in terms of reports appears to be a good idea (or, at least, it would be a good idea if the Reports Premise were true). It is not clear that there must be a precise point at which the victim first feels pain or at which the victim's pain first gets worse, and it would be begging the question to argue in a way that assumes that there is. But it does not seem similarly controversial to assume that there must be a precise point at which the victim's answer first changes from a clear "no pain" to something else.

3.10 Reports and Vague Boundaries

But is this actually the case? I now want to suggest that it is not in fact obvious that the reports in HT have sharp boundaries.¹⁰⁹ It is surely right that there will have to be some first point at which the words the victim says in answer to the question "are you in pain?" must be different than just "no."

¹⁰⁸ I do not mean to suggest that this is what Kagan would say if asked "can you just run your argument directly in terms of pain sensations rather than by invoking reports and if not, why?" But whether or not he would say this, it seems quite clear that it is the apparent indisputability of the claim that reports cannot have vague boundaries that is behind the appeal of putting the argument in terms of reports.

¹⁰⁹ I am indebted to Niko Kolodny for raising this question of whether the boundaries between reports might be vague.

That is, it is clear that there are no vague boundaries between saying one phrase and saying another. But this is not what Kagan means when he says “differences in reports.” When Kagan says that a difference in reports implies a difference in the experiences themselves, he does not and cannot mean by “reports” simply the *words* coming out of the victim’s mouth. Suppose he did mean this, and imagine that when asked “are you in pain?” the victim says “not at all” at s_0 and “not in the slightest” at s_1 . Kagan would have to say that this was a difference in reports between s_0 and s_1 and so he would have to conclude that there must be a perceptible difference in feeling between these two states. The Reports Premise on this reading of “reports” would be *obviously* false. The words I use to describe my perceptual experience in two states can certainly be different without there being a perceptible difference between them: I could *mean* the same thing by my words. What Kagan means by “a difference in reports implies a perceptible difference” is that a difference in what the victim is *reporting* – that is, in the *content* of the victim’s answer – implies a perceptible difference in the experiences. Indeed, Kagan makes this clear when he says “if our reactions are immediate and spontaneous reports of our own experience...then differences in the *content of those reports* indicate differences in the experiences themselves.”¹¹⁰

Once we make this distinction between the words used in reporting and the content of the reports and realize that what Kagan means is the *content*, it no longer seems at all obvious that there are no vague boundaries. Kagan’s description of how the reports might go in HT helps bring this out. Kagan explains that what the victim says would likely progress extremely gradually:

I...do not mean to suggest that at 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)...What we might expect, instead, is a slow hedging of characterization. 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...The difference between the ways that two adjacent states are characterized will presumably be quite slight, befitting the fact that

¹¹⁰ Kagan, “What Difference Do I Make?”, p. 139, my emphasis.

the difference in the level of pain, although perceptible, is extremely slight.¹¹¹

If the difference in what the victim *says* (the words, the manner) in giving his reports for two adjacent settings is only very tiny, it is not at all clear whether we can assert that the victim is *reporting* a different pain state for those settings. If what he says progresses very gradually, perhaps there are no sharp boundaries between his reporting one amount of pain and reporting another.

It seems to be taken for granted (even if only implicitly) both in Kagan's argument and in similar arguments elsewhere,¹¹² that no matter how gradual the change is in what the victim says, the victim's reports cannot have vague boundaries: what he *says* (the words) must be different for some adjacent settings, so this means that the victim must be giving a different report. But the suggestion here is that just because the words the victim is using cannot be the same for every pair of adjacent settings, this doesn't mean that there are sharp boundaries between his reporting one thing and his reporting something else. If the change in words is so gradual that the phrases he uses are only slightly distinguishable for any pair of adjacent states, this may be a case in which reports involve vague boundaries. If this is so, then *even if* the Reports Premise were true, it could not play the role that Kagan needs it to play in his argument.

Perhaps Kagan could avoid this issue by fixing things so that gradual changes in the victim's reports are not allowed: perhaps the victim must answer either "yes" or "no". But if Kagan does this, it becomes even harder to see how he could make any case for the Reports Premise. So, it seems Kagan's options are: either allow extremely gradual, just barely distinguishable changes in what the victim says, which opens up the objection "is it really true that the boundaries between his reports cannot be vague?", *or* do not allow the changes in what the victim says to be subtle and gradual, which makes it even less plausible that differences in reports are perfect indicators of what is going on in the experiences themselves.

3.11 Quasi-Triggering Cases

Kagan's argument attempts to prove that for any run through the *HT* series there must be some *i* such that the victim feels worse at s_i than s_{i-1} . For

¹¹¹ Ibid., pp. 134 – 135.

¹¹² See, for example, Frank Arntzenius and David McCarthy, "Self Torture and Group Beneficence." See also Ken Binmore and Alex Voorhoeve, "Transitivity, the Sorites Paradox, and Similarity-Based Decision Making."

the reasons already discussed, I do not think Kagan succeeds in establishing this. But let us *suppose*, for the sake of argument, that Kagan's argument for this conclusion is successful. Would this be enough to show that HT must be a triggering case?

If HT is a triggering case, then for each morally relevant change in outcome, an individual act of dial turning *triggers* the change. If this is so, then, as long as it is possible for the torturers to collectively stop turning the dial exactly at a triggering point, it is *possible* that each torturer's turn will make a difference (though likely only a tiny one) for the worse.¹¹³ But is showing that there is some i such that the victim feels worse at s_i than he did at s_{i-1} enough to show that some act of turning the dial triggered the change for the worse? For this, Kagan would need to have shown that the *reason* a change in pain sensation occurs at s_i , rather than – say – at s_{i-1} or s_{i+1} , is that the dial was turned up a single setting further.

Kagan says nothing explicit about this, but Alex Voorhoeve and Ken Binmore – who give an argument fairly similar to Kagan's – are more explicit about this move. After arguing that there must be particular points in a run through the series at which the pain sensations get worse, they draw the following conclusion: “we must reject the popular idea that there is a ‘least-noticeable difference’: a magnitude of physical change so small that human beings always fail to detect a difference between situations in which a change smaller than this magnitude has and hasn't occurred.”¹¹⁴ They think that if – for example – a victim wasn't in what he would say was very bad pain at s_{823} but is at s_{824} , it must be that the minuscule change in voltage triggered a change in how the victim feels, and that thus the victim is *detecting* the difference between s_{823} and s_{824} . It must be that the voltage at s_{823} was not enough (in this particular run through the series) for the victim to feel very bad pain, but the voltage at s_{824} was enough, and thus the difference in voltage is what made the difference to the victim. They illustrate their point with this analogy:

Consider a device that registers charges only in whole kilovolts. If we hooked this device to a machine that administered a current of varying voltage, started with the dial at 0 and kept increasing the charge by small increments [let's say, one nanovolt at a time], then at some point

¹¹³ If, for example, in a given run through the series, the turn to s_{120} triggers the first feeling of pain, then, were no other torturer to act after the turn to s_{120} , each person who did act would have made a difference: had that person not turned the dial, it would have stopped at s_{119} and so the feeling of pain would not have been triggered.

¹¹⁴ Binmore and Voorhoeve, “Transitivity, the Sorites Paradox, and Similarity-Based Decision Making.” p. 104.

the device will change from registering ‘0 kV’ to registering ‘1 kV’. This implies that there are no just-noticeable differences in the sense under discussion for this device, even though its capacities of discrimination are limited.¹¹⁵

I think their reasoning here is at least partially mistaken. It is a mistake to take the fact that at some point the device registers 1 kV, when it didn’t immediately before, to imply that it was the last tiny nanovolt increase – or any single nanovolt change – that made the difference. For any measuring or sensing device in the world there is a *margin of error*. If the quantity being measured is within the device’s margin of error of some value p then whether or not the device registers it as p will depend on accidental internal or environmental factors. Imagine that you are working with this machine that registers charges only in whole kilovolts, increasing the current applied to it nanovolt by nanovolt. Eventually the current will be within the margin of error of a kilovolt. So, the machine could change from registering 0 kV to registering 1 kV at any moment. But, given that you are within the margin of error of 1 kV for that device, it would be a mistake to think that, at the moment when it actually does register 1 kV, this is due to the last minuscule increase in voltage that you made. It *is* due to the fact that many increases were made, such that the current is in some rough, very close vicinity of 1 kV. If the current had not been within the margin of error the machine would not have registered it as 1 kV. But that it registered 1 kV at the precise point in your adding nanovolts that it did is most likely due to mechanical or environmental factors and not to the addition of some single nanovolt. This means, I think, that we cannot say that *had* you not added that last nanovolt, the machine would not have registered 1 kV.¹¹⁶

The same goes in the HT case. Even if we suppose that Kagan is right that there are precise spots at which the victim’s pain gets worse, it would be a mistake to assume that the precise location of these changes is due to the fact that some precise voltage was reached on the torture machine and not a minuscule amount less. We presumably have something like the margin of error that a measuring device has: within a rough, but close, range of voltages I could “register” any of some close variety of pain sensations. Which one I do register and where exactly it begins and where exactly it changes – *supposing*, for the sake of argument, that Kagan is right that there must be exact such points – may be due to any number of psychological, neurological or

¹¹⁵ Ibid., p. 104.

¹¹⁶ This idea was influenced by and relates to points found in Diana Raffman’s paper “Vagueness Without Paradox.”

environmental factors. Or, to use Diana Raffman’s words, “the exact location of the shift” from one sensation to another “is determined by brutally mechanical (‘sub-personal’) factors.”¹¹⁷

Kagan might reply that we can prove that the change in voltage is the trigger by simply thinking of how his argument would go if we ran through the series *counterfactually* rather than over time. Instead of running the argument through the Collective Harm Case itself, where torturers are taking turns increasing the setting over some period of time, “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?’”¹¹⁸ This way we can hold fixed everything *except* for the voltage. On this formulation, if Kagan is right that there must be some precise setting s_i at which the victim feels worse than he would have at s_{i-1} , it would *have* to be the minuscule increase in voltage at s_i that explains this difference, since – by stipulation – the increase in voltage is the only thing that is different between the two counterfactual scenarios.

It seems to me, though, that this at most shows that a turn of the dial may be among the potential triggers in the HT case. That is, I think it at most shows that it cannot be ruled out *a priori* as a possible trigger. In HT, we are dealing with people turning the dial over some period of time, and the question is whether, in this scenario, a single turn might make a difference between one morally relevant outcome and another. When we turn back to this problem, after having considered the counterfactual version of Kagan’s argument, it still remains the case – supposing Kagan is right that there are sharp boundaries between the pain states – that *other* factors (“brutely mechanical” changes in the brain, environmental changes, and so on) would be among the possible triggers. Indeed, it may be most likely that it would be one of these other factors that serve as the trigger, rather than a minuscule further change in voltage. So, even if the counterfactual argument shows that the turn of the dial could not be ruled out *a priori* as possibly a trigger, it still may be highly unlikely to be the trigger.¹¹⁹

What does this mean for Kagan’s argument? It suggests, for one thing, that it isn’t quite right that triggering and non-triggering cases (and combinations of the two) exhaust the possible ways Collective Harm Cases may be able to go. There may be a category between the two, which we could call “quasi-triggering” cases. In quasi-triggering cases, we cannot say – as we

¹¹⁷ Raffman, “Vagueness Without Paradox”, p. 56.

¹¹⁸ Kagan, “Do I Make a Difference?”, p. 133.

¹¹⁹ The sense of “possible” here is whichever notion of possibility comes into discussions of expected utility – plausibly, epistemic possibility.

can in triggering cases – that if there are enough acts to bring about a change in morally relevant outcome, then an act of the relevant type triggers the change. But we can say – unlike in non-triggering cases – that if there are enough acts to bring about a change in morally relevant outcome, then an act of the relevant type *could* be what triggers the change. If we accept Kagan’s argument, it seems the most we could conclude from it is that the *HT* is a quasi-triggering case.¹²⁰ (Of course, if we do not find his argument convincing, the possibility that *HT* is a non-triggering case is left open.)

If *HT* is a quasi-triggering case, would this be enough for the consequentialist to apply the expected utility solution? As I said, if *HT* is a quasi-triggering case, it may be most likely that other factors (neurological, psychological, environmental) are what actually trigger a change for the worse, once the voltage is within any particular rough vicinity. So, if *HT* is a quasi-triggering case, then each torturer may have only an *extremely tiny chance* of making a *barely perceptible difference* in pain to the victim. This chance of making a barely perceptible difference in pain would be the probability that all three of the following conditions hold: (1) a single increase in voltage (rather than some other factor) will serve as the trigger for the last change for the worse; (2) the torturers will stop turning the dial at exactly this triggering point; and (3) no other factor would trigger this change for the worse *were* that increase in voltage not to occur.¹²¹ This would, it seems, be a very remote possibility. Keeping in mind that this expected disutility will have to be balanced against whatever benefit the torturer will gain from participating, it seems very unlikely that the result will yield a negative expected utility overall. A very remote possibility that turning up the dial will make the difference between one pain state and a barely distinguishable pain state would surely be very easily outweighed in a calculation of expected utility. So, *even* granting Kagan his argument, it seems it would still only be enough to obtain this very weak conclusion for the consequentialist. Consequentialists would then be confronted with the following alternatives: either give an explanation of how the miniscule chance of a tiny difference for the worse could generate a

¹²⁰ I say “the *most* we could conclude from it” because perhaps we would not even be able to conclude that much. While perhaps a single turn of the dial cannot be ruled out a priori as possibly a trigger, it might be demonstrated scientifically that it would never, in the actual world, be the trigger. If this were so, we would still have that *HT* is a non-triggering case – a case in which a single act of the relevant type can never in fact make the difference between one morally relevant outcome and another.

¹²¹ Condition (3) would not hold if, were the last turn of the dial not performed, some environmental or internal factor would trigger the change instead. We can think of this as a “cross-factor,” rather than “cross-person,” version of “if I don’t do it, someone else will” (i.e., “if someone doesn’t do it (trigger the change), *something* else will”).

negative moral verdict, *or* bite the bullet and accept that they do not condemn the torturers' acts.

3.12 Summary

So, Kagan's argument that HT must be a triggering case does not succeed. For one thing, it relies on an implausible premise about the reliability of pain reports. Furthermore, even if the Reports Premise were true, this is not enough. It needs to be that the boundaries between the victim reporting one pain state and reporting another cannot be vague, which is not at all obvious. Finally, even supposing that, despite these objections, Kagan succeeds in establishing his argument's conclusion (that there must be perceptible differences between some adjacent states) this is still not enough to establish that HT must be a triggering case. First, a perceptible difference between some adjacent states need not be a *worsening of pain*. Second, even supposing that it is, this would at most allow us to obtain that HT is a quasi-triggering case. In that event, while there would be a chance of making at least a tiny difference for the worse, this chance may be only very slight and so unlikely to yield a negative expected utility on balance.

It is important to note that even had Kagan's argument in the HT case succeeded, this would not have helped with the non-phenomenal cases. This is because if Kagan's Reports Premise is true, it is true specifically because it is about experiential reports. Kagan does not think (and rightly so) that it is true of *all* reports that they are accurate in that way. Thus, the non-phenomenal cases remain problematic regardless of the verdict about Kagan's HT argument.

3.13 Denying the Description: the State of the Project

Where does this leave us with respect to the proposal that we can solve the Problem of Collective Harm by denying the description? Is it just that Kagan's arguments, and similar arguments given by others, do not succeed? Might it be that with a different argumentative strategy we could show that there is always at least a chance of making a difference?

Our discussion cast serious doubt on the plausibility of this project. We have seen that to argue that in all Collective Harm Cases there is at least a chance of one's individual act making a morally significant difference *is* to argue that there cannot be any cases that have a *non-triggering* structure. And what this requires showing is that there cannot be vague boundaries between morally significant outcomes. There is no way of doing this without entering

into the debate on the broader topic of vagueness and taking a highly controversial position within that debate.¹²² One would need to argue for a position like Epistemicism, the view that there are sharp but unknowable boundaries – a view that is highly contested – if one wanted to take this approach in both phenomenal cases and non-phenomenal cases. If our ability to make sense of the morality of Collective Harm Cases *depends* on being able to show that such a theory about vagueness is correct, this is not a good position to be in. It seems to me, therefore, that we should proceed on the assumption that there are genuine non-triggering cases.

In summary, we are not in a position to take a denying-the-description approach as a general solution to the Problem of Collective Harm. We cannot say that we are always making a mistake when we think that an individual act of the relevant sort won't make a difference in Collective Harm Cases.

A further thing to note is that even when we are in a triggering case, or a quasi-triggering case, and so even when we *can* say that there is at least some chance of an individual act of the relevant sort making a difference, this chance of making a difference might be only extremely slight. When this is so, it is not clear that we *should* take this to mean that we are making a mistake when we think that an individual act won't make a difference. That is, it is not clear that it really negates the claim that we are making when we say that an act won't make a difference to point out that there is a *highly remote* chance that it will make a difference. Take the case of a large national election. Replying to the “it makes no difference” argument with “that’s not true; your vote *might* make a difference because there might be a tie!” hardly suggests that we are wrong in our assessment that a single vote won't make a difference. The chances of an actual tie or one-vote-win are so remote that we do not take them seriously in deliberation, and it is not clear that we should. Outside the context of trying to defend consequentialism from the Problem of Collective Harm, I don't think it should seem plausible that the reason I have to vote in a large national election is that there is a miniscule chance of a tie or a one-vote-win, and thus a miniscule chance that the election will *come down to me*.

¹²² This point is not at all appreciated in the literature. Instead the issue is regarded as one about *imperceptibility*. Discussions of whether we can get out of the problem by denying the claim that an individual act won't make a difference have been focused on the question of whether there can be imperceptible harms and benefits. But what we saw in sections 3.5 and 3.6 is that accepting that there can be imperceptible harms and benefits does not itself avoid the problem. To argue that individual acts might make a morally significant difference in all Collective Harm Cases *requires* denying that there can be vague boundaries between morally significant outcomes, and this is so whether or not there can be imperceptible harms and benefits.

The point is that in triggering cases and quasi-triggering cases, the chance of making a difference could be overwhelmingly slim, and, when this is so, it's not clear that we can appeal to this chance to give a plausible reply to the Problem of Collective Harm that arises. Thus, not only are we not in a position to take a denying-the-description approach in purported *non-triggering* cases, such an approach might not work in many triggering cases and quasi-triggering cases either.¹²³

¹²³ The Oxfam Case might be a good example of a quasi-triggering case. There are, we might say, precise points in the 'series of donations' where things change with respect to the morally significant outcome. We can imagine that while donations are rolling in, meetings are had and decisions are made, and these decisions entail changes in the aid given to those in need. The fact that the donations are already in some rough vicinity could affect these decisions. But it seems unlikely that the decision-making process would be affected by a difference of, say, \$20 in the donations received so far. Then again, we cannot say *a priori* that the fact that the total amount of money that has come in so far is \$X rather than \$(X – 20) *couldn't* have affected the decision-making process.

Chapter 4

Helping Without Making a Difference

4.1 Introduction

In this chapter, I turn to my own proposal for how to solve the Problem of Collective Harm. My proposal begins from the observation that underlying the problem is a standard, intuitive assumption. The assumption is that if your act cannot *make a difference* to an outcome – if things will be the same with respect to that outcome whether or not you do it – then your act cannot play a *significant* causal role in bringing that outcome about. What I will propose is that, as intuitive as this assumption is, it is mistaken and that the key to solving the problem is to see this.

The chapter will be structured as follows. In sections 4.2 and 4.3, I will discuss the assumption, why we should reject it, and what the task of rejecting it involves. In 4.4, I will draw on examples of Non-Collective Harm Cases to help motivate the idea that there is something wrong with the assumption. In 4.5 – 4.7, I will develop an account of what it is to play a significant, non-superfluous causal role in the occurrence of an outcome, where this does not require making a difference, and I will use this to explain our reasons for action in Collective Harm Cases. In 4.8, I will give an error theory that explains why the standard assumption seems right, even though it is mistaken. And in the final section, 4.9, I will briefly discuss a question that arises with respect to my account that I am leaving open.

4.2 The Standard Assumption

The Problem of Collective Harm arises because when it seems that acting in a certain way won't make a difference, it's hard to see how there could be any point in doing so. As Frank Jackson writes,

Typically our actions make a difference. Had they not been performed, things would be different from the way they in fact are. Indeed, there

would be little point in acting if this were not so. Why do something if things would be *exactly the same* whether or not you did it?¹²⁴

As concerned as I might be about climate change, if things will be just as bad whether or not I drive, there doesn't seem to be any point in inconveniencing myself by refraining – or, at least, there doesn't seem to be any point in connection with the problem of climate change. Similarly, in Drops of Water, as concerned as I might be about the suffering men, if it won't make *any difference* to them whether or not I add my pint, then what reason could there be to add it? If things will be exactly the same for them with or without my pint, then adding it seems to be a mere waste.

Behind this line of thought (which generates the Problem of Collective Harm) is a standard, intuitive assumption. The assumption is that if your act won't make a difference with respect to an outcome, then it cannot play a significant causal role in bringing that outcome about. In other words, *helping* to bring about an outcome requires *making a difference* to it.

As I'm using the phrase, by 'helping to bring about' I mean playing a *non-superfluous* causal role in the occurrence of the outcome. Whatever it is that is actually required for an act to help – or, play a non-superfluous causal role – helping is a causal relationship that is of significance in practical and moral deliberation, and that can itself have reason-giving force. If by acting in a certain way I can help to bring about a good outcome, or help to prevent a bad one, this itself can give me reason to do it.

Whether your act *makes a difference* is a counterfactual matter: it is a matter of whether had you not done it, things would have been different. This, at least, is the notion of 'difference-making' that comes into the Problem of Collective Harm. The problem arises because it seems each person can say, "had I acted differently, things wouldn't be any better."

The standard assumption, that helping requires making a difference, is highly natural. If my act *makes no difference*, this means that things would have been the *same* whether or not I had done it, and this seems to be exactly what it is to be purely superfluous and thus not helpful. But, as natural as this assumption is, I think the key to solving the Problem of Collective Harm is to reject it.

Now, if we make the standard assumption, then if we want to solve the Problem of Collective Harm, rather than just bite the bullet and accept that – for instance – no individual has reason to add her pint in Drops of Water, or to buy a fuel-efficient car, or to vote in a national election, etc., then there are only two options, both of which – as we have already seen – are problematic.

¹²⁴ Jackson, "Group Morality", p. 94, my emphasis.

The first option is to accept that there are cases in which each act cannot itself make a morally significant difference but to argue that there are other reasons to do it, *apart from* anything to do with the individual causal role your act will play in bringing about the outcome. So, if one takes such an approach, one is saying: even though your act *can't help* to bring about the better outcome, there is this other kind of reason that you have to do it. In Chapter 2, we found that such a strategy isn't going to work. We saw that views that try to show that there are other sorts of reasons for action cannot get off the ground as solutions to the Problem of Collective Harm *unless* they can claim that your act could play a significant causal role.

The second option is to challenge the description of the cases: to argue that while it may *seem* that each act cannot make a difference, individual acts do, or at least might, make a morally significant difference in all the cases in question. In Chapter 3 we found that we are not in a position to obtain a general solution to the problem along these lines. Surely in some particular cases we are mistaken in thinking that an individual act won't make a difference, but we cannot say that we are *in general* mistaken when we think this. We should, instead, proceed on the assumption that in at least some Collective Harm Cases, an individual act won't make a difference.

This leaves us with one option, if we are to solve the problem: we need to reject the assumption. We need to show that your act can *help* without making a difference.

If we *can* reject the assumption, then for one thing, the proposals we saw in Chapter 2 can work. This would fill in the main hole that we found with those views. But, *moreover*, if we can reject it and show that an act can help in Collective Harm Cases even if it makes no difference, then that itself can be what provides us with reason for action. If my act can *help* to alleviate the men's suffering in the Drops of Water case, for example, *that* would seem to be a good reason to do it. It's not clear that we would need to appeal to any further considerations, like fairness or group membership, to explain why I have reason to add my pint. Indeed, this seems to be a central reason we take ourselves to have when we are motivated to act in Collective Harm Cases. While I have talked about the way in which the 'it makes no difference' argument arises in deliberation and can influence our choices, the other side of the story is that many of us are – at least some of the time – motivated to act in the relevant ways. Many people try to reduce the amount they drive or fly, refrain from purchasing factory-farmed meat, make modest donations to charity, avoid using disposable plastics, vote in national elections, and so on. When we are motivated to act in these ways, things seem quite different than they do when we are struck by the impression that an individual act won't make a difference. It is not as though we are thinking, 'yes this is superfluous –

it won't do any good at all – but there are other reasons to do it.' When we do these acts, we typically take ourselves to be non-superfluously contributing to the outcomes in question. Individuals frequently choose to act in the relevant ways precisely because they think that doing so will help to bring about the beneficial outcome, or help to prevent the harmful one. You might donate \$20 to Oxfam because you want to help to reduce the suffering of those in need. You likely recycle in order to help prevent environmental damage. *But*, since we make the standard assumption, this picture is taken to be inconsistent with the claim that an individual such act won't make a difference. The claim that your act won't make any difference seems to contradict the thought that in doing it you would be causally contributing in a significant way, and so it threatens this motivation for action.

My proposal is that there is no real inconsistency here. An act can help to bring about an outcome, even if it cannot make any difference, and the central task of solving the Problem of Collective Harm is to make sense of this. If we can, then we can get that our primary reason to act in these cases *is* the commonsense one: that by doing so, one could help to bring about the good outcome, or help to prevent a bad one.

4.3 Part of the Cause versus Helping

It may seem that if this is the view I want to put forth then what I need to do is argue against certain counterfactual theories of causation: that I need to show that an act can be part of the cause of an outcome even when it makes no difference to it.¹²⁵ This, however, is not what I need to do and it is important to see why.

The problem in Collective Harm Cases is *not* that no individual act seems to be part of the cause. On the contrary, it is a *starting point* – a point from which the problem gets going in the first place – that if the morally relevant outcome occurs, then individual acts of the type in question will have been part of what caused it. In Drops of Water, for example, it's part of the set-up of the case that if enough of us donate our pints to the cart, this will

¹²⁵ In his paper "Why Citizens Should Vote," Alvin Goldman takes such an approach. I think Goldman and I share a central intuition: that there can be reason for action having to do with the causal role your act could play, even if it won't make a difference. But what Goldman – who is focused specifically on the voting case – takes this to mean is that what he needs to do is show that a single vote for a candidate can be *part of the cause* of that candidate's win, even when it makes no difference. So, Goldman's paper is concerned with rejecting the simple counterfactual model of causation and with looking for models of causation that can capture an individual vote as causally involved. This, in my view, is not what needs to be done. It misses what is actually at issue in these cases.

result in the men's suffering being alleviated. So, in that situation, acts of water donation will have been part of what caused the beneficial outcome. That's not under debate. Indeed, without this causal relationship between acts of pouring water into the cart and the alleviation of the suffering, the question of why anyone has reason to donate their water would not have come up in the first place. Similarly, in the voting case, there is no question that if Y wins the election, acts of voting for Y were involved in bringing about this outcome. It is not as though we are under the impression that votes are only causally involved in elections when there is a tie or a one-vote-win.

So our challenge is not one of showing that individual acts of the relevant type can be causally involved, even if they make no difference. That is something we can – and really, must – take for granted. We must take it for granted to set up the Problem of Collective Harm to begin with. This isn't to say that it is not a difficult challenge of its own to come up with an account of causal involvement that works. But the point is that this is not our challenge. Our challenge, rather, is to show that one's act can be a *non-superfluous*, or *helpful*, causal component, even if it cannot make a difference.

To illustrate this distinction between *merely being part of the cause* and *helping*, consider this variation on the Drops of Water case. Suppose that ten thousand people each donate a pint and the cart is at maximum capacity. It is about to be driven out into the desert, but then I come along, with my own supply of water – ten thousand pints worth. I take a power hose from my supply, lower it into the cart, and turn it on full blast – forcing in the water from my supply. Because the cart was full, ten thousand pints overflow onto the ground, and we can suppose that (due to the placement of the hose, perhaps) what in fact happens is that the majority of what overflows is the original water, and the majority of my supply ends up in the cart. The cart is then driven out into the desert, the water distributed, and the men's suffering is alleviated.

It is clear that in this case my act is not helpful. It is purely superfluous. And, since it involves spilling a large amount of water, it is horribly wasteful. I surely had no reason to do it. Or at least, I did not have as a reason that doing so might help to alleviate the men's suffering. Still, my act is part of what in fact *causes* their suffering to be alleviated. After all, the majority of the water that the men actually drink is from my supply. So, my act is certainly causally involved in bringing about the outcome. It was just not a genuine way of

helping to bring it about, and thus its potential causal role in bringing about the beneficial outcome could not itself have given me reason to do it.¹²⁶

So, while my act is causally involved, it is not – in our sense – helping.¹²⁷ It is also true that it does not make a difference to the outcome: had I not power-hosed my water into the cart, things would be just as good for the men. Under the standard assumption, the fact that my act makes no difference is *why* it is unhelpful. Making a difference, on this assumption, is what separates a helpful part of the cause from a purely superfluous one. Under this assumption, from a causal standpoint, individual acts of donating one’s pint of water in the original Drops of Water case fall into the same category as my act of power-hosing in the Power Hose case: each is a purely superfluous causal component, since it makes no difference. This is the picture that we need to reject.¹²⁸

¹²⁶ Compare this to a case where I simply dump out my water supply upon realizing that the cart is already full. That would be a similarly wasteful act, but not one that is causally involved in bringing about that the men’s suffering is alleviated.

¹²⁷ On Goldman’s use of the word ‘helping’ my act of power-hosing *is* helping, because ‘helping’ for him just is ‘part of the cause’. He is using the term differently than I am. Goldman proposes two different models of causation (a version of Mackie’s INUS account and what he calls a “vectorial causal system”). As they should if they are accounts of causal involvement, both models of causation count my vote for *Y* as part of the cause of *Y* winning the election *and* my act of power-hosing as part of the cause of the men’s suffering being alleviated. Neither does anything to explain how the vote can be – as I am going to argue – non-superfluous, even though the power-hosing *is* purely superfluous and unhelpful. So, no light is shed on why one has reason to do the former but not the latter.

¹²⁸ In addition to focusing on what seems to me to be the wrong question – focusing on arguing for what we can take for granted in setting up the problem to begin with – I think Goldman gets things somewhat backward, possibly due to the fact that he is not thinking of the voting example as among a class of similar cases that raise the same problem (Collective Harm Cases). Goldman appeals to the fact that we would hold someone morally and legally responsible in overdetermined contexts in order to give a case against the counterfactual theory of causation. For instance, he takes the example of a firing squad, in which 10 people fire simultaneously at one victim, and writes “is it not bizarre to declare that none of their individual actions has any causal influence on the outcome? Moral responsibility for the death is presumably contingent on playing at least some causal role in the death, and so if we declare each shooter causally irrelevant to the death, we commit ourselves to absolving each shooter from any moral responsibility. That seems misguided.” (Goldman, “Why Citizens Should Vote”, p. 205) I agree with both Goldman’s conclusions here, but I think he gets things backward. He is appealing to verdicts about moral responsibility to justify his claim that individuals in the firing squad were causally involved in the death. It seems to me that the issue goes the other way around. We *know*, before asking any questions about moral responsibility that at least some of the individuals were causally involved in the death and that each could have been causally involved. Indeed, this is *why* we are asking about moral responsibility. From there a hard moral question arises: if each made no difference, it seems each is entirely superfluous to what happened, and so how can we hold him responsible? If

4.4 Helping Does Not Require Making a Difference

As tempting as it is to think that helping requires making a difference, it doesn't take very much to see that this cannot be quite right. Leaving aside Collective Harm Cases for now, there are many examples of Non-Collective Harm cases in which it is uncontroversial that an act helps, even though it very well might not make any difference.

Here is one example. Suppose you need some groceries. You can get to the store by going up either Shattuck or College Avenue. Neither route is better than the other. You go up College. But if you had not gone up College, you would have gone via Shattuck.¹²⁹

In this case, did your act of driving up College make a difference to the outcome in question (having the groceries)? No, it seems it did not: had you not gone up College, you still would have reached the store and gotten the groceries – you would have gone via Shattuck. The outcome in question would have been the same either way. Nonetheless, it's clear that your act of driving up College helped you get the groceries. There was nothing superfluous about it. And the reason you had to drive up College was that doing so could help to bring about this desired outcome. So we have an example in which one's act obviously helps to bring about an outcome even though it does not actually make a difference to it – even though the outcome would have been the same whether or not you had done it.

Now, what I have just said is not quite right. It is at least *possible* that your act of going up College made a difference. It is possible that had you not gone up College, something would have gone wrong. There could have been, for instance, an accident on Shattuck. So, we cannot say that it definitely makes no difference whether or not you go up College. But we can say that it *most likely* makes no difference. And this is all that is needed in order to obtain the conclusion that helping cannot require making a difference. Your going up College only had a *slight chance* of making a difference. Most likely things would have been the same whether or not you did it. But it really did help. It did not only have a slight chance of helping. It helped *regardless* of whether or not it made a difference.¹³⁰

what *I* do is entirely extraneous, why am I to be blamed? And what reason did I have act otherwise? *Merely superfluous* causal involvement does not seem to provide reason to refrain.

¹²⁹ This example is inspired by an example of the same type from Niko Kolodny, "Instrumental Reasons", pp. 23 – 24.

¹³⁰ A reply that might come to mind is: "But it *does* make a difference whether or not you travel to the store *some way*. That is, it makes a difference whether you go up *either Shattuck or*

College.” The suggestion here is that your act makes a difference under another description. It likely doesn’t make a difference under the description “going up College.” But it does make a difference under the description “going up Shattuck or College,” or “traveling to the store.” I think this is true, but it doesn’t affect the point being made. First, notice what this reply is pointing out. It is pointing out that whether or not your act makes a difference is relative to a *description*. If we ask ‘did it make a difference that I went up College?’ the answer is likely ‘no’. But if we ask ‘did it make a difference that I went up either College or Shattuck?’ the answer is likely ‘yes’. So, the answer to whether or not an act makes a difference is *relative* to how you describe it – it is relative to the *act type* you are interested in. This makes difference-making a weaker sort of notion than it might have initially appeared. There’s no simple fact about whether or not the act you performed made a difference; it depends on how we describe it. This, it seems to me, actually helps break down the force of the impression that making a difference is crucial for helping. Second, notice that it is still *true* that it most likely made no difference that you *went up College* and it is still true that, nonetheless, going up College obviously *helped* you get the groceries – it was not superfluous to go up College. So, under the description ‘going up College’, your act most likely does not make a difference, but, regardless, it does help. So, helping cannot require making a difference.

Next you might reply: “but it only helps in virtue of the fact that (a) going up College is an act of going up Shattuck or College, and (b) it makes a difference whether or not you do *that*. That is, your going up College helps because there is *some description* under which it makes a difference. So, helping does require making a difference *under some description*.” There are several problems with the suggestion that helping requires *making a difference under some description*. But the main thing worth pointing out here is that it just isn’t true that what explains why your act helps in this example is that it makes a difference under the broader description ‘going up Shattuck or College’. To see why, consider this alternate case, which follows the same pattern as the next example in this chapter. As before, you need to acquire certain groceries in time to make dinner tonight. Suppose your roommate says to you: “if you don’t get the groceries, I will get them a little later today.’ You go up College and get the groceries. Was your act of going up College causally superfluous with respect to the desired outcome of having the groceries in time for dinner? Clearly not. But it most likely made no difference. Your roommate is reliable, and so chances are that had you not gotten the groceries, she would have. So, your going up College most likely made no difference, but it did help. Here we cannot reply: “but it made a difference that you went up *College or Shattuck*”. It did not. Had you not gone any way at all, the outcome in question would have most likely been the same.

Now, at this point you might be inclined to press on: “but it made a difference whether or not *someone* did an act of the type *going up Shattuck or College*.” That may or may not be true, depending on what else we say about the example. But in any case, at this point we have traveled too far away from what was causing the trouble in Collective Harm Cases in the first place. In many Collective Harm Cases, there is some description of your act such that it very well might make a difference whether or not *someone* does an act of *that type*. For instance, in Drops of Water, suppose I happened to have added the 99th pint to the cart. Eventually enough people add their pints, and the suffering is relieved. In this case, there is a description of my act under which it makes a difference whether or not *someone* does such a thing. Had *someone* not *added the 99th pint* to the cart the suffering would not have been relieved. But this doesn’t show that your act made a difference in the original sense that got the problem going.

Here's another example. Judith's birthday is coming up on Wednesday and we have made her a card. It would be ideal if she received the card in time for the big day. It's Sunday; if we drop the card in the mailbox any time today, it will be picked up tomorrow morning at around 8am, and in all likelihood it will get to her on time. You are busy this morning, but you say to me, "If you don't mail the card this morning, I can do it this afternoon. It's no problem." It turns out, though, that I do have time and, wanting to do something useful with it, I go off to the mailbox and mail the card. The card is picked up at 8am the next morning, and arrives at Judith's on Tuesday.

Did my act of mailing the card make a difference with respect to the outcome of concern (the card arriving at Judith's and arriving in time for her birthday)? It is *possible* that it did: perhaps had I not mailed the card, you wouldn't have done so either. Perhaps you would have forgotten, or had a bout of akrasia. Perhaps on your way you would have dropped the card down a sewer. But *most likely*, none of this would have happened. You are highly reliable. Chances are had I not mailed the card, you would have done so later in the day, and it would have been picked up the next morning just the same. So, most likely my mailing the card made no difference with respect to the outcome in question. But it obviously helped to bring it about. It was not at all superfluous. It helped to bring about the desired outcome *regardless* of whether or not it made a difference.

Notice that while there was a *chance* that my mailing the card would make a difference, this didn't necessarily come into my motivating reasons at all. The chance of it making a difference has to do with the fact that there is a chance that you won't mail the card if I don't. But since I think of you as highly reliable, we can imagine that this sort of consideration doesn't come into my deliberations. Nonetheless, the reason I take myself to have for going and mailing the card *is* that doing so is a useful thing to do with respect to our goal. That is, I'm acting because I think it will likely *help* to bring about our desired outcome. So, my reason is *not* that doing so might make a difference. But my reason *is* that doing so might help.

Now, the point is not that there is a complete analogy between these sorts of cases and Collective Harm Cases; there isn't.¹³¹ These examples are

¹³¹ This is, in a way, *itself* part of the point. What this chapter is aimed at showing is that an act can fail to make a difference *for different reasons*, only some of which entail that one's act is superfluous – that it is not playing a morally or prudentially significant causal role. In the two-routes-to-the-grocery-store example, my driving up College most likely did not make a difference because if I hadn't done it, I would have done something else equally effective. In the birthday card example, my mailing the letter most likely made no difference because if I hadn't done it, you very likely would have done so instead. In triggering Collective Harm Cases, my act makes no difference, because the threshold is not hit exactly. In non-triggering

not supposed to themselves show that an individual act can play a non-superfluous causal role in Collective Harm Cases. What they are meant to show is that an act can help to bring about an outcome, quite apart from whether or not it makes a difference with respect to that outcome. What comes into focus when looking at examples like these is that making a difference is a matter of things *turning* on your act. Your act makes a difference to an outcome, if the outcome would have been *different* had you acted otherwise. But things need not turn on your act, in order for your act to help. Things need not turn on whether or not you do *this particular thing*, or even on whether or not *you* do something, in order for *what you actually do* to be genuinely helpful.

4.5 Conditions for Helping

If making a difference is not what separates a helpful part of the cause from a merely superfluous one, then what does?

Sticking with Non-Collective Harm Cases for now, let's consider a situation in which one's act is obviously superfluous, even though it's causally involved. Suppose it's 5:00pm and I want to park in a particular spot. I find that the meter is already paid until 6:00, and after 6:00 today, parking is free. Despite this, I put a dollar in. This advances the time on the meter to '7:00'. The police come by at 5:30, see that the meter reads '7:00' and so do not give me a ticket.

In this case, my act of adding a dollar to the meter played a causal role in bringing it about that I did not receive a ticket. After all, I don't receive a ticket because the police *see* that the time on the meter is 7:00, and it is 7:00 in part due to my addition of a dollar. But still, even though my act is causally involved, it's purely superfluous with respect to the goal of not getting a ticket. It did not help. And I did not have as a reason to add the money that doing so might help prevent me from getting the ticket.

Why is putting in the dollar superfluous? It is superfluous because before I put it in, there is already enough money in the meter to ensure that I will not get a ticket. While adding money could be part of what in fact causes the desired outcome, that outcome is already guaranteed. In general, it's clear that for an act to help to bring about an outcome, it must still be an open possibility that the outcome will fail to come about. If, at the time at which you

Collective Harm Cases, my act makes no difference, because there is no sharp threshold. None of these, in my view, actually block your act from playing a non-superfluous, significant causal role. The next section spells out what I think *would* block this.

act, failure is already not possible – if the outcome is already guaranteed – then your act is purely superfluous with respect to it.

But suppose the outcome is not guaranteed to occur. Your act could still be superfluous. Continuing with the parking meter example, it may not be true that, at the time at which I put the dollar in, it is guaranteed that I won't get a ticket. Maybe I'm parked over the line, and it is possible that I will get a ticket for that reason. Nonetheless, my act of adding a dollar is superfluous. Why is that? It is because, even if it's possible that I will get a ticket, it is not possible that I will get a ticket due to a lack of money in the meter. There is already enough money in the meter, and so if I get a ticket, it will have to be for some other reason. Had it been possible that I would receive a ticket due, at least in part, to a lack of money added, then adding money would not have been superfluous, and might have helped.

Compare this to the Drops of Water case. Suppose you add your pint well before the cart is full. In this case, at the time at which you act, it is still possible that the men's suffering will not be relieved. And, in particular, it is possible that it will not be relieved due, at least in part, to there not having been enough acts of adding water. Yes, your single pint can't itself make the difference between one level of relief and another. But more water is still needed if the suffering is to be relieved and there is no guarantee that enough people will add their pints. So, while it's still open that the good outcome will come about, it also might not. And it could be that it will fail to come about at least partly as a result of a lack of water added to the cart. This, it seems to me, is enough to make your act of adding a pint non-superfluous, and thus potentially helpful. If the men's suffering is alleviated, your act will have played a causal role in bringing this about – it will have been part of what caused this good outcome. And the causal role your act plays is *non-superfluous* because, at the time at which you do it, there's a real risk that the outcome will fail to come about as a result of not enough people acting in exactly that way.

In the power hose case, on the other hand – in which I come along and replace the water with my own supply – it's not possible, at the time at which I add my water, that the men will continue to suffer as a result of there not having been enough additions of water. There is already enough water in the cart. So, if the men's suffering is not alleviated, this will have to be a result of something else. It cannot be due, in any part, to a lack of water added. This is why my act of power-hosing is purely superfluous, and thus not helping to bring about the outcome.

Here is my proposal, in general terms:

Suppose that acts of a certain sort – acts of *X-ing* – could be part of what cause outcome *Y*. In such a case, your *act of X-ing* is non-superfluous and so could help to bring about *Y* if and only if, at the time at which you *X*,

(*) It is possible that *Y* will fail to come about due, at least in part, to a lack of *X-ing*.

Contained in this account of helping are three conditions that are worth separating out. First, contained in the supposition that *acts of X-ing could be part of what cause Y*, we have that:

(1) It must be possible that *Y* will occur.

An act cannot be potentially part of the cause of *Y* if *Y* is impossible.

Contained in (*) – the condition for non-superfluity – we have that:

(2) It must be possible that *Y* won't occur, *and*

(3) possible that it won't occur, at least in part, as a result of there not having been enough acts of *X-ing*.¹³²

My proposal is that when these conditions hold and when the outcome in question is a good outcome, or the avoidance of a bad one, then one has as a reason to act in the relevant way that doing so might *help* to bring it about. (This reason is a morally relevant one as long as the outcome in question is good or bad from a *moral* perspective – rather than, say, only from a prudential perspective, in which case the consideration would be purely prudential.¹³³)

Notice that your act could satisfy the conditions and yet not actually help to bring about the outcome in the end. For instance, the outcome could fail to come about, in which case – of course – your act didn't help to bring it about. For your act of *X-ing* to *in fact* help to bring about *Y*, it must be that the above conditions hold, *Y* in fact comes about, and your act of *X-ing* is in fact

¹³² The notion of possibility that comes into this account needs to be one that we use in contexts of practical and moral deliberation. For the purposes of this dissertation, I am largely going to remain agnostic about how we should cash out this notion of possibility, but in the final section of this chapter, section 4.9, I will briefly discuss this question.

¹³³ The outcomes that we are concerned with in Collective Harm Cases are morally relevant outcomes. So, in such cases, reasons for action having to do with how doing so could help are morally relevant reasons.

part of what causes \mathcal{Y} .¹³⁴ But when it comes to deciding whether or not you have *reason* to X , the question that needs to be asked is whether your X -ing *could* help to bring about \mathcal{Y} , and not whether your X -ing will have in fact helped to bring about \mathcal{Y} in the end.

Let's apply the account to the voting case. Suppose that there are two candidates running for the presidency of a large country. One of them is Dr. Evil – someone who will cause a lot of harm were he to be president. The other candidate would be much better – not at all evil, and quite a good leader. Let's call him "Mr. Powers". On my proposal, a primary reason you have to vote for Mr. Powers is that doing so could help prevent disaster. You have this reason as long as there have not yet been enough votes to settle who will have the majority. As long as the election could still go either way, then more votes in favor of Mr. Powers are needed if he is to win. So, it's possible that Mr. Powers will not win due, at least in part, to a lack of votes in his favor. Thus, your act of voting for him is not superfluous, and you have as a reason to do it that it could help.

Now, suppose that more than enough votes have been cast in favor of Mr. Powers to *ensure* that he will have the majority. This does not guarantee that the disastrous outcome will be avoided. It is possible, for instance, that Dr. Evil will attempt to do something evil, and try to bring it about that he is declared president either way. Still, at this point, your vote cannot *help* to prevent the bad outcome. This is because what is not possible at this point is that Dr. Evil will become president due, in any part, to a lack of votes for Mr. Powers. Mr. Powers already has more than enough votes to win, and so if he does not become president, this will have to be for some other reason. So, voting for Mr. Powers can no longer help to prevent the disastrous outcome.

This might seem problematic. Surely – you might think – one has reason to vote even if there have already been enough votes to settle the winner?¹³⁵ But, first, it is important to notice that as long as you don't *know*

¹³⁴ Here's an example of a case in which all but the last of these points hold: suppose I am adding money to the meter when there *isn't enough* – that is, when it is possible that I will get a ticket as a result of not enough money in the meter. In this case, my adding money could help to prevent me from getting a ticket. But suppose that the police just never come by. If this is so, while my act *might have helped*, and while I did not get a ticket, it did not in fact help. It did not in fact help to bring about the good outcome because my having added money was not, in the end, part of what caused the good outcome, even though it could have been.

¹³⁵ Goldman raises this as a concern about his account. He writes:

At this point an important objection must be confronted ... [In elections it will often be that] by the time many voters cast their ballots, the outcome has already been decided. This is particularly striking in the case of presidential elections in the United States, where several time zones are involved. By the time voters in California or

that one of the candidates has more than enough votes to ensure a majority, you would still have as a subjective reason to vote that doing so might help to prevent the bad outcome. Even if you are aware of the fact that the Mr. Powers is way ahead in the polls, you would still have this subjective reason to vote if, as far as you know, it is still possible that the election will turn around in Dr. Evil's favor. Second, importantly, you could have other reasons to vote. You could, for instance, have as a reason to vote for Mr. Powers that doing so could help to bring about some further outcome, apart from, or over-and-above, Mr. Powers winning. Suppose it matters not just whether Mr. Powers wins, but whether he wins by what is regarded as a large margin. Even if your single vote cannot itself make a difference to whether or not the margin is regarded as large, your vote could *help* to bring it about that the margin is regarded as large. This could give you reason to vote.¹³⁶ What is true on my account is that once the winner is settled, you no longer have the *same* objective reasons to vote that you had when the election could still go either way.

4.6 Reasons to Refrain

In this chapter, I have been focusing on Collective Harm Cases in which enough acts of a certain sort will bring about a good outcome or prevent a bad one. But in many Collective Harm Cases it is, rather, that unless enough people *refrain* from acting in a certain way, bad consequences will result. That is, there are many cases in which enough acts of a certain sort will collectively cause some bad outcome, and what we need is an explanation as to why each individual has reason to refrain from acting in this way. I have

Hawaii go to the polls, the earlier-voting states may already have determined the victor. How can voters in these later-voting districts rationalize their voting? Do their votes really count as partial causes of the outcome (even if they vote for the winner)? (Goldman, "Why Citizens Should Vote?", p. 212)

Goldman's response is to argue that one's vote still counts as part of the cause of the outcome, even if the election is already determined by earlier votes. I think this is reasonable. It seems reasonable to say, for instance, that a vote for Y is part of the cause of Y winning as long as it was among the votes counted. But this is beside the point: it does not entail that you have reason to vote. Even if it will be part of the cause, voting at this point would be a superfluous part of the cause.

¹³⁶ There may also be other sorts of reasons to vote, including reasons that do not – in my view – do anything to help resolve the Problem of Collective Harm but that still can come into Collective Harm Cases. For instance, you might vote because you want to express, even if only to yourself, your solidarity with a certain political party, or symbolically affirm your commitment to certain values.

been fairly loose with the distinction between acting and refraining in the dissertation so far, often using ‘act in the relevant way’ for both acts and omissions. But, at this point, it seems to matter that sometimes what we are looking for is a *reason to refrain*. In Harmless Torturers, for instance, if there are enough acts of dial-turning the victim will suffer tremendous pain. So, we want to be able to explain why each individual has reason *not* to turn the dial. But we cannot say that the reason is that *not* turning up the dial will help to prevent the harm. One’s *not* doing something – one’s omission – cannot play a causal role in bringing about an outcome (at least not in the sense of ‘cause’ that comes into the account of helping.¹³⁷) So, how does my account handle such cases?

The first thing to note is that, just as you can help to bring about a good outcome or help to prevent a bad one, you can help to bring about a bad outcome or help to prevent a good one. Helping – an act’s playing a non-superfluous, deliberately significant, causal role in the occurrence of an outcome – is not necessarily a “positive” thing. Whether it is or not depends on the outcome.¹³⁸ We have said that if by acting in a certain way I might help to bring about a good outcome or help to prevent a bad one, this gives me reason to do so. But, similarly, if by acting in a certain way I might help to bring about a bad outcome, or help to prevent a good one, this gives me reason to *refrain* from acting in that way. If my act might make a non-superfluous causal contribution to bringing about an avoidable bad outcome, this gives me reason not to do it.

In Harmless Torturers, we know that if there are enough acts of dial turning, this will cause the victim to suffer. As long as *more* dial turns are needed if the victim is to suffer (or to suffer more), turning the dial is non-superfluous with respect to this harm. It would still be possible that the suffering will be avoided due to a lack of dial-turns, and so, turning the dial is non-superfluous with respect to bringing about this bad outcome. This gives you reason to refrain.

This nicely captures how we often do think of our reasons in such cases. One might refrain from using a plastic cup in order to avoid helping to bring about environmental damage that is caused by the use of disposable plastics. One might refrain from purchasing factory-farmed meat because one thinks

¹³⁷ Even if one thinks that there is a sense in which omissions can be causes, it doesn’t seem that they can be causes in the sense of *cause* that comes into my account. In “Two Concepts of Causation,” Ned Hall argues that omissions can be causes in one sense of the term – the counterfactual sense – but also that there is another notion of causation, “production” that we use. Something like ‘production’ is the notion relevant in my account.

¹³⁸ Note, as well, that I don’t rule out that it could be positive from a moral perspective but negative from a purely prudential perspective, or vice versa.

that such a purchase plays a part in sustaining an industry that causes animals to suffer horrendously. One might refrain from driving because one wants to reduce one's contribution to climate change. My account allows us to make sense of these reasons, even if it is true that a single such act cannot make a difference. Even if the state of climate change – or, the harm that results from it – will not be different depending on whether or not I drive, my act of driving could play a non-superfluous causal role with respect to it. In general, even if my act cannot itself make the difference between one morally significant outcome and another, it could amount to taking a non-superfluous step closer to a certain morally significant outcome, and this can explain why I have reason to refrain (or, reason to do it, depending on what outcome we are talking about).¹³⁹

The content of the reason to refrain is that acting in the relevant way could help to bring about the bad outcome (or help prevent the good one). Now that we understand this, we can feel free to speak a bit loosely. We can say things like, “*help prevent environmental damage*: reduce the amount you drive.” When we say this, I think the consideration we are getting at is that an act of driving might help cause further environmental damage. But, for simplicity, we can be loose and describe the reason the other way around. That is, when it is simpler or more natural to do so, we can think of *not* doing something that could help cause a bad outcome, as ‘doing something that could help prevent the bad outcome’.

4.7 Triggering versus Non-Triggering Cases

On my view, the primary reason individuals have to act in the relevant way in Collective Harm Cases is the *same* whether we are dealing with a triggering case or a non-triggering case. Notice that in the voting case, the reason that one has to vote has nothing in particular to do with the chance of your vote making a difference, with the chance of the election *turning* on your vote. Your vote helps because, at the time at which you vote, more votes for Mr. Powers are needed to prevent the disastrous outcome, and there is no guarantee that there will be enough such votes. So, it's possible that the bad outcome will not

¹³⁹ This also seems to capture how we often evaluate others in such cases. When one sees a truck idling unnecessarily, spewing a cloud of black exhaust into the air, one might think that the driver ought to turn off her truck because of what she is, right there, *doing* to the environment. My view makes sense of this. Even if she is not herself making a difference for the worse by idling (even if environmental damage would be just as bad had she not left the truck idling there for 10 minutes), this act is playing a non-superfluous role in bringing about further damage.

be avoided due, at least in part, to a lack of votes for Mr. Powers. This has nothing in particular to do with there being a chance that your vote will make or break a tie. And it is exactly like the reason you have to add your pint in the Drops of Water case, a non-triggering case.

There is, though, one notable difference between these reasons as they occur in triggering cases and these reasons as they occur in non-triggering cases. In triggering cases there are precise – even if unknown – points at which your act *first* can no longer help to bring about the relevant outcome. In the Dr. Evil case, suppose that more than half the population eligible and able to vote has voted for one of the two candidates. When the first vote past this halfway point is cast, this is the point at which your vote *first* can no longer play a non-superfluous role in preventing Dr. Evil from being elected. So, this is a precise – though unknown – point at which, your objective reasons to vote change. Even if the polls are still open, you no longer have as an objective reason to vote for Mr. Powers that doing so could help prevent Dr. Evil from being elected. (But as long as it is unknown that this point has already been reached, you still have the subjective reason to vote.) In non-triggering cases, on the other hand, there are no precise such points. In Drops of Water, when *close to* ten thousand people have donated, there will be enough water in the cart for the men to receive the *same* benefit that they would receive from the full ten thousand pints worth. Once this has occurred, even though there is still some room left in the cart, adding more water can no longer help to alleviate their suffering because it's not possible that part of what will stand in the way of full relief is a lack of water donations.¹⁴⁰ But because there is no precise threshold at which this occurs – since the boundary between enough pints and not enough is vague – there is no precise point at which adding a pint *first* ceases to help. So there can be no precise point at which one goes from having as an objective reason to add one's pint that doing so might help to not having this reason. This is not problematic, though. Indeed, it seems exactly right. For one thing, as long as it is clear that more water is still needed for full relief –

¹⁴⁰ Notice that my view gives opposite verdicts about our reasons for action 'around the edges' of Collective Harm Cases from at least some of the other views on offer, including Fairness and Membership. Take the Fairness approach. As we saw in Chapter 2, Cullity does not think that the Fairness approach kicks-in *unless* others have already started adding their pints; he also says that – according to Fairness – individuals are obligated to add their pints *until the cart is full*. (See Cullity, p. 20) So, if no one has yet added their pints, the Fairness approach does not tell you to add your pint; but if the cart only has room for one more pint, it still says that you ought to add it. My views outputs the opposite verdicts at these two borders: as long as it is *possible* that others will add their pints, there is reason to add the *first* pint, but when there is only room for one more pint, there is no reason to add any more water to the cart.

which it will be until the cart is *close* to full – each person with a pint will have the reason in question to add it. As the cart gets close to full, there will be no precise cut off point at which you first lose this reason. But if this case really is a non-triggering case, that makes sense: if there is no sharp boundary between enough pints in the cart and not enough, then why would we expect there to be a sharp boundary at which the reason to add a pint goes away? And, as with triggering cases, as long as it is unclear whether or not there have been enough acts of the relevant type, one would still have as a subjective reason to act that doing so could help.

4.8 Why it *Seems* As Though Helping Requires Making a Difference

The problem in Collective Harm Cases was that if you cannot make a difference by acting in the relevant way, it's hard to see how there could be any reason to do so. My proposal is that not being able to make a difference is not actually the determining factor: it does not block you from having a reason for action, and in particular a reason that has to do with the causal role your act could play.

What *would* block you from having this sort of reason would be if the outcome were already settled *or* if there were no chance that the better outcome would fail to come about at least partly as a result of there not having been enough acts of the type in question.

In addition to explaining why one does have reason to act in the relevant ways in Collective Harm Cases, this account explains why it *seemed* as though not being able to make a difference would block being able to help. It explains why the standard assumption – that helping requires making a difference – seems right, even though it is mistaken.

Typically – in ordinary, Non-Collective Harm cases – if an act cannot make a difference, this is a situation in which that act cannot help. Indeed, in ordinary cases of an act being unable to make a difference, your act is unable to make a difference as a *result* of the fact that the conditions for helping do not obtain.

Think of ordinary, unproblematic cases in which you realize that doing something won't make a difference and then conclude from this that you have no reason to do it. Suppose, for example, that a student is taking a class on a pass/fail grading scheme, rather than for a letter grade. It's the end of the semester, and before the exam, she does some calculations to find out where she stands. What she discovers is that she will pass the course no matter what; even if she gets a zero on the final, she will still pass. So, taking the exam

cannot make a difference to whether or not she passes. The outcome will be the same whether or not she takes it. She might very well conclude from this that she has no reason to take the exam. And, whether or not she has other reasons to take the exam, she would be right to conclude that she does not have as a reason to take the exam that doing so might help her pass the course. But this is *because* this is a situation in which failure is not possible. Condition (2) is not satisfied. The desired outcome – passing the course – is already guaranteed. That is what explains why taking the exam cannot help. The fact that taking the exam cannot make the difference between a pass and a fail is *also* explained by the fact that the good outcome is already guaranteed. It is settled in advance that taking the exam will not make a difference to whether or not she passes precisely because it is already settled that she will pass.

Here's another example. I have a casserole recipe that calls for blanching the vegetables first. I have made the recipe doing the blanching step and I have made the recipe skipping it, and so far it has come out the same either way. It seems to me, then, that it makes no difference to the success of the dish whether or not I blanch the vegetables. So I have concluded that I don't have reason to do it. Blanching seems to be a waste of my time. Now, if I am *right* that blanching cannot make a difference to the success of the dish, then I am right that there is no reason to do it – or, at least, no reason having to do with how blanching could help to make the dish more successful. But this is *because* this would be a situation in which Condition (3) fails to hold. It's possible that the casserole won't come out – maybe I'll accidentally leave out an ingredient. But if the casserole doesn't come out this will not be due, in any part, to not having blanched the vegetables. This is why blanching is superfluous and cannot possibly help. This is *also* why it cannot make a difference. It is settled in advance that the quality of the casserole will not differ depending on whether or not I blanch the vegetables *because* not blanching the vegetables cannot be any part of what stands in the way of a superior casserole – that is, because blanching is a superfluous part of the recipe.

Now, what if I was mistaken? What if blanching *could* make a difference to the success of the dish? In this case, the fact that the casserole came out the same in my attempts so far – the fact that blanching *hasn't made a difference* – would have to be a result of something else other than the failure of the conditions for helping. Perhaps, for instance, I accidentally over-baked the casserole, and over-baking somehow negated the improving effects that blanching-in-advance-of-baking can have. In this case, my blanching the vegetables *could* have helped make the casserole a good one, and I did have reason to do it. But it did not make a difference, and – *more generally* – it did not help, because something else that I did afterwards (over-baking) undid the

good effects of the blanching and so prevented the better outcome from coming about.

The point is that in typical Non-Collective Harm cases, when an act cannot make a difference, it *also* cannot help. But this is not because the only way to help *is* to make a difference. It is because in these ordinary cases, one's act is unable to make a difference *because* the conditions for helping do not obtain. This explains why it is easy for it to look as though helping requires making a difference. In these typical, familiar kinds of cases, being unable to make a difference *reliably indicates* that an act is unable to help.

This particular point actually carries over to *triggering* Collective Harm Cases. In a triggering case, if it's not *at all possible* that an act will make a difference, it will also be that it's not possible for it to help. Take the voting case. When you vote, there is usually at least a chance – no matter how extremely tiny – of your vote making a difference, because there is at least some chance of a tie or a one-vote-win. On my view, your reason to vote does not come from the fact that there is this miniscule chance of making a difference.¹⁴¹ Instead, it comes from the fact that your vote could help get the better candidate elected. And it can help, even when it does not make a difference. But suppose there is *no chance* of your vote making a difference. Could your vote still help? Well, when is there no such chance? The only scenarios in which a tie or a one-vote-win are impossible are scenarios in which the winner is already settled, or – at least – in which it's already settled which candidate will have the majority of the votes. As long as it is an open possibility that either candidate could have the majority, then it is still possible that there will be a tie, and so possible, though extremely unlikely, that your vote will make a difference.

Thus, even in the voting case, if your vote cannot possibly make a difference, then it will also be true that it cannot possibly help. But this isn't because a vote can only help *by* making a difference. The election need not *turn* on your vote in order for your vote to help. It is because when there is no chance at all of making a difference in a triggering case, this means that there is no chance of the threshold being hit exactly. And the only way for that to be impossible is for things to be *settled* with respect to whether or not the threshold will be met. That is exactly when the conditions for helping are not satisfied. So, in triggering cases – like ordinary Non-Collective Harm cases – when there's no chance of an act making a difference, this is a *result* of the fact that the conditions for helping do not hold. When the Problem of Collective Harm arises in triggering cases, though, the conditions for helping do hold, and so

¹⁴¹ Or, at least this is not your main reason to vote in the case of a large-scale election.

there will be at least some chance of making a difference, whether or *not* this chance comes into the reasons that you have to act at all.

When we turn to non-triggering cases, though, when the Problem of Collective Harm arises, there is no chance of making a difference, and this does *not* result from the failure of the conditions for helping. Instead, in non-triggering cases, one's act has no chance of making a difference as a result of something else: namely, a *lack of sharp thresholds* – the fact that there is no precise number of acts of the relevant type needed to bring about the outcome in question. So, here – unlike in triggering cases and ordinary non-Collective Harm Cases – the fact that your act has no chance of making a difference does *not* indicate that it cannot help.

Either way, whether one is in a triggering case or a non-triggering case, when the Problem of Collective Harm arises, the conditions for helping hold. It's not as though the outcome is already settled. Even if one's individual act will not *itself* make a difference, things could still go either way. Dr. Evil could come into power, or he could be defeated. Environmental damage, and resulting harm to humans and animals, could worsen or it could be reduced. Consumer patterns could shift in a way that decreases global poverty, or they could not. As long as it's still possible that the bad outcome will not be prevented due, in part, to there not having been enough acts of the relevant type, then doing such an act is not at all superfluous. Acting in such a way could *help* prevent the bad outcome, even though what happens with regard to that outcome does not turn on whether or not you act in that way.

4.9 The Notion of Possibility

In this final section, I want to briefly discuss the question: how are we to understand the notion of possibility that comes into the account of helping? Not just any notion of possibility will do, of course. In particular, it needs to be one that we use in contexts of practical and moral deliberation. We can rule out some suggestions quickly. For instance, it is obviously not logical possibility or metaphysical possibility.¹⁴² But it is a difficult question which notion(s) of possibility we do use in such contexts, and I do not have a settled opinion about how to cash out the notion of possibility that comes into the account of helping. I do, however, want to set out two constraints on what it could be.

The first constraint is that what is *possible* and what you have *reason to believe is possible* must come apart. For instance, if it is 6pm and I know that the

¹⁴² It is logically and metaphysically possible that by clapping our hands we will alleviate the men's suffering, but as far as the notion of possibility that comes into deliberation is concerned this is not possible.

store's business hours are 8am - 8pm, then I have reason to believe that it is possible for me to pick up groceries on my way home. So, I have reason to believe that going up College *could help*. But if I get there and find that due to a family emergency the store actually closed at noon today, I would not take myself to be discovering that while, as of a minute ago, it was possible that I would get the groceries, it is suddenly no longer possible. I would, rather, take myself to be discovering that, while I had good reason to think as I did, I was wrong. It was *not* in fact possible for me to pick up the groceries, and so – contrary to what I had reason to believe – it *couldn't* have actually helped to go up College. While I must base my decisions on what I believe is possible, in contexts of deliberation I can and do distinguish between what I have reason to believe is possible and what actually is possible. So, the notion of possibility cannot be 'what you have reason to believe is possible,' nor can it be 'what as far as you know is possible'.¹⁴³ However, this is not meant to rule out that there may be some other epistemic notion of possibility that respects this constraint and that would work.

The second constraint is that the notion of possibility must be *deliberation-independent*: it cannot hold fixed what the relevant agents will choose to do.¹⁴⁴ Suppose we know that nothing could go wrong if you attempt to go up College: the streets are totally clear; your car is in great condition; you are an exceptional driver; friendly, magical fairies will guide you to ensure you arrive safely and within the needed time-frame to get the groceries.¹⁴⁵ Does this mean that it is *not possible* for you to fail to get the groceries due, in part, to a lack of going up College? If so, this would mean that, on my account, going up College is purely superfluous. But clearly it is not superfluous; going up College can help, whether or not your success in doing so is assured. My response to this example is: *of course* it is possible that you will fail to get the groceries due, in part, to not having gone up College. If you haven't yet gone up College, it is open for you to *choose* not to go and to act on that choice. Even if you are three quarters of the way up College, you could choose to stop and turn around. Unless these fairies are not as friendly as we thought and are actually dragging you up College, or unless you are otherwise externally forced to go, it is possible – under a standard notion of possibility used in contexts of

¹⁴³ Indeed, the fact that what as far as I know is possible and what is possible might come apart is a consideration that I sometimes take into account in deliberating. For example, I might think, 'as far as I know it is not possible to fly direct from Oakland to Toronto, but I might be wrong; I should look into this more thoroughly.'

¹⁴⁴ Thank you to Sergio Tenenbaum, who originally pointed out to me that I needed deliberation-independence.

¹⁴⁵ Thank you to Niko Kolodny for this great example.

deliberation – for you to choose not to go, or for you to change your mind.¹⁴⁶ To capture this, we need a notion of possibility that does not hold fixed what agents will choose to do. That is, we need a deliberation-independent account of what is possible.¹⁴⁷

¹⁴⁶ If you are being externally forced up College, by fairies or otherwise, it's not clear that we would call your going up College an act of yours at all. If that's right, then my account of helping (which specifically has to do with acts) would not apply. But suppose we do think that it is an act of yours. We could then ask, "is this something you have reason to do, even though in fact you have no choice in the matter?" But if we ask that, we must extrapolate from the fact that you have no choice in the matter. We are asking about whether there is reason to do it *independent* of the fact that you are going to do it whether or not there is any reason (i.e. because you are being forced.) To figure out the answer to that question, we must imagine the case as one in which you *do* have choice.

¹⁴⁷ *Conditional possibility*, which is an account of which *actions* are possible for an agent, is a notion of possibility specifically designed to be deliberation-independent. According to this account, the actions that are possible for me are those that I would do, if I chose to do them. (For an argument in favor of this account, see, for example, Bok, "Freedom and Practical Reason".) But this is just about which *actions* are possible for an agent. We need a general notion of possibility. Perhaps, though, we can extend this notion of conditional possibility. Perhaps, for instance, we can say: \mathcal{Z} is possible *if and only if* there is at least one set of choice-assignments (i.e. assign a choice to each relevant person) such that if these choices are made, then \mathcal{Z} will come about. This might be a promising start, though further qualification would likely be needed. (For instance, one problem with conditional possibility that would carry over here is that there may be some things that, due to facts about your psychology, you cannot choose, but that *if* you did choose them, you would do them. In that case, we don't want to say that it is possible for you to act in that way, even though the conditional analysis says that it is.) In any case, I am not committed to this proposal, and want to leave it as an open question which account of possibility works best.

Chapter 5

Concluding Remarks

To conclude, I want to highlight some noteworthy features of my proposal and comment briefly on a couple of questions that I have left open for future work.

5.1 Separability

The first point worth noting is that my view has two parts that can be *separated*: one could accept the first part, and put it to work, without accepting the second part. The first part is my proposal concerning how it is that an act can *help* – or, play a significant, non-superfluous causal role – without making a difference. The second part is the claim that this itself explains why there is reason for action in Collective Harm Cases: the main, morally relevant reason that we have to act in such cases *is* that doing so could help to bring about the good outcome, or help prevent the bad one. Whether or not you are convinced of this second point, my account of how it is that you can help without making a difference can be used to fill out other views about our reasons for action in Collective Harm Cases, including each of the views discussed in Chapter 2. So, one need not accept my *whole* picture about what is going on in Collective Harm Cases to get mileage from my proposal.

5.2 Other Reasons

Of course, the fact that one wants to adopt another proposal about our reasons for action does not mean that one should reject the second part of my proposal. It is entirely consistent with my proposal that there can be *other* reasons for action besides those that helping gives you. For instance, with my account, we can say – if we want to – that one has as reasons to add one’s pint in Drops of Water *both* that doing so could help to alleviate the men’s suffering *and* that not doing so would be unfair to the others who have donated. I don’t think we can say that keeping your pint is unfair without my account; it is an upshot of Chapters 2 and 3 that unless we can explain how one can help without making a difference, appealing to fairness in many Collective Harm

Cases does not make sense. But *with* my account, it does make sense: if adding my pint could help to alleviate the men's suffering, then it at least *makes sense* to say that if I don't add it while others do, then I am not pulling my weight in what we all ought to be doing.

Now, my own view is that fairness does not provide the main, or central, reason for action in Collective Harm Cases. On my proposal, the main reason to donate to Oxfam is *not* that doing so would be fair to others who have donated. The main reason to donate is that doing so could help to improve the situation of those in need. Similarly, on my view, the main reason to recycle is to help to prevent environmental damage,¹⁴⁸ and not that doing so would be fair to others who have recycled. But even if you agree with me about this, the point is that further considerations (such as fairness) can provide *additional support* for acting.¹⁴⁹

5.3 Extendability

The third point worth noting is that my account of helping can be easily extended so as to apply to relations that acts can stand in to outcomes other than *causation*. I focused only on causation for simplicity, but it is important to see that the role really need not be a *causal* one for my account to work. For instance, in addition to playing a non-superfluous causal role, we can easily extend the account to cover playing a non-superfluous *constitutive* role. We can add:

Constitutive-Helping

Suppose that acts of a certain sort – acts of *X-ing* – could be part of what *constitute Y*. In such a case, your *act of X-ing* would be *non-superfluous* and so could *help* to bring it about that *Y if and only if*, at the time at which you *X*,

¹⁴⁸ Or, more accurately, to avoid helping to cause environmental damage.

¹⁴⁹ Just to be explicit, it should be clear that one could reject the 'main reason' part of my proposal, but keep the rest. Suppose, for instance, that you think that considerations of fairness give rise to a *moral obligation* to add your pint. You might, then, want to say that the fact that you are *morally obligated* to add your pint is the main reason that you have to add it. Still, to get this story about the moral obligation going, you need it to be the case that by adding your pint you could *help* to alleviate the men's suffering. And you might agree that the fact that you could help gives you at least some morally relevant reason to add the pint. That is, you could agree with me that helping is not just a *precondition* for the fairness story to get off the ground, but also that it is a morally relevant consideration of its own in favor of acting.

(*) It is possible that Y will fail to come about due, at least in part, to a lack of X -ing.

We can either think of this as a separate notion of helping: helping-to-constitute rather than helping-to-cause. Or, we can expand our concept of *helping* to cover both. Helping, on the expanded understanding, is playing a non-superfluous *causal or constitutive* role in the occurrence of an outcome. Either way, we can explain how one can help in a *constitutive* sense, without making any difference.

For an example to illustrate why this would be useful: suppose we think that it is important that we have a flourishing, active democracy. Having a large portion of the population vote in the national election does not *cause* a flourishing democracy but rather is partially *constitutive* of it. Now, a single vote does not seem to make a difference to whether or not our democracy is flourishing. If our democracy is flourishing, taking a single vote away will not change this, and if it is not flourishing, adding a single vote cannot be enough to make it flourish. But, even if this is right, on my account, a single vote can *help* to bring it about that our society is, or continues to be, a flourishing democracy. It would help – or play a non-superfluous *constitutive* role – as long as it is possible that the democracy will fail to flourish, or flourish as much, due in part to a lack of voting.

We can, in the same way, extend the account so as to capture notions of playing a non-superfluous role in other relevant relations that acts can stand in to outcomes. For another example, take *preventing a preventer*.¹⁵⁰ Preventing a preventer of Y is not the same as causing Y . If you are going to block the ball from going into the net, but I knock you over before you can, then I have not caused the ball to go into the net; I have prevented you from preventing it from going in. We can easily extend my account of helping to cover playing a non-superfluous preventing-a-preventer role:

Helping to Prevent a Preventer

Suppose that acts of a certain sort – acts of X -ing – could be part of what *prevent a preventer* of Y . In such a case, your *act of X-ing* is non-superfluous and so could *help* to prevent this preventer of Y *if and only if*, at the time at which you X ,

¹⁵⁰ I would like to thank Seth Yalcin for pointing out that there could be a Problem of Collective Harm involving preventing a preventer, and that my account of helping (done in terms of ‘causal role’) would not apply in such a case.

(**) It is possible that *the preventer of Y will fail to be prevented* due, at least in part, to a lack of *X-ing*.

Extending the account in this way allows us to explain how it is that one can *help* to prevent a preventer, even if one cannot make any difference to whether or not the preventer is prevented. For a quick example, just imagine that the ‘victim’ in Harmless Torturers is an evil villain who will prevent the delivery of crucial life-saving supplies to a hurricane ravaged region *unless* he is electric-shocked into submission. In this case, you could have as a (morally relevant) reason to turn the dial on the torture machine that doing so could help to prevent this evil villain from preventing the delivery of the life-saving supplies.

5.4 Individual Over Time Cases

In Chapter 4, in giving the error theory (the explanation of why it seems as though helping requires making a difference even though it does not), I talked about “*typical*” Non-Collective Harm Cases. I didn’t explain, though, what exactly I had in mind as “*non-typical*” cases. The “non-typical” Non-Collective Harm Cases that I had in mind are not actually uncommon; rather, they are non-typical only because they are *analogous* to Collective Harm Cases, and so they give rise to a largely analogous problem. These are cases in which, instead of a *group* of people doing acts of a certain sort, we have a *single individual* doing acts of a certain sort over a period of time. By acting in this way over time, an outcome of significance (whether from a moral perspective or some other perspective, such as self-interest) results. But – as in Collective Harm Cases – no *single* such act seems to itself make a difference with respect to this outcome, and thus it is difficult to explain why there is reason for that individual to do it (or, depending on the case, to refrain) on any given occasion. We can call these cases “Individual Over Time” cases.

Warren Quinn turns Parfit’s Harmless Torturers into an Individual Over Time Case, which he calls “the Self-Torturer”.¹⁵¹ The “self-torturer” is hooked up to a Harmless-Torturers-style shock machine. Each week he has to make a choice: turn the dial on his machine up a single notch and receive a monetary reward, or leave the dial where it is and receive nothing. (He cannot turn the dial back). The problem is that each week he can argue: it won’t make

¹⁵¹ Quinn, “The Paradox of the Self-Torturer.” For some other articles that link Collective Harm Cases and Individual Over Time Cases, see: Andreou, “Environmental Damage and the Puzzle of the Self-Torturer”, Elster, “Rationality, Morality, and Collective Action”, and Arntzenius and McCarthy, “Self Torture and Group Beneficence.”

any difference whether or not I turn up the dial – either way, I will feel exactly the same as I currently do, however bad that is – and so there doesn't seem to be any point in forgoing the extra money. But if he follows this reasoning each week, eventually he will be in excruciating pain, and would “gladly relinquish his fortune and return to 0” if he could.¹⁵²

This is not a Collective Harm Case. But there is a clear analogy. And, as Quinn points out, Individual Over Time Cases arise in everyday practical deliberation. He writes:

The self-torturer is not alone in his predicament. Most of us are like him in one way or another. We like to eat but also care about our appearance. Just one more bite will give us pleasure and won't make us look fatter; but very many bites will. And there may be similar connections between puffs of pleasant smoking and lung cancer, or between pleasurable moments of idleness and wasted lives.¹⁵³

Also, while Individual Over Time Cases are typically presented as a problem for Rational Choice Theory, rather than moral theory, we can easily construct examples that are clearly morally relevant. Simply take the Self-Torturer, for instance, and imagine that instead of turning the dial on himself week-by-week, he is turning the dial on another person. Or, take the smoking case and imagine that the smoker in question is, each time, exposing someone else to second-hand smoke.

A nice feature of my account is that it works in Individual Over Time Cases just as well as in Collective Harm Cases. Take the smoking example. Even if it is true that a single cigarette – or if we want, following Quinn, a single puff – cannot make the difference between one state of health and a worse one, on my account, it could *help* to bring it about that one is in a worse state of health. Acts of smoking could be part of what cause illness. So, according to my account, an act of smoking could play a *non-superfluous* causal role in bringing it about that one is sick (or, more sick) as long as, at the time at which you smoke the cigarette, it is still possible that one won't get sick (or won't get *more* sick) due, at least in part, to a lack of smoking. If this is the case then one has as a reason to *refrain* from smoking another cigarette – or, taking another puff – that doing so could help to bring it about that you become ill. Smoking even just one cigarette could play a non-superfluous role in this

¹⁵² Quinn, “The Paradox of the Self-Torturer,” p. 79.

¹⁵³ *Ibid.*, p.79.

undesirable and avoidable outcome, and so there is prudential reason (and, depending on what else we say, perhaps moral reason) not to smoke it.¹⁵⁴

5.5 Strength of the Reason

An important question that I have left open with respect to my proposal is: how are we to evaluate the strength of the reasons that one gets from helping? It would be problematic if these reasons were so weak in Collective Harm Cases that they were easily outweighed by other considerations. I am not going to give a proposal here for how to evaluate the strength of the reason. Instead, I will leave this as an area in which future work is needed. But I do want to say something very brief about it.

I don't think there is any reason to assume that these reasons would be easily outweighed in Collective Harm Cases. There was a concern along these lines with the appeal to one's chance of making a difference in triggering cases, but we are dealing with a different sort of consideration here. Suppose, for example, that it turns out that Harmless Torturers is a triggering case. In particular, suppose that each act of turning the dial makes a real, though miniscule and barely perceptible difference in the pain level of the victim. With your turn of the dial, the victim would be slightly worse off than he would be without it. Now, while this would – I think – be a consideration in favor of refraining from turning the dial, this consideration might be very easily outweighed. If turning the dial could make a small but clearly perceptible and not insignificant difference for the better to someone (e.g. your Aunt Neta – who has been feeling rather stiff recently – would receive a rejuvenating massage), this might outweigh the miniscule, barely perceptible difference in pain to the victim. But, on my proposal, there is a *distinct* sort of reason to refrain from turning the dial: doing so might play a significant, non-superfluous role in bringing about very serious harm. By turning the dial, you are taking a non-superfluous step closer to the production of a serious and avoidable bad outcome (namely, someone being in severe pain), and this gives you reason to refrain. This is a reason that you have quite apart from whether or not your individual act itself makes a tiny difference to the level of pain. And at least on the face of it, there is no reason to think that it would be

¹⁵⁴ Not every proposed solution to the Problem of Collective Harm could even be proposed as a solution to the problem in Individual Over Time Cases. In particular, the Fairness approach (which has to do with there being a group of people under a collective obligation) and the Participation approach (which has to do with intentions to participate in collective action) clearly could not be proposed as solutions to that problem.

outweighed by the fact that turning the dial would give Aunt Neta this small but not insignificant benefit that she would not otherwise have received.

Now, it is not obvious how we should evaluate the strength of reasons having to do with how an act might help, nor is it obvious how they should weigh up against other considerations, including difference-making considerations. Some factors have obvious effects on the strength of the reason. For instance, the more serious the outcome in question, the stronger the reason you would get from helping. If the men in Drops of Water are suffering mild-to-moderate dehydration, the reason will not be as strong as it is if they are suffering severe dehydration. It also seems plausible that the larger the causal (or constitutive, etc.) role your act could play, the stronger the reason to act in the relevant way will be. It's not clear, though, how exactly these two factors should weigh up against one another. And there are many other factors that could affect the strength of the reason and that require more careful investigation. For instance, one possibility is that as the chance of the relevant outcome coming about gets closer to 50%, the reason that one gets from being able to help to bring it about (or, depending on the case, help to prevent it) gets stronger. That is, it could be that the more up in the air things are with respect to whether or not the outcome will occur, the stronger the reason to act in the relevant way would be. But it is not obvious that this is the right way to go.

5.6 Moral Obligation

A related question that I have left open is: should we say that one *ought* – or is morally obligated – to act in the relevant ways in Collective Harm Cases? As I explained in Chapter 1, I think it is important that we tackle the Problem of Collective Harm by *first* trying to figure out how it is that there is at least some reason for action.¹⁵⁵ That is the basic question that the Problem of Collective Harm poses, and if we can answer it, then it seems to me that the most puzzling and unique part of the problem is resolved. But this does not mean that we shouldn't then go on to discuss questions concerning our individual moral obligations in these cases.

What one says about such questions, though, will likely depend on one's wider views about moral obligations and how they are generated. One thing worth exploring is whether and how my approach could be incorporated into various moral theories. It seems to me that my account of how an act can help without making a difference provides a crucial tool needed for many *non-*

¹⁵⁵ Not just any reason, of course. We needed a reason that is morally relevant and that connects in some appropriate way with the morally significant outcome in question.

consequentialist theories to be able to handle Collective Harm Cases. It should be clear from Chapter 2 why I think this. If one's turning up the dial on the torture machine (in order to get a free massage) could play a non-superfluous, significant role in harming the victim, it makes sense to say – as the Kantian would want to – that in doing so one is treating him as a *mere means* to one's ends. If by adding your pint you are *helping* to alleviate the wounded men's suffering, then it is not implausible to propose – as Virtue Theorists might want to – that this is what the virtuous person would do in these circumstances. And if by using pesticides in my garden I would be helping to cause serious illness in a portion of the population, then we could say – as the Scanlonian Contractualist would want to – that each of the harmed individuals has an objection to a principle permitting me to use pesticides in my garden.¹⁵⁶ So, it seems to me that my proposal allows a range of non-consequentialist theories to get up and running in Collective Harm Cases.

A different question is what consequentialists should think about my proposal. Of course, there are various consequentialist views. But suppose that you are a committed consequentialist just in the sense that you are committed to the view that all that matters morally are consequences, or states of affairs. So far, what you have taken this to suggest is that what one ought to do is maximize expected outcome. My discussion calls into question whether this is what you – the person who thinks that all that matters morally are states of affairs – should think. In particular, I have argued that we need a broader understanding than the difference-making conception of what counts as a morally significant relation between an act and an outcome. On my proposal, things need not turn on whether or not you act in that particular way, in order for your act to play a morally significant causal role in bringing about a certain state of affairs.¹⁵⁷ If you agree with this proposal, then 'maximize expected outcome' cannot be the right way to go.

In any case, I mention these points only to note what I have not discussed. I am saving a proper discussion of questions concerning my account

¹⁵⁶ This is quite hand-wavy, but is sufficient for the present purposes. There are subtle questions to ask about how Scanlon's Contractualism handles Collective Harm Cases, and there is more to say about why my view would be useful there. Relevant to this topic are: Scanlon, *What We Owe To Each Other*, Chapter 5; Parfit, "Justifiability to Each Person"; and Scanlon, "Replies", pp. 431 – 434.

¹⁵⁷ Note that if you are a *non-consequentialist*, who thinks that other things matter besides states of affairs, this point applies to those other morally significant things as well. It is, I have argued, a mistake to think that one must *make a difference* to something in order to have contributed to it in a morally significant, non-superfluous sense. This applies whether we are talking about states of affairs or something else.

and moral obligation, and how my account interacts with various developed moral theories for another time.

Bibliography

- Andreou, Chrisoula. "Environmental Damage and the Puzzle of the Self-Torturer." *Philosophy & Public Affairs* 34 (2006): 95-108.
- Arntzenius, F. and D. McCarthy. "Self Torture and Group Beneficence." *Erkenntnis* 47 (1997): 127-144.
- Binmore, K. and A. Voorhoeve. "Transitivity, The Sorites Paradox, and Similarity-Based Decision-Making," *Erkenntnis* 64 (2006): 101-114.
- Bok, Hilary. "Freedom and Practical Reason," *Freedom and Responsibility*. Princeton: Princeton University Press, 1998, 130 – 166.
- Cullity, Garrett. "Pooled Beneficence," *Imperceptible Harms and Benefits*, Ed. M. Almeida Dordrecht: Kluwer Academic Publishers, 2000, 1 – 23.
- Eggleston, Ben. "Does Participation Matter? An Inconsistency in Parfit's Moral Mathematics." *Utilitas* 15 (2003): 92 – 105.
- Eggleston, Ben. "Should Consequentialists Make Parfit's Second Mistake?: A Refutation of Jackson." *Australasian Journal of Philosophy* 78 (March 2000): 1–15.
- Elster, Jon. "Rationality, Morality, and Collective Action." *Ethics* 96 (October 1985): 136 -155.
- Fara, Delia Graff. "Phenomenal Continua and the Sorites." *Mind* 110 (2001): 905 – 935. (Originally published under the name 'Delia Graff'.)

- Glover, Jonathan. "It Makes No Difference Whether or Not I Do It." *Proceedings of the Aristotelian Society* 49 (1975): 171 – 90.
- Goldman, Alvin I. "Why Citizens Should Vote: A Causal Responsibility Approach." *Social Philosophy & Policy* 16.2 (1999): 201 – 217.
- Gruzalski, Bart. "Parfit's Impact on Utilitarianism." *Ethics* 96 (1986): 760 – 83.
- Hall, Ned. "Two Concepts of Causation", *Causation and Counterfactuals*. Ed. Collins, Hall, and Paul. Cambridge, Mass: MIT Press, 2004, 225 – 276.
- Jackson, Frank. "Group Morality." In *Metaphysics and Morality: Essays in Honour of J.J.C. Smart*. Ed. Philip Pettit, et al. Oxford: Blackwell, 1987, 91 – 110.
- Jackson, "Which Effects." In *Reading Parfit*. Ed. Jonathan Dancy. Oxford: Blackwell, 1997, 42 – 53.
- Kagan, Shelly. "Do I Make a Difference?" *Philosophy and Public Affairs* 39.2 (2011): 105 – 141.
- Kant, Immanuel. *Groundwork of the Metaphysic of Morals*. Trans. and Ed. Mary Gregor. Cambridge University Press: Cambridge, 1997.
- Keefe, Rosanna. "Phenomena Sorites Paradoxes and Looking the Same". *Dialectica*. 65.3 (2011), pp. 327 – 344.
- Kolodny, Niko. "Instrumental Reasons." Draft, 11/16/11.
- Kutz, Christopher. *Complicity*. Cambridge: Cambridge University Press, 2000.

- Otsuka, Michael. "The Paradox of Group Beneficence." *Philosophy and Public Affairs* 20 (1991): 132 – 49.
- Parfit, Derek. "Comments." *Ethics* 96 (1986): 832 – 872.
- Parfit, Derek. "Justifiability to Each Person," *Ratio* XVI 4 (December 2003), 368 – 389.
- Parfit, Derek. *Reasons and Persons*. Oxford: Clarendon Press, 1984.
- Parfit, Derek. "What We Together Do." Draft, March 1988.
- Quinn, Warren. "The Paradox of the Self-Torturer," *Philosophical Studies* 59 (1990): 79-90.
- Raffman, Dianna. "Vagueness Without Paradox", *Philosophical Review* 103 (1994): 41-74.
- Regan, D. "Perceiving Imperceptible Harms: With Other Thoughts on Transitivity, Cumulative Effects, and Consequentialism", in M. Almeida (ed.), *Imperceptible Harms and Benefits*, Kluwer Academic Publishers, Dordrecht, 2000. pp. 49-73.
- Scanlon, Thomas. "Replies," *Ratio* XVI 4 (December 2003), 431 – 434.
- Scanlon, Thomas. *What we Owe To Each Other*. Cambridge, Massachusetts: Harvard University Press, 1998, Chapter 5.
- Scheffler, Samuel. "Individual Responsibility in a Global Age." *Social Philosophy and Policy* 12 (1995): 219 – 36.

Sinnott-Armstrong, Walter. *It's Not My Fault: Global Warming and Individual Moral Obligations*” Perspectives on Climate Change: science, Economics, Politics, Ethics. *Advances in the Economics of Environmental Research*, Volume 5 (2005): 293-315.

Strang, Colin. “What if Everyone Did That?” *Durham University Journal*, 53 (1960): 5 – 10.

Tännsjö, Torbjörn. “The Morality of Collective Action,” *The Philosophical Quarterly* 39 (April 1989): 221 – 228.