October 2018

The Mismatch Problem for Act Consequentialism

Robert Gruber
University of Massachusetts Amherst

THE MISMATCH PROBLEM FOR ACT CONSEQUENTIALISM

A Dissertation Presented

by

ROBERT GRUBER

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2018

Philosophy

THE MISMATCH PROBLEM FOR ACT CONSEQUENTIALISM

A Dissertation Presented

by

ROBERT GRUBER

Approved as to style and content by:

Fred Feldman, Chair

Peter Graham, Member

Katia Vavova, Member

Lyn Frazier, Member

Joseph Levine, Department Chair
Philosophy

DEDICATION

for Mom and Dad

ACKNOWLEDGMENTS

I would first like to thank my friends from the philosophy graduate program at UMass, especially Miles Tucker, Luis Oliveira, Scott Hill, Tricia Magalotti, Jordan Kroll, Julie Rose, and John Robison. I have learned so much from enjoyable discussions and paper swaps with each of them.

I am grateful to my many wonderful professors at UMass. In particular, I'd like to thank Ernesto Garcia for helping me think more carefully about group action, Christopher Meacham for clearing up my thinking about expected utility in Chapter 3, Alejandro Pérez Carballo for offering productive comments on an early draft of Chapter 5, and Philip Bricker for providing invaluable advice on how to understand vagueness in Chapter 6.

I would like to thank the members of my dissertation committee for their dedication to this project. I would like to thank Katia Vavova for her supportive, kind, and insightful suggestions along the way. Katia always knows how to best organize and strengthen an argument; I've learned so much about writing philosophy from her. I would like to thank Peter Graham for offering challenging cases for me to mull over and for helping me to sharpen my reasoning in all of these chapters. And I would like to thank Lyn Frazier for her service as an outside reader.

I am so thankful to friends and family for their support throughout my time as a graduate student. Specifically, Will Steffen for always being there for me, my parents who are always just a phone call away, and my partner Gabriella for celebrating each milestone with me, for lovingly supporting me when the going was tough, and for inspiring me to take time away from my desk to bike with her to the garden.

I save the most important for last. I could not have done this project without Fred Feldman, my mentor, philosophical role-model, and friend. I absolutely cannot thank Fred enough for his unwavering encouragement and dedication. While all the others have shaped my thinking, Fred has taught me how to do philosophy. In every sentence I write, I strive for the clarity, rigor, and persuasiveness that I admire in Fred's work. His influence may be found on every page of this dissertation.

ABSTRACT

THE MISMATCH PROBLEM FOR ACT CONSEQUENTIALISM

SEPTEMBER 2018

ROBERT GRUBER

B.S., UNIVERSITY OF WISCONSIN–MADISON

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Fred Feldman

I present the mismatch problem for Act Consequentialism, and I critically evaluate some popular solutions before offering my own solution to a specific version of the problem. The mismatch problem arises for Act Consequentialism when a group could have done better, but no individual in the group had an alternative with a better outcome. In such cases, the theory delivers mismatched verdicts: it condemns what the group does, but it cannot condemn any of the individual acts. In the first chapter of the dissertation, I explain exactly how this problem works. In the next four chapters, I identify a variety of cases that give rise to the mismatch problem, and I explain why the most popular strategies for modifying Act Consequentialism do not get around it. In the final chapter, I introduce a novel taxonomy of problem cases, and I introduce a 'cautious' version of Act Consequentialism that doesn't encounter the mismatch problem for a certain class of case.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

EXPLAINING THE MISMATCH PROBLEM

1.1 Introducing the Problem

If you're like me, you sometimes worry about contributing to certain large scale environmental problems. Take climate change, for example. When you drive home from work, you join billions of people throughout the world in the collective production of an alarmingly high concentration of atmospheric CO_2. Extreme weather events now occur more frequently, rising sea levels wash out developed coastlines, heat waves happen more often and with greater intensity.[1] People overall are suffering more than they otherwise would have under a stable and cooler climate system. And if we would just substantially reduce our emissions now, we would prevent some truly horrifying scenarios from happening in the future: the melting of the polar ice caps, widespread extinctions, and the displacement of billions of future people. So isn't it obviously wrong for you to carry on with business as usual, continuing to spew CO_2 from your tailpipe? Shouldn't you pursue green alternatives whenever possible?

A familiar puzzle arises here in connection with many forms of consequentialism. Under most ways of formulating the theory, you would have a moral obligation to reduce your emissions on some occasion only if doing so would have a better outcome than your not reducing emissions. But because of the scale of the problem, your tiny contribution doesn't seem to make a difference. What's the result of *your* driving to work instead of biking? It's seemingly negligible. Furthermore, no matter how

[1] For a comprehensive source of information on the effects of climate change, see Stocker et al. (2013).

you would act, not enough *others* would reduce their emissions; the bad outcomes of climate change would apparently be exactly the same. Climate change results because of the aggregate effects of trillions of tons of greenhouse gases building up in the atmosphere. Your emissions are apparently too small to have an impact. Though a large group of emitters has an alternative with a better outcome—and though you are a member of that group—it's not clear that *you* have an alternative with a better outcome. Traditional versions of consequentialism apparently cannot condemn your individual act.

Of course, this illustration of the situation assumes that you don't have much influence over the behavior of others. To be sure, in some cases an individual's act apparently makes a significant contribution to certain environmental problems. For example, if a popular celebrity goes on joyrides in her flashy sports car, others may be influenced by her example. The celebrity's contribution may ultimately have a big impact on the atmospheric concentration of CO_2 and the resulting harms of climate change. But in most cases, it seems that an individual's act of contribution makes no difference. In extreme cases, it might be that *no* individual's act makes a difference. No matter how any given individual would act—whether, for example, he or she would bike to work instead of driving—not enough others would act differently; the bad outcomes of climate change would still be just as bad. In such cases, each participant seems to have a plausible excuse: "the result would have been the same even if I had made no contribution."

This extreme version of the climate change case is just one of a family of related cases that highlights a difficulty for many forms of consequentialism. In all such cases, a group of people acts together to bring about a bad outcome, and though the group could have done something much better, no individual member of the group could have done any better by acting differently. In these cases, many forms of consequentialism deliver mismatched verdicts: the theories condemn the behavior of the group, but they cannot condemn any of the component individual acts. It is my aim in this dissertation to evaluate some of the various solutions that have been offered to address this problem. But first, it's important to get clear on the exact nature of the problem that I mean to be discussing. In this chapter, I will explain the problem that I'm interested in, and I'll distinguish it from several related problems.

1.2 Some Fundamental Assumptions

First, we may wonder: can it really be the case that a group of people brings about a suboptimal outcome while no individual could have made things better? Indeed, if there are such cases, we may recognize an immediate worry for one popular version of consequentialism. According to *Act Consequentialism*, an act is morally permissible just in case there's no alternative with a better outcome. Suppose a group of people brings about a bad outcome, and the group could have acted differently to produce a better outcome. Under a natural extension of Act Consequentialism that issues moral evaluation on the acts performed by groups, the group acts wrongly. But if no individual member of the group could have done better, then each member of the group acts permissibly according to Act Consequentialism. If such cases are possible, then Act Consequentialism paradoxically delivers mismatched verdicts between different levels of evaluation. The theory condemns the behavior at the group level, but this condemnation fails to trickle down to any of the individuals involved.

To see exactly how the problem may arise, consider the following case:

> *Snake Bite:* You and I are the only doctors working during the night shift at a hospital. A kid comes in who's been bitten by a venomous snake. Time is running out: we both know that the child will almost certainly live only if you run the diagnostic tests while I sprint to get the antivenin kit. Neither of us can do both of these things, though *you* could easily run the tests, and *I* could easily get the kit. But we don't do our respective tasks. I'm lazy and I would not get the antivenin even were you to run the tests. Since you know this, you don't bother to do your part in saving the child: doing so would be pointless since I wouldn't help out. So, instead, you administer a pain reliever. But I know that you are lazy as well; I know

> that you would not run the tests even were I to get the antivenin. Since I know this, I don't bother to do my part in saving the child either—you wouldn't help me out. So, instead, I sing a comforting song while the child dies. While it is very bad that the child dies, it would have been worse had only one of us acted in the way we in fact act. Had you run the tests while I sang a comforting song, the child would have suffered more. And had I sprinted to get the antivenin kit while you administered the pain reliever, the child would have suffered more.

We may represent the situation in Snake Bite by the following table.

	you run tests	**you give pain reliever**
I get antivenin	*best*	*worst*
I comfort victim	*worst*	***bad***

Table 1.1. The possible outcomes in Snake Bite

The very top row of the table illustrates your possible acts. What you actually do is administer the pain reliever (in bold), but you could have run the tests. The very left column of the table illustrates my possible acts. What I actually do is comfort the victim (in bold), but I could have retrieved the antivenin. Each combination of possible acts corresponds to an outcome, the value of which is displayed in the table. In those worlds where you run the tests and I get the antivenin, the child lives—the *best* outcome. In those worlds where you run the tests and I comfort the victim, the child dies without pain reliever—the *worst* outcome. In those worlds where you administer pain reliever and I get the antivenin, the child dies without a comforting song—tied for the *worst* outcome. And in the actual world, where you administer pain reliever and I sing a comforting song, the child dies—a *bad* outcome. Each of these outcomes is accessible to the two of us together; we could have brought about any of them. We actually bring about the *bad* outcome in the bottom right box (in bold), but we could have brought about the better outcome in the top left box had we both acted differently.

It may seem that this case has little to do with the case of anthropogenic climate change. But the relationship between the outcomes available to the group and the outcomes available to the individuals in Snake Bite is in many ways structurally similar to the relationship between the various climatic outcomes available to the group and the climatic outcomes available to the individual emitters. By focusing on the essential features that give rise to this relationship in Snake Bite, we may discover exactly how the problem for consequentialism may arise more generally. This will show that the climate change case is just one of many suitably related cases, all of which give rise to the same problem for consequentialism.

To explain why Snake Bite meets the conditions of the sort of case that I'm interested in, I must highlight some important assumptions. First, I assume that an act has an alternative if there's something else that the agent who performs the act could have done. In Snake Bite, I assume that your act of administering the pain reliever has an alternative because you could have done something else; you could have run the tests. Similarly, I assume that my act of comforting the victim has an alternative because I could have done something else; I could have retrieved the antivenin.

Second, I assume that the outcomes of alternatives are to be determined counterfactually: an act has an alternative with a better outcome if, were that alternative performed, the result would have been better than the result of the act that's actually performed. In Snake Bite, to determine whether your act has an alternative with a better outcome, we must consider what would have resulted had you run the tests. Similarly, to determine whether my act has an alternative with a better outcome, we must consider what would have happened had I retrieved the antivenin. Given these assumptions, notice that neither of our individual acts has an alternative with a better outcome in Snake Bite. Suppose you had run the tests. Since I would have comforted the victim, that locates the outcome of your alternative in the bottom left box of the table. So you don't have an alternative with a better outcome. Had *you* acted differently in Snake Bite, the result would have been worse. Or suppose that

I had retrieved the antivenin. Since you would have administered the pain reliever, that locates the outcome of my alternative in the top right box. So I don't have an alternative with a better outcome either. Had *I* acted differently in Snake Bite, the result would have been worse.

Third, I assume that groups of people do things and that what a group does can be understood as a set of individual acts. For convenience, I'll use the term "group act" to refer to any set of individual acts. There are certainly more sophisticated accounts of group action, but for now, I'm going to proceed as if it's perfectly fine to reduce what a group does to the set of what the individual members do. A deeper metaphysical analysis of group action will have to be studied elsewhere. For our purposes here, I assume that groups of people do things and that sometimes they could have done other things, and talking about group acts is an easy way to cash out these assumptions. In Snake Bite, we perform this group act: [I comfort victim, you give pain reliever]. But we could have done something else—in particular, we could have done this group act: [I get antivenin, you run tests]. Along with the foregoing assumptions, this reveals that we together have an alternative group act with a better outcome: we could have performed the group act [I get antivenin, you run tests], and had we done so, the resulting outcome would have been better than the result of what we actually do.

So we can see how the sort of case that allegedly highlights a difficulty for consequentialism is possible. As Snake Bite and the accompanying assumptions illustrate, there are cases in which a group of people acts together to bring about a bad outcome, though they could have done something much better, and yet no individual member of the group could have done any better by acting differently.

The problem that I mean to discuss concerning such cases arises in connection with Act Consequentialism. The problem results under the assumption that Act Consequentialism delivers verdicts on the moral status of group acts. If we make this assumption, then Snake Bite contains a morally impermissible group act with no accompanying impermissible individual acts. The group act [I comfort victim, you give pain reliever] is morally impermissible because it has an alternative with a better outcome. But according to an intuitively plausible assumption, if a moral theory says that what a group does is morally wrong, then something must be said against the individuals involved. Accordingly, proponents of consequentialism want to reject Act Consequentialism in favor of a version of consequentialism that is able to condemn, in some way, the individual acts.

In general, the *mismatch problem* arises for a moral theory whenever the theory delivers mismatched verdicts between a group act and the individual acts that compose it. As I've just demonstrated, under some intuitive assumptions the mismatch problem arises for Act Consequentialism in connection with cases in which a group of people acts together to bring about a bad outcome, though they could have done something much better, and yet no individual member of the group could have done any better. For convenience, let's say that any such case in which a group could have done better though each individual could not have done any better is a *collective harm.*[2] I can then more succinctly state the problem I intend to discuss : I am interested in the mismatch problem for Act Consequentialism in connection with collective harms.

It will be helpful to consider even more carefully the assumptions under which this problem arises. One foundational assumption is that groups of people perform actions. The possibility of collective harms requires this assumption. But once it is recognized that groups can do things, and that these things have consequences, a second foundational assumption emerges: it is natural to assume that a slightly extended version of Act Consequentialism should be true for group acts. Just like

[2] I borrow this terminology from Julia Nefsky; see Nefsky (2012b).

individuals, groups should do the best they can. According to *Act Consequentialism+*, any act (whether it be an individual act or a group act) is morally permissible just in case there's no alternative with a better outcome. In this chapter, it will be helpful from here on out to emphasize that the mismatch problem for Act Consequentialism is really more perspicuously understood as a mismatch problem for Act Consequentialism+. In the rest of this introductory chapter, when I engage with authors who mean to be discussing consequentialist moral theory as it applies simply to individuals, I'll use 'Act Consequentialism'. When I engage with authors who mean to be discussing the natural extension of consequentialist moral theory to groups, I'll use 'Act Consequentialism+'. It's only under Act Consequentialism+ that one and the same consequentialist moral theory may deliver mismatched verdicts between a group and individual level. We must assume that something like Act Consequentialism+ provides a normative evaluation of the group act. Without this assumption, the mismatch problem cannot arise.[3]

Both of these foundational assumptions might be challenged. For example, someone might object "But groups don't do anything—only *agents* do things, and groups are not agents."[4] I'm not sure what to make of this objection. It seems fairly plain that all sorts of groups do things. We talk this way all the time. The German football team won the 2014 World Cup. The United States electorate put Trump in office. We comforted the victim while she died in Snake Bite. But someone might press that the common ways of talking are misguided; we shouldn't infer from the way that we speak about groups whether the aforementioned groups of people actually do things. As I see it, defending this revisionist approach constitutes a solution to the problem

[3] In subsequent chapters, I drop this notational distinction. In most places in this dissertation, it will be clear from the context when I am discussing a version of Act Consequentialism that is extended to cover the evaluation of group acts.

[4] Stephanie Collins, for example, has defended some restrictions on group agency; see Collins (2013). Though Collins doesn't deny the possibility of group action in some cases, she would deny that *we* do anything in Snake Bite.

of mismatched verdicts, though—since I believe that groups do perform actions—it's not a solution that I'm interested in pursuing.

A more compelling objection to the second foundational assumption of the mismatch problem comes from Michael Zimmerman.[5] The mismatch problem arises under the assumption that Act Consequentialism should be naturally extended to deliver verdicts on the moral status of group acts—that something like Act Consequentialism+ is true. But this seems to require that groups have moral obligations. According to Zimmerman, however, groups of people probably do not have moral obligations. Zimmerman grants that groups can achieve outcomes with so-called "deontic value": groups can bring about various states of affairs with the sort of value that consequentialists think must be maximized. But Zimmerman disputes the idea that group acts can be right or wrong—at least in the sense that implies obligation. Here's the relevant passage:

> ...it might be that we should say that an agent, to be the subject of obligations, must be a *moral* agent; if not, we may have to accept that nonhuman animals have obligations (for they, too, may be able to achieve outcomes that have a certain deontic value). What is involved in an agent's being a moral agent is unclear, but arguably it must have the capacity to conceive of its action in moral terms, and it seems safe to say that no group with more than one member (as opposed to individuals in the group) has this capacity. Furthermore, some have contended that, for something to be obligatory, it must be something about which decisions can be made, and that joint outcomes achievable by groups with more than one member (as opposed to the individual component outcomes achievable by the individual members of the group) are not entities of this sort. If either of these points is correct, then, despite the fact (if it is a fact) that groups with more than one member can achieve an outcome that has deontic value and, often, has a degree of deontic value that no outcome achievable by some proper subgroup has, they cannot be obligated to achieve this outcome.[6]

[5] Zimmerman discusses the mismatch problem in Chapter 9 of Zimmerman (1996).

[6] Zimmerman (1996, 262)

The quoted passage contains several considerations against attributing wrongdoing to what the two of us together do in Snake Bite. Perhaps *we* cannot conceive of our action in moral terms (though *I* can, and *you* can). Perhaps *we* cannot make decisions about what *we* will do (though I can make decisions about what *I* will do, and *you* can make decisions about what *you* will do). If either of these conditions counts against attributing a moral obligation to our group, then this would undermine the possibility of a mismatch problem for Act Consequentialism+ in connection with the case. A mismatch arises in Snake Bite if the case contains a group act that's wrong. But if Act Consequentialism+ is false, and only an individualistic version of consequentialist moral theory is true, then this verdict does not emerge. If talk of morally wrong group acts make no sense, then there's no possibility of a mismatch—there's no group level verdict to conflict with the individual level verdicts.

I will not evaluate the arguments against the possibility of morally wrong group acts here. Rather, at this point it's important for us to see that if group acts are not subject to moral evaluation under a natural extension to Act Consequentialism, then this constitutes a solution to the problem of mismatched verdicts—though not a solution that I'm interested in pursuing. Perhaps a related problem persists even if there's no mismatch, but then this problem is distinct from the problem that I want to discuss.[7]

1.3 What's Not Essential to the Mismatch Problem

It's important to reflect on the features of the problem that I want to discuss so as to distinguish it from related but distinct problems. We may proceed by making a couple observations about Snake Bite.

[7] Zimmerman seems to think that there's still a problem, even if it's not the mismatch problem. See Chapter 9 of Zimmerman (1996).

1.3.1 An Observation about Permissibility 'All the Way Down'

First, notice that if we imagine a case similar to Snake Bite, but in which you are the only doctor in the hospital, then if you had behaved as you do in Snake Bite, your act would not have been wrong. Similarly for me. To see this, suppose that you are the only doctor in the hospital when the child comes in—say that I'm on vacation. Just as in the original Snake Bite case, suppose that you administer the pain reliever. Since the child will almost certainly live only if someone runs the diagnostic tests while someone else gets the antivenin kit, and since you cannot do both of these tasks, it is not possible for you to save the kid without me around. Thus, in the modified version of Snake Bite we are now imagining, the situation looks like what's depicted in Table 1.2.

	you run tests	**you give pain reliever**
	worst	***best***

Table 1.2. The possible outcomes in the modified Snake Bite

The very top row of Table 1.2 illustrates your possible acts. What you actually do is administer the pain reliever (in bold), but you could have run the tests. The very left column of the table illustrates the fact that I'm not around. Accordingly, there's nothing that I can do to affect the possible outcomes in this situation. In those worlds where you run the tests, the child dies without pain reliever—the *worst* outcome. In those worlds where you administer pain reliever, the child dies in less pain—in this situation, that's the *best* outcome. Each of these outcomes is accessible to you; you could have brought about either of them. You actually bring about the *best* outcome in the right box (in bold), but you could have brought about the worse outcome in the left box.

Notice that, in this situation, your act is morally permissible according to Act Consequentialism+. You don't have an alternative with a better outcome. So, when we think about the original Snake Bite case, we notice that your action of administering the pain reliever is morally permissible ‘all the way down’: not only do you actually act permissibly according to Act Consequentialism+ when I’m around, but also *you would have acted permissibly had I not been around.* A similar line of reasoning establishes that my action is also morally permissible ‘all the way down’.

1.3.2 An Observation about Act Types

A second feature of Snake Bite that’s important to notice depends upon a distinction between act tokens and act types. An *act token* is a concrete action taken by a specific agent at a specific time and place. Act tokens are not repeatable. Snake Bite involves several act tokens. One of these is your administering the pain reliever to this particular child shortly after she arrives at the hospital. Another act token is my singing some particular comforting song in the room where the child dies. A third is our performing the group act composed of the two of our act tokens taken in a set together. Notice that there’s very little in common between your act token and my act token. You neglect to run tests by administering a pain reliever, and I neglect to get antivenin by singing a song. Thus, our act tokens do not appear to be of the same *act type*, or general way of acting. We can imagine a variant of Snake Bite in which a third doctor is also at the hospital with us, and we can imagine that she also neglects to run tests by administering a pain reliever. In this variant of the case, two act tokens of the same act type are performed; there are two ‘tokenings’ of the type *administering pain reliever to a child who will soon die.* But in the original Snake Bite case, it may be difficult to come up with some single description under which both of our individual act tokens fall.

In some situations, more than one person performs an act token of the same act type, and though a single tokening of this act type would not be bad, the result of some large number of people tokening this act type is bad overall. Some situations of this sort include the lawn crossing problem (many people take a shortcut across a lawn on campus, killing the grass and damaging the overall appearance of the landscape), the voting problem (more than enough people fail to vote for the better of two candidates, leading to the victory of the inferior candidate), and the vegetarianism problem (many people purchase factory farmed chicken, thereby creating a market demand for cheap meat, and many new chickens suffer in factory farms as a result). Say that an act type is *repeatedly bad* if the performance of one act token of that type doesn't have any bad results, but enough repeated performances of act tokens of that type has bad results. Each of the just-mentioned problems involves a repeatedly bad act type: crossing the lawn, not voting, and purchasing factory farmed chicken.

The noteworthy feature of Snake Bite is that, under several natural assumptions, the case doesn't involve a repeatedly bad act type. Because I would not get the antivenin otherwise, you perform an act token of this type: *administering pain reliever to a child who will soon die.* Suppose that many others perform act tokens of this type. We may imagine either that many others alleviate the pain of many other children, all of whom will soon die, or that many others administer a dose of pain reliever to one and the same dying child. In the first case, the result is not obviously a bad outcome—the pointless suffering of many children is averted. In the second case, the result is not obviously a bad outcome either—overdosing on pain relievers doesn't harm a child who will die soon anyway.

My act token does not appear to be of a repeatedly bad act type either. Because you would not run the diagnostic tests otherwise, I perform an act token of this type: *singing a comforting song to a child who will soon die.* If many others sing to many other dying children, the result is not obviously bad. And if many others sing to the same dying child, the result is probably better than if it's only me singing.

On the other hand, one might describe each of our act tokens in Snake Bite as a type of *behaving neglectfully.* Perhaps this is a repeatedly bad act type? But notice that, as is represented in Table 1.1, the child dies under even one instance of neglectful behavior in Snake Bite. This shows that even one tokening of behaving neglectfully has bad results, and so this act type is not *repeatedly* bad.

1.3.3 Two Lessons

In summary, we may draw two lessons about the mismatch problem, one from each observation of Snake Bite. The first lesson is that the problem I'm interested in doesn't depend upon our individual acts being wrong on their own—the problem of mismatched verdicts arises in Snake Bite though neither of us would have acted wrongly were the other absent from the situation. The problem arises even though each act is permissible 'all the way down'. The second lesson is that the problem I'm interested in doesn't require the presence of repeatedly bad act types. In the following section, I will refer to each of these lessons in order to distinguish the problem I mean to be discussing from related problems.

1.4 Related Problems

There are several related problems that must be carefully separated from the problem that I am interested in. Many philosophers seem to be discussing the problem of mismatched verdicts, but they do so by considering cases that involve inessential and distracting features. Accordingly, these philosophers tend to veer off into discussing different, but related problems—or perhaps these philosophers intend to be discussing different problems in the first place, though they use cases that are similar to the ones I am interested in discussing. In this section, I draw from the two lessons above to distinguish the mismatch problem from related problems.

Recall that the problem of mismatched verdicts for Act Consequentialism+ arises in connection with collective harms: cases in which a group could have done better though each individual could not have done any better. In one widely discussed collective harm case, each of two individuals performs an act that's thought to be wrong. But because of the way the two acts overdetermine a bad outcome, each act becomes morally permissible under Act Consequentialism+ when both acts are performed together. Thus Act Consequentialism+ apparently allows two wrongs to cancel each other out. Consider

> *Two Shooters*: X and Y are sharpshooters. They shoot at me simultaneously, their bullets striking the same fatal location in my chest. Neither of them could have prevented the other from shooting, and either shot is sufficient for my immediate death. In other words, were Y not to shoot, X still would, and I would immediately die. Similarly, were X not to shoot, Y still would, and I would immediately die.

Like Snake Bite, Two Shooters meets the conditions of the sort of case that we're interested in. It's a collective harm because the bad outcome that X and Y together bring about (my death) could not have been prevented by either of X or Y individually, but could have been prevented by X and Y together. When we think about what would have happened had X holstered his weapon, we are to imagine a counterfactual world in which Y shoots me just as he does in the actual world. Similarly, when we think about what would have happened had Y holstered his weapon, we are to imagine a counterfactual world in which X shoots me just as he does in the actual world. Accordingly, Act Consequentialism+ fails to condemn either one of their individual acts. Y doesn't have an alternative with a better outcome: X would shoot and instantly kill me even if Y were to holster his weapon instead. And X doesn't have an alternative with a better outcome either: Y would shoot and instantly kill me even if X were to holster his weapon instead. I die no matter what either of X and Y, individually, does. And yet their acts together bring about my death. There's an alternative group act with a better outcome, and so X and Y together act wrongly according to Act Consequentialism+.

1.4.1 Parfit, Jackson, and Two Shooters

Derek Parfit discusses Two Shooters in the third chapter of *Reasons and Persons* in an attempt to dispel a mistake in moral mathematics.[8] The so-called "Second Mistake" involves one of the standard assumptions underlying Act Consequentialism. The assumption (which I mentioned earlier) is that the outcomes of alternatives are to be determined counterfactually: an act has an alternative with a better outcome if, were that alternative performed, the result would have been better than the result of the act that's actually performed. In Two Shooters, this assumption leads to the conclusion that X's shooting me doesn't have an alternative with a better outcome; same with Y's shooting me. Parfit argues that this assumption is mistaken. In connection with Two Shooters, it leads to an "absurd conclusion":

> Suppose that we make the Second Mistake. We assume that, if an act is wrong because of its effects, the only relevant effects are the effects of this particular act. Since neither X nor Y harms me, we are forced to the absurd conclusion that X and Y do not act wrongly.[9]

Parfit apparently thinks that Two Shooters poses a problem for the standard way of thinking about Act Consequentialism. The problem is that standard Act Consequentialism delivers an absurd conclusion: X and Y do not act wrongly.

In order to see whether Parfit is discussing the mismatch problem, there are some important unclarities in his presentation that we must sort out. First, we must become clear about what the absurd conclusion is. Second, we must understand why Parfit believes it is absurd.

Under the assumption that groups do things, there are three acts in Two Shooters, and it will be helpful to carefully distinguish them. There are two individual acts: X shoots me, and Y shoots me. And there's the group act, [X shoots me, Y shoots me]. Thus, when we think about what X and Y do, we may be thinking about the

[8] See Parfit (1984).

[9] (Parfit, 1984, 70)

two individual acts, or we may be thinking about the group act. Accordingly, when Parfit says that Act Consequentialism's absurd conclusion in Two Shooters is that "X and Y do not act wrongly", there are two possibilities:

> *No Individual Wrongdoing (NIW)*: neither individual act—neither X's shooting me nor Y's shooting me—is morally wrong.
> *No Group Wrongdoing (NGW)*: the group act, [X shoots me, Y shoots me], is not morally wrong.

Parfit is not discussing the mismatch problem if he believes that Act Consequentialism implies NGW. The mismatch problem arises for Act Consequentialism+ in Two Shooters only if the theory says that the group act is morally wrong. If Parfit takes the absurd conclusion to be NGW, then Parfit is not concerned with the problem that I'm interested in discussing. Thus, in order to see whether there's an interpretation of the quoted passage according to which Parfit is discussing the mismatch problem, let's assume that he believes Act Consequentialism implies NIW—not NGW.

An unclarity remains: why does Parfit believe that NIW is absurd? Parfit himself doesn't elaborate on why he rejects NIW. Since different reasons for thinking that NIW is absurd give rise to slightly different problems in connection with Two Shooters, it's important that we reflect on some possible reasons. Only one of the reasons I will consider corresponds with the mismatch problem.

Frank Jackson takes Parfit's reluctance to accept NIW to be a matter of its counterintuitive nature.

> It is evident that something wrong *happens*... but more than that is evident: something wrong is *done*. (It would be quite wrong to think of [Two Shooters] as being one of a natural misfortune, like a flood.) But if what X does and what Y does is not wrong in the over-determination case..., what actions *are* wrong...?[10]

[10] (Jackson, 1987, 100)

As Jackson points out, the situation in Two Shooters certainly smacks of wrongdoing. (Perhaps it's because, as I will point out in the next section, each shooter would have acted wrongly were the other not around to overdetermine the outcome of his act.) But Act Consequentialism cannot locate the wrongdoing in X's or Y's individual acts. As Jackson sees it, the problem in Two Shooters is that Act Consequentialism implies NIW, which apparently conflicts with the intuition that something wrong is done.

Jackson's interpretation of the problem in Two Shooters is distinct from the mismatch problem. This is best demonstrated by reflecting on Jackson's proposed solution. Jackson argues that the intuition about wrongdoing in Two Shooters can be accommodated by accepting the existence of morally wrong group acts:

> What foxes us when first presented with [Two Shooters]. . . is tunnel vision. We are immediately struck by the intuition that something wrong is done; we then search around for a wrong action, restricting ourselves, without fully realizing it, to the individual actions in the cases, and so have no choice but to say that X's and Y's actions. . . are wrong. There are no other starters among the individual actions. But if we enlarge the class of actions which may be morally evaluated to include group actions as well as individual actions, we can say that the agents' group actions, though not their individual actions, are wrong.[11]

Jackson argues that NIW is not counterintuitive after all. The intuition that something wrong is done in Two Shooters can be accommodated by the thought that the group act [X shoots me, Y shoots me] is morally wrong. Jackson endorses Act Consequentialism+. According to Jackson, this resolves the absurdity of NIW that Parfit is concerned with.

Ironically, Jackson's *solution* to what he takes the problem to be in Two Shooters is what *gives rise* to the problem that I'm interested in discussing. I think that the counterintuitive nature of NIW doesn't get to the heart of the problem in Two Shooters. In my view, NIW appears absurd because it results in a mismatch of verdicts

[11] (Jackson, 1987, 100)

within Act Consequentialism+. Suppose we agree with Jackson that the group act [X shoots me, Y shoots me] is morally wrong according to Act Consequentialism+. If Act Consequentialism+ also implies NIW, then the result is a paradoxical mismatch of verdicts within the theory: Act Consequentialism+ says that a group act is wrong though no individual act is wrong. It's very natural to believe that if a moral theory says that what a group does is morally wrong, then something must be said against each of the involved individuals individually. It may appear absurd to suggest otherwise.

Thus, there are at least two different reasons why Parfit may believe that NIW is absurd. The first conforms with Jackson's interpretation of the problem in Two Shooters: NIW is absurd because it conflicts with the intuition that something wrong is done. The second conforms with my characterization of the mismatch problem: NIW is absurd because it conflicts with the fact that the group acts wrongly.[12] It's important to see that Jackson's problem is distinct from the mismatch problem. Of course, the problems are related. But in some sense the mismatch problem is deeper. Thus, I'm not directly interested in Jackson's problem. Instead, I'm interested in the mismatch problem.

1.4.2 The Cancellation Problem

T. Zamir offers a third reason to think that NIW is absurd. According to Zamir, the problem in Two Shooters depends on the fact that shooting me would be wrong were it performed on its own. Because X and Y each overdetermines the bad outcome of the other's act, this paradoxically makes each otherwise wrong act morally permissible under Act Consequentialism.

[12] In earlier sections of reasons and persons, Parfit seems to be discussing the mismatch problem (though he classifies things somewhat differently). For example, see section 21 of Parfit (1984).

To see this more clearly, imagine what the situation would be like in Two Shooters if Y weren't a part of it; suppose we erase Y from the case. In the resulting 'One Shooter (X)' case, Act Consequentialism says that X acts wrongly in shooting me. Had X holstered his weapon in One Shooter (X), I would not have died, so in this version of the case X has an alternative with a better outcome. Similarly, in the 'One Shooter (Y)' version of the case, Act Consequentialism condemns Y's act. Reflection on these variations of the case shows that each of the individual acts in Two Shooters would be wrong were it performed on its own. But because of the way the two individual acts interact, Act Consequentialism says that each of X and Y acts morally permissibly when they both shoot—their would-be-wrong acts cancel each other out. This seems odd to Zamir, especially because it doesn't seem like there's any morally relevant difference between what X does in One Shooter (X) and what X does in Two Shooters. In both cases, X promotes a bad outcome. The only difference between the cases is that, in the latter case, Y promotes this same bad outcome. But an Act Consequentialist shouldn't think two promotions of the same bad outcome makes for two permissible acts.[13]

Notice that Zamir's problem rests on the idea that there's no morally relevant difference between what X does in One Shooter (X) and what X does in Two Shooters. But surely an Act Consequentialist will reject this idea. The accessible outcomes are different between the cases, and this should make all the difference to the moral evaluation of X's act under Act Consequentialism.

Furthemore, notice that Zamir's problem is not the mismatch problem. Recall the first lesson we drew above regarding Snake Bite. The problem of mismatched verdicts arises in Snake Bite though neither of us would have acted wrongly were the other

[13] (Zamir, 2001, 160)

absent from the situation. Thus, unlike Zamir's problem, the problem I'm interested in doesn't depend upon the individual acts being wrong on their own.

1.4.3 The Repeated Performances Problem

I've been discussing Two Shooters, but there's another widely discussed type of case in which the mismatch problem may be seen to arise. This type of case involves a large number of people, each of whom has the ability to perform the same repeatedly bad act type. If enough of them act in this way, the result will be bad. But no individual tokening of the repeatedly bad act type produces an effect that's sufficient on its own to make a difference. In a recent article, Shelly Kagan centers his discussion of the mismatch problem on this type of case.

> These cases appear to have the following structure: A certain number of people—perhaps a large number of people—have the ability to perform an act of a given kind. And if a large enough group of people do perform the act in question then the results will be bad overall. However—and this is the crucial point—in the relevant cases it seems that it makes no difference to the outcome what any given individual does. And this is true regardless of whether others are doing the act or not. Thus, if enough people do perform the act the results are bad overall; but for all that, it remains true of each individual agent that it makes no difference to the overall results whether or not they perform the act in question.[14]

In this passage, Kagan highlights the essential features of the cases that he's interested in discussing: (1) each of several people could perform an act of the same act type; (2) the performance of one act token of that type doesn't have any bad results, but enough repeated performances of act tokens of that type has bad results—the act type is repeatedly bad; (3) it doesn't matter how many others will token the repeatedly bad act type, it appears to make no difference to the overall results whether or not one more act token of the repeatedly bad act type is performed.

[14] (Kagan, 2011, 107)

Almost every discussion of the mismatch problem I've seen in print satisfies at least condition (1) of Kagan-style cases.[15] But notice that Snake Bite doesn't satisfy any of these conditions. In Snake Bite, we don't appear to be performing two tokens of the same act type. And, according to the observation in section 1.3.2 above, even if we do perform the same act type, it's not repeatedly bad. So none of the three conditions above is a necessary condition of the mismatch problem.

More importantly, the satisfaction of all three of Kagan's conditions is not sufficient for the mismatch problem. Notice that in the quoted passage, Kagan doesn't make it clear whether we are supposed to be imagining only those cases in which a bad outcome actually results, or whether Kagan also means to include cases in which a bad outcome doesn't actually result, though it could. For a case of the latter sort, consider

> *Drought*: Because it hasn't rained in a while, the town of Amherst issues a watering ban. If enough people continue to water their lawns, the town reservoir will soon be depleted. You and your neighbor have long driveways; you are hidden from the road by trees. The two of you continue to water your lawns, and no one finds out. Everyone else in Amherst obeys the watering ban. The result is that the reservoir remains full enough throughout the period of the drought that no one is negatively affected.

Drought satisfies the three conditions of a Kagan-style case. First, there are repeated performances of the same act type—you and your neighbor each performs, several times, act tokens of the type *watering*. Second, this act type is repeatedly bad. One act of watering doesn't have any bad results, but enough repeated performances will deplete the reservoir. Third, it doesn't matter how many others will water, it appears to make no difference to the health of the reservoir whether or not one more act of watering is performed.

[15] Such cases include Two Shooters, Regan (1980)'s button-pressing cases, Zimmerman (1996)'s Vincent and Virgil cases, Feldman (1980)'s thin ice and fumigation cases, and Pinkert (2015)'s Two Factories case. Other such cases include Parfit (1984)'s Harmless Torturers, and the widely-discussed lawn crossing problem, voting problem, and vegetarian problem.

But notice that Drought cannot give rise to the mismatch problem for Act Consequentialism+. Recall that the mismatch problem arises for a moral theory whenever the theory delivers mismatched verdicts between a group act and the individual acts that compose it. In Drought, neither you nor your neighbor could have done any better. You both keep your lawns healthy, and the reservoir remains healthy even though each of you waters. Accordingly, Act Consequentialism+ says that you together have acted morally permissibly. Furthermore, you don't have an alternative with a better outcome in Drought, and neither does your neighbor. So Act Consequentialism+ says that each of you individually acts morally permissibly. Since Act Consequentialism+ delivers the same type of verdict on the group act and the individual acts in Drought, the mismatch problem does not arise.

That's not to say that Drought contains no problems whatsoever for Act Consequentialism. There is, of course, a common intuition that arises in connection with Drought and all other Kagan-style cases. This is the Kantian intuition captured in the question 'What if everyone did that?' In Drought, it would be natural for other residents in Amherst to be upset with what you do. It is wrong for you to water your lawn, they may think, because it would be disaster if everyone did it. The fact that Act Consequentialism fails to capture this intuitive thought is often seen to be a defect in the theory.

It's important to note, however, that the inadequacy of Act Consequentialism to accomodate the Kantian intuition is a different problem from the one that I mean to discuss. The Kantian intuition arises only in cases that feature repeatedly bad act types. Recall the second lesson we drew above regarding Snake Bite. The problem of mismatched verdicts arises in Snake Bite though neither of us performs a repeatedly bad act type. Thus, the mismatch problem has nothing to do with a failure of Act Consequentialism to capture the Kantian intuition.

It will perhaps be helpful to see an example of a Kagan-style case in which the mismatch problem does arise. Consider

> *Beans*: In a village, 100 people are about to eat lunch. Each has a bowl containing 100 beans. Suddenly, 100 hungry bandits swoop down on the village. Each bandit takes the contents of the bowl of one villager, eats it, and gallops off. Next week, the bandits plan to do it again, but one of their number is afflicted by doubts about whether it is right to steal from the poor. These doubts are set to rest by another of their number who proposes that each bandit, instead of eating the entire contents of the bowl of one villager, should take one bean from every villager's bowl. Since the loss of one bean cannot make a perceptible difference to any villager, no bandit will have harmed anyone. The bandits follow this plan, each taking a solitary bean from 100 bowls. The villagers are just as hungry as they were the previous week, but the bandits can all sleep well on their full stomachs, knowing that none of them has harmed anyone.[16]

The mismatch problem arises for Act Consequentialism+ in connection with what the bandits do on the second raid. The bandits together bring about a bad outcome, and they could have done something else with a much better outcome—had none of them participated in the second raid, the villagers would not have gone hungry. Thus, the group act [Bandit1 steals, Bandit2 steals, Bandit3 steals, . . . , Bandit100 steals] has an alternative with a better outcome, and so it's morally wrong according to Act Consequentialism+. On the other hand, one bean more or less is never enough to make a difference to well-being. No bandit could have made the situation better by not participating. Thus, no individual act—not Bandit1's theft, not Bandit2's theft, not Bandit3's theft, . . . , and not Bandit100's theft—has an alternative with a better outcome, and so no individual act is morally wrong according to Act Consequentialism+.

Beans is a fine example of the mismatch problem. But we should be careful when evaluating the solutions offered in connection with this case and any other Kagan-style case. Suppose some proposed resolution to the mismatch of verdicts in

[16] Peter Singer discusses this case in Singer (1998), and this case was first discussed by Jonathan Glover in Glover (1975).

Beans depends upon the fact that all the bandits perform an act of the same type—that Beans satisfies condition (1) of Kagan-style cases. Or suppose some resolution depends upon the fact that *stealing one bean from each villager* is a repeatedly bad act type for which individual tokenings don't make a difference—that Beans satisfies conditions (2) and (3) of Kagan-style cases. Under either supposition, the proposed resolution will not be generalizable to all instances of the mismatch problem. For as I have demonstrated, the three conditions of Kagan-style cases that we've been considering are neither necessary nor sufficient for the mismatch problem. They are merely distractions from the essential nature of the problem that I mean to discuss.

1.5 A Note about Prisoner's Dilemmas

I'm interested in the mismatch problem for Act Consequentialism+ as it arises in connection with collective harms. A collective harm, recall, is a situation in which a group could have done better though each individual could not have done any better. One natural thought is that collective harms arise in connection with prisoner's dilemmas. But, as I will demonstrate in this section, collective harms and prisoner's dilemmas come apart. Not every prisoner's dilemma gives rise to the mismatch problem for Act Consequentialism+.

Before proceeding, it will be helpful to consider an example of a prisoner's dilemma.

> *Two Hats*: You have a blue hat, and you'd prefer a red one. I have a red hat, and I'd prefer a blue one. Each of us strongly prefers two hats to any one. For this reason, neither of us would reciprocate were the other to give up their hat. Furthermore, each of us prefers either of the hats to no hat at all. Accordingly, you decide to keep your blue hat; had you given it to me, I would not have given you my red hat in exchange, and you would have ended up worse off. Similarly, I decide to keep my red hat; had I given it to you, you would not have given me your blue hat in exchange, and I would have ended up worse off. As a result, each of us gets stuck with an undesirable hat. Each of us values having a single, undesirable hat at +1. Each of us values having no hat at all at −1. Each

> of us values having two hats at +6. And each of us values having a single, preferred hat at +2.[17]

We may represent the situation in Two Hats by Table 1.3. The very top row of the table illustrates your possible acts. What you actually do is keep your hat (in bold), but you could have given it up. The very left column of the table illustrates my possible acts. What I actually do is keep my hat (in bold), but I could have given it up as well. Each combination of possible acts corresponds to an outcome, and the value for each of us under each outcome is displayed in the table. The worlds in which you and I give have a value of +2 for you and a value of +2 for me. The worlds in which you keep though I give have a value of +6 for you and a value of −1 for me. The worlds in which I keep though you give have a value of −1 for you and +6 for me. And the worlds in which you and I keep have a value of +1 for you and +1 for me. Each of these outcomes is accessible to the two of us together; we could have brought about any of them. We actually bring about the outcome in the bottom right box (in bold), but we could have brought about the outcomes in any of the other boxes had we acted differently.

	You give	**You keep**

I give	you(+2)
	me(+2)
		you(+6)
		me(−1)

I keep	you(−1)
	me (+6)
		you(+1)
		me(+1)

Table 1.3. The possible outcomes in Two Hats

Notice a few features of the situation in Two Hats. First, the case introduces a distinction between value *for* an agent and *overall* value. Roughly, the value of an outcome for an agent is a measure of how well that agent fares under that outcome. The outcome in the bottom left box is better for me than it is for you. The overall

[17] I adapt this example from Steven Kuhn's Stanford Encyclopedia of Philosophy article on Prisoner's Dilemmas, Kuhn (2017).

value of an outcome is a function of how everyone fares under that outcome. It is a more impartial measure of the value of an outcome than the value of that outcome for any particular agent. We may understand the calculation of the overall value of an outcome as follows: write down the value of the outcome for each of the agents involved in it, and then add up all of these values. The outcome in the top right box has the same overall value as the outcome in the bottom left box; each has an overall value of $6 - 1 = 5$.

Second, the structure of the case allows for dominance reasoning. For ease of presentation, we may focus on your decision situation, bearing in mind that my decision situation is the same in all relevant respects. You are better off keeping your hat no matter what I do. If you keep your hat, the worst that can happen is that I too keep my hat, which is much better than the worst that can happen if you give your hat to me—in which case you could end up with no hats at all. Keeping your hat also allows for the possibility that you achieve the best outcome for you in which you end up with two hats. In this way, what you actually do dominates giving your hat to me.

On the basis of these two features, I will assume that Two Hats is a paradigmatic instance of a prisoner's dilemma. There are many different but related problems that arise in connection with prisoner's dilemmas. One problem has to do with the self-interest theory of rationality:

> A common view is that the puzzle illustrates a conflict between individual and group rationality. A group whose members pursue rational self-interest may all end up worse off than a group whose members act contrary to rational self-interest.[18]

According to the *Self-Interest Theory of Rationality* (SIR), it is rational for an agent to perform some act just in case the agent has no alternative with an outcome that's better for her. In Two Hats, SIR says that each of us acts rationally. You keep your

[18] Kuhn (2017)

hat, which gets you +1. Had you pursued your alternative, you would ended up with no hat at all, -1. So you don't have an alternative with an outcome that's better for you. Similar remarks apply to my keeping my hat.

The problem arises once we see that it's natural to extend SIR to evaluate the actions taken by groups. Assume that it makes sense to talk about the welfare level of a group. As a simple way of doing this for the purposes of illustration, let's assume that the welfare level of a group of people is the sum of the welfare levels of the individuals in the group. Under this assumption, it makes sense to extend SIR to groups: according to *SIR+*, it is rational for an agent (whether an individual or a group) to perform some act just in case the agent has no alternative with an outcome that's better for the agent. In Two Hats, we together do have an alternative that's better for our group. In fact, we have three. Had we done [I give, you give], [you keep, I give] or [I give, you keep], the result would have been better for our group. Because we actually do [I keep, you keep], we actually end up with a welfare level of $1 + 1 = 2$, but we could have ended up with a welfare level of $2 + 2 = 4$ or $6 - 1 = 5$ had we acted differently. Thus, SIR+ delivers mismatched verdicts in Two Hats: the theory says that our group has acted irrationally though it says that each of us has acted rationally.

We are now in a position to see how prisoner's dilemmas and the problem I'm interested in come apart. Notice that, unlike SIR+, Act Consequentialism+ makes the normative evaluation of acts a function of overall value. Like most theories of right and wrong, Act Consequentialism+ purports to honor the idea that morality is impartial. Sometimes morality requires you to pursue an outcome that's not best for you. Under the Act Consequentialist+ picture, morality often demands self-sacrifice in the interest of maximizing overall value.

For this reason, Two Hats does not give rise to the mismatch problem for Act Consequentialism+. According to Act Consequentialism+, each of us acts morally impermissibly. To see this, we must consider the situation in Two Hats from the perspective of overall value. See Table 1.4. The top left box (representing the outcome in which each of us gets a desired hat) has an overall value of $2 + 2 = 4$, which is second best to the overall value of $6 - 1 = 5$ in the top right and bottom left boxes (representing the outcomes in which one of us gets two hats). The bottom right box (representing the outcome in which each of us keeps an undesired hat) has the worst overall value, $1 + 1 = 2$. What actually happens (in bold) is that I keep my hat and you keep your hat, putting us in the bottom right box. Accordingly, each of us has an alternative with a better outcome. Had you given your hat to me instead, I would have kept my hat, putting us in the bottom left box: the *best* outcome. Similarly, had I given my hat to you, you would have kept your hat, putting us in the top right box: the *best* outcome. Thus, each of us acts wrongly according to Act Consequentialism+. And this verdict lines up with the verdict at the group level. We together have three alternatives with a better outcome than the *worst* outcome that we actually bring about; so the group act is also wrong under Act Consequentialism+. No mismatch.

	You give	**You keep**
I give	*2nd best*	*best*
I keep	*best*	***worst***

Table 1.4. The possible outcomes in Two Hats in terms of overall value

The foregoing discussion demonstrates that not every prisoner's dilemma gives rise to the mismatch problem for Act Consequentialism+. On reflection, this shouldn't be surprising. Prisoner's dilemmas give rise to a mismatch problem for SIR+, but SIR+ and Act Consequentialism+ deliver different types of normative evaluations, and they do so under different conditions. Thus, it's important to be clear that I'm not interested in discussing problems for prisoner's dilemmas. Instead, I'm interested in discussing the problem of mismatched verdicts for Act Consequentialism+ in connection with collective harms.

1.6 Conclusion

In this chapter, I've explained the problem that I'm interested in discussing: the mismatch problem for Act Consequentialism+ in connection with collective harms. This problem arises under a natural set of assumptions about acts, alternatives, and group acts. It arises under a natural extension of Act Consequentialism that evaluates the moral status of group acts. Given these assumptions, Act Consequentialism+ delivers mismatched verdicts in Snake Bite, Two Shooters, and Beans. Related problems crop up in connection with the latter two cases. It will serve us well to keep these related problems in mind. Many philosophers seem to be discussing the problem of mismatched verdicts, but they do so in connection with cases like Two Shooters or Beans. Accordingly, it's important to be on alert; it's important to ask whether a proposed solution to the mismatch problem applies to all instances of the problem, or whether it applies merely to those cases that share the same additional, inessential features.

1.7 Summary of Chapters

The rest of the dissertation proceeds as follows:

In Chapter 2, I focus on the mismatch problem for consequentialism as it arises in connection with anthropogenic climate change. In certain cases, it seems as though someone has misbehaved, but if no individual makes a difference, then it's not clear how to assign wrongdoing. Several philosophers have suggested an intuitive solution. The idea is to assign disutility to a group act and then distribute a share of this disutility to each of the participants. I explain why the idea about participation-adjusted utility hasn't been developed in a satisfactory way. I identify several difficulties for the approach, and I explain why I don't think the approach can be made to work in the climate change case.

In Chapter 3, I consider the popular idea that an appeal to expected utility will solve the mismatch problem. I explain why appealing to expected utility cannot work. And in Chapter 4, I present Shelly Kagan's particular version of the expected utility approach. I explain why Kagan's version cannot succeed.

In Chapter 5, I reject another popular attempt to resolve the mismatch problem. Several philosophers have assumed that an essential feature of all mismatch problem cases is the failure of the individuals to cooperate. These philosophers then go on to suggest that uncooperativeness constitutes a moral failing under a suitably extended version of consequentialism. I explain why this strategy does not work as a general solution to the problem; I present a version of the problem case that does not involve uncooperative individuals.

In Chapter 6, I offer a partial solution to the mismatch problem. A particularly nasty version of the mismatch problem arises in cases where a mismatch appears to persist on every assumption about how many individuals will participate in the morally impermissible group act. This happens when the disvalue of the bad outcome accumulates, without a sharp boundary between any neighboring outcomes. Perhaps the most famous illustration of accumulation is Derek Parfit's case of the Harmless Torturers. I identify a distinguishing axiological feature of the Harmless Torturers: unlike certain standard versions of the mismatch problem, the case involves indeterminate value comparisons. On this basis, I show how Act Consequentialism can be modified in order to solve the problem. I conclude with some ideas about how the solution can be applied to some versions of the anthropogenic climate change case.

One final note: I originally prepared the following chapters as stand-alone papers. For this reason, there's redundancy in a few places, particularly in the introductory parts of the chapters in which I describe the mismatch problem. In some chapters, I've attempted to remove discussions that are repeated elsewhere. In other places, I’ve added footnotes describing when sections may be skipped provided that readers have familiarity with previous chapters.

CHAPTER 2

CLIMATE CHANGE, CONSEQUENTIALISM, AND DIFFICULTIES FOR THE PARTICIPATION-ADJUSTED UTILITY SOLUTION

2.1 Introducing the Climate Change Case

There are apparently many ways in which an individual American may contribute to the rising concentration of atmospheric CO_2: driving a gas-guzzling SUV, flying somewhere for vacation, clear-cutting a wooded lot, purchasing a steak. Through the performances of acts like these, each of millions of Americans participates in a collective activity that's causing dangerous climate change. Erratic and extreme weather events occur more frequently, rising sea levels wash out developed coastlines, and heat waves disrupt crop growth. People overall suffer more than they otherwise would under a stable and cooler climate system.[1]

In some cases, an individual's act apparently makes a significant contribution to the problem. If a popular celebrity goes on joyrides in her flashy sports car, others may be influenced by her example. The celebrity's contribution may have a big impact. But in other cases, it seems that an individual's act of contribution makes no difference. The emissions produced by the daily commute of an ordinary American, for example, are seemingly negligible. Climatic changes apparently will not occur from the little bit of CO_2 that comes out of a single vehicle's tailpipe, and no one will be influenced to commute more by your regular commute. In extreme cases, it might

[1] For a comprehensive source of information on the effects of climate change, see Stocker et al. (2013).

be that *no* individual's act makes a difference. No matter how any given individual would act—whether, for example, he or she would bike to work instead of driving—not enough others would act differently; the bad outcomes of climate change would still be just as bad. In such cases, each participant seems to have a plausible excuse: "the result would have been the same even if I had made no contribution."

There's a familiar problem for consequentialism that arises in connection with the extreme cases. It seems as though someone, or perhaps *something*, has misbehaved. But Act Utilitarianism is notorious for being unable to capture the intuition when focused solely on an individualistic level of moral evaluation. We assume that each individual act maximizes utility: the consequence of a given act of contribution is no worse than the consequence of not contributing. So if a consequentialist remains fixated on an individualistic level of evaluation, it is unclear how to accommodate the intuition that a wrong has been committed.[2]

We may distinguish between two broad approaches consequentialists have taken in order to accommodate the intuition of wrongdoing in connection with anthropogenic climate change and related cases. Under the first approach, the consequentialist widens the scope of moral evaluation to include the behavior of groups. Under this 'group-based' approach, the consequentialist attempts to capture the intuition about misbehavior from within the traditional Act Utilitarian moral framework. The *group act* doesn't maximize utility; the consequence of what the group does is worse than many other consequences accessible to the group. Thus, it is possible to identify at least one morally wrong act: a group act. The second approach is to remain fixated on individualistic morality—to leave group acts out of it. Typically, those who pursue the

[2] See, for example, Sinnott-Armstrong (2005) and Christian Barry (2015). See also Lawford-Smith (2016).

‘individualistic’ approach abandon the Act Utilitarian moral framework for something else.[3]

Can the group-based approach be successfully developed in connection with the case of anthropogenic climate change? In what follows, I elucidate and evaluate a familiar version of the group-based approach. It involves assigning disutility to the group act performed by the American emitters and then distributing a share of this disutility to each of the participants. I show that this idea hasn’t been developed in a satisfactory way in connection with the case of anthropogenic climate change. I explain why I can’t see how this could be done.

2.2 Adjusting Utility for Participation

Perhaps the biggest obstacle for group-based approaches is the mismatch problem. While it is an entirely natural thought that the group of American emitters has acted wrongly, the success of an appeal to group wrongdoing requires that we have some way of distributing the group’s guilt to the individual members of the group. Otherwise we will end up with a moral framework that paradoxically condemns what the group does without being able to say anything against what the individual members of the group have done. Since the group has an alternative with a greater utility, the group has acted wrongly under Act Utilitarianism. But since no individual contributor has an alternative with a greater utility, each individual member of the group has acted permissibly under Act Utilitarianism. The theory delivers mismatched verdicts; it condemns the behavior at the group level, but this condemnation fails to trickle down to the behavior of any of the individuals involved. So if we are to pursue the group-based approach to consequentialism in connection with the case of anthropogenic

[3] Some pursue a hybrid version of consequentialism. Dale Jamieson is one such philosopher; see Jamieson (2007). Others embrace virtue ethics or Kantianism as the best way to capture the intuition of wrongdoing. Such philosophers include Casey Rentmeester and Marion Hourdequin; see Rentmeester (2010) and Hourdequin (2010).

climate change, we need a slightly modified version of Act Utilitarianism: one that is able to accommodate the intuition that there is wrongdoing but that will not lead to mismatched verdicts.

Some philosophers have suggested, in various ways, that this may be accomplished by appeal to a natural thought: individuals who participate in a suboptimal group act ought to get assigned some discredit for participating.[4] Many philosophers find it hard to accept that a harmful group act could be composed of harmless individual acts.[5] Peter Singer, for example, has remarked that it is absurd to deny that we are each responsible for a share of the harms we collectively cause. He has claimed that "[a]n act may contribute to a result without being either a necessary or sufficient condition of it, and if it does contribute, the act-utilitarian should take this contribution into account."[6]

For an illustration of how this idea has been presented in connection with anthropogenic climate change, consider John Nolt's calculation of the climate-related harms attributable to an average American.[7] Through its collective emissions of greenhouse gases, a large group of Americans has brought about serious injury to millions of people. This group—call it the 'American Emitter collective' (AE)—could have done much better had the group pursued more sustainable, less carbon-intensive ways of living. And yet we are assuming that in the extreme cases no individual member of AE makes a difference by his or her personal emissions. The climate-related injuries

[4] See Lyons (1965), Singer (1972), Glover (1975), Goldman (1999), and Nolt (2011) for suggestions of this sort. See also Horwich (1974), Regan (1980), Parfit (1984), and Kernohan (2000) for presentations and criticisms against the idea. In a later footnote, I explain why my criticism of the idea is importantly different from the criticisms mounted by this latter group of philosophers.

[5] See, for example, Roberts (2011). Relatedly, see Zimmerman (1992) and Kagan (2011) for a reluctance to accept that a morally wrong group act may be composed entirely of morally permissible individual acts.

[6] The quotation is from (Singer, 1972, 103). See also Singer (1998).

[7] See Nolt (2011) and Nolt (2013). For other climate ethicists who seem to adopt a similar procedure, see Broome (2012) and Hiller (2011).

are not brought about by any individual emitter, but by the group. Nolt nonetheless pursues a calculation that attributes harm to the emissions of an individual American. He estimates the total greenhouse gas emissions associated with AE's group act, and he estimates the total harms resulting from these total emissions. He also estimates an average individual American's lifetime emissions. Based on the proportion of individual lifetime emissions to the emissions of the group, he concludes that an average American's lifetime of emissions does harm equivalent to the serious suffering and/or deaths of two future people. In this way, Nolt assigns to the acts of individual members of AE a share of the total disutility of what AE does. Apparently, this assignment is based on the fact that each individual American participates in the group act. The assignment of extra disutility carries normative implications. Presumably, it makes it so that each individual American acts wrongly if he or she doesn't reduce personal emissions.[8]

It is important to list the steps that would be involved in a complete development of the idea. First, a group act with significant disutility is identified: this is AE's group act. Next, each individual member of AE is identified as a participant in the group act, and so each individual act gets assigned a share of the disutility associated with the group act. Assuming that each individual could have acted differently in a way that wouldn't have counted as participation, we hold off from assigning any extra disutility

[8] In connection with related cases, other philosophers have pursued a solution of the same sort. Consider Jonathan Glover's case of the bean-stealing bandits in Glover (1975). A group of 100 bandits raids a village, each stealing an imperceptible amount of beans. Though no individual act makes a difference, the group act makes the villagers hungry. Glover suggests that each bandit's theft should inherit some of the total disutility associated with the group act performed by the bandit collective. The harm each individual does "should be assessed as a fraction of a discriminable unit, rather than as zero... [and] in cases where harm is a matter of degree, sub-threshold actions are wrong to the extent that they cause harm, and where a hundred acts like mine are necessary to cause a detectable difference I have caused 1/100 of that detectable harm." (Glover, 1975, 174) Each bandit participates in the raid. Accordingly, each individual bandit's act of theft gets assigned an additional amount of disutility: 1/100 of the disutility associated with the group act. So each bandit acts wrongly because each has an alternative which does not receive the adjustment for participation. (The case is also discussed in Jackson (1987) and Singer (1998). Derek Parfit discussed a similar case involving one thousand torturers in Parfit (1984); see Kagan (2011) as well.)

to the individuals' non-participating alternatives. Act Utilitarianism is modified: according to *Participation-Adjusted Act Utilitarianism* (PAAU), an act is morally permissible just in case there's no alternative with a greater participation-adjusted utility. Each individual American has an alternative with a greater participation-adjusted utility. So each acts wrongly. In this way, a slight modification to Act Utilitarianism allows the group-based approach to evade the mismatch problem.

Note the data that must be accommodated if PAAU is to be successfully developed: in the case of anthropogenic climate change, AE is a group; the group performs a group act; each individual act counts as participation in it; and each individual has an alternative under which he or she wouldn't have participated. In the following sections, I proceed to show that there are no satisfactory accounts of groups and of participation that will serve to substantiate this data.[9] In section 2.3, I present several accounts of what it is for a group to be capable of acting. I raise problems for each account. In section 2.4, I pursue a different way of thinking of group acts that will allow us to move forward with the discussion: I introduce a distinction between cohesive group acts—acts that are performed by the group *acting as a group*—and non-cohesive group acts. I raise problems for two plausible accounts of cohesive group acts. In section 2.5, I propose a causal account of participation under which each individual contributor to anthropogenic climate change participates in the group act by causing some change to an underlying dimension. I present several difficulties for the account. There is no suggestion in the literature, and no suggestion I can think of, that will make it work right. I conclude that either consequentialists should not

[9] In a previous footnote, I alluded to several philosophers who present and then reject PAAU. These philosophers include Paul Horwich, Donald Regan, Derek Parfit, and Brian Kernohan. They write in such a way as to suggest that they assume that PAAU can be formulated adequately; they assume that the data can be accommodated in related cases. They then show that the theory would deliver counterintuitive verdicts in connection with certain other cases. It seems to me that the advocate of PAAU would be able to respond to these objections by insisting that the oppponent has not formulated PAAU properly. But it is my aim in this paper to identify the difficulties that arise in trying to formulate PAAU in the first place.

pursue the group-based approach, or they should not attempt to resolve the mismatch problem by appealing to participation-adjusted utility.

2.3 Searching for an Adequate Account of Group Acts

To accommodate the data, the advocate of PAAU needs an account of groups that establishes that AE is a group capable of performing group acts. There are two different approaches one might take. The first, *conservative approach* proceeds by comparing AE with certain paradigmatic act-performing groups. The goal is to identify some general feature of these groups that makes them capable of performing group acts and such that AE has this feature. The second, *liberal approach* proceeds by simply taking any set of individuals to make up an act-performing group. Each approach encounters difficulties.

2.3.1 The Conservative Approach

Start by reflecting on paradigmatic act-performing groups: corporations, armies, senate committees, sports teams. We are looking for some general feature of these groups that makes them capable of performing group acts; perhaps AE shares such a feature with the paradigmatic act-performing groups.

There's a certain way of thinking about the paradigmatic act-performing groups that has intuitive appeal. The thought goes like this: only *agents* perform acts; so if a group is to be capable of performing acts, there must be some feature of the group that makes it appropriate to think of the group as an agent. According to Stephanie Collins, the characteristic feature of the paradigmatic act-performing groups is the possession of a decision-making procedure.[10] The procedure takes in and processes moral reasons. The result of the group's following the procedure is the issuing of instructions to the individual members of the group. Suppose some large corporation

[10] See Collins (2013).

is in engaged in morally problematic environmental practices; suppose the production of a certain product is causing unnecessary environmental harm. The board members of the corporation may get together according to customary procedures in order to figure out how to address the problem. The board may decide to change a certain aspect of production, and it may then issue instructions to individuals within the corporation regarding how to implement the changes.

But under the Collins-inspired account, we discover that AE is not capable of performing group acts. Unlike corporations and the other paradigmatic act-performing groups, AE lacks a decision-making procedure. Environmentalists lament the fact that AE is not able to process moral reasons—if only there were some way to command the group to emit less! Unfortunately, there is no governing body to receive and process an appeal by environmentalists to reduce emissions. Furthermore, no individual American would be issued instructions to behave in any particular way by AE. So the advocate of PAAU needs a different way of accounting for groups in order to accommodate the data in the climate change case.

Perhaps a different feature of the paradigmatic act-performing groups will do. Notice that members of the paradigmatic act-performing groups listed above belong to their respective groups in virtue of some formal membership condition. The employees of a corporation have their names appear on a payroll; the members of an army have to be accepted into its ranks. In order to serve on a senate committee, one must be officially appointed; in order to be a member of a sports team, one must make it in the tryouts. In general, each of the paradigmatic act-performing groups apparently has some formal condition of admission. We may assume that it is in virtue of this feature that we may see the group as capable of acting.

But AE obviously lacks formal admission criteria. There is no official procedure for inducting individual Americans into AE. There's no list associated specifically with the group.

Perhaps *formal* admission conditions are not necessary for a group to be capable of performing group acts. Suppose several individuals spontaneously gather in a park and toss a ball around. Two teams emerge; suddenly the play becomes competitive. If one of the two teams wins the game that has developed, we will want to ascribe the victory to a group. But no formal membership condition exists among the members of the winning team. There is no roster; there were no tryouts. Apparently, the only membership condition in this case is that each self-identifies with the team. So it might be suggested that in certain circumstances it is sufficient for being a group capable of performing group acts that each member merely identifies himself or herself with the group.

Of course, AE fails to possess even this feature. Suppose Danny the Driver drives his gas-guzzling SUV excessively; he often takes the long way home from work just because it gives him an excuse to spend some leisure time behind the wheel. But suppose that Danny denies that he is contributing to climate change. He doesn't even recognize that an increase in atmospheric CO_2 is a byproduct of burning gasoline. Suppose that even if Danny accepted that a group activity is causing climate change, Danny would not self-identify as a member of the group. Intuitively, Danny is a member of AE, as are hundreds of thousands of other Americans who also deny their roles in causing climate change. So the idea that a group is capable of performing group acts if each individual identifies as a member is not able to accommodate the data in the climate change case. The advocate of PAAU will have to look elsewhere.

There is apparently no feature of the required sort that AE shares with the paradigmatic group acts. AE lacks sufficient structure, and its members fail to meet a recognizable formal or informal membership condition. Some advocates of the conservative approach will conclude that AE is not a group of the appropriate sort—that it is not capable of performing group acts. They will conclude that the idea about participation-adjusted utility cannot be adequately developed.

2.3.2 The Liberal Approach

But other philosophers have taken a different, liberal approach toward thinking about whether a group is capable of performing group acts. In his discussion of the mismatch problem for Act Utilitarianism, Frank Jackson is "inclined to count any old mereological sum of individual actions (or of group actions) as a group action (or as a *further* group action). 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."[11] Consider any random collection of individuals. According to the Jackson-inspired account, the collection is a group, and the group is capable of performing group acts. Jackson understands the group act as a 'mereological sum' of the individual actions. For our purposes here, we may understand a group act as simply any set of individual acts. Accordingly, any set of people is capable of performing a group act.

While this liberal approach identifies AE as an act-performing group, it generates several unacceptable normative implications when trying to develop PAAU. Toward explicating these unacceptable consequences, consider the enormous number of group acts that are performed in the climate change case under the liberal approach. First, there's the group act performed by AE. This is a very large set of individual acts. Suppose we remove one individual act from the set. Another, distinct set of individual acts results. Under the liberal approach, this is also a group act performed by a different group: AE-i_1. And if we put the individual act back and take away a different individual act, we get a third set of individual acts corresponding to a third group act performed by a third group: AE-i_2. There are millions of such groups that may be generated in this way.

[11] (Jackson, 1987, 93). See also Killoren and Williams (2013).

Furthermore, instead of removing acts from the group act performed by AE, we can also add individual acts at random. Consider some innocent bystander, b. Suppose he acts in a way that doesn't have anything to do with what each of the members of AE does. Nonetheless, there is a set of individual acts corresponding to AE's group act plus the bystander's individual act. Under the liberal approach, this is a group act performed by a different group: AE+b. By substituting in different bystanders, we apparently generate millions—even *billions*—of related groups.

Of course, the sheer profusion of groups does not by itself pose a problem for developing PAAU. The problem arises when we attempt to calculate the participation-adjusted utility of an individual act. The existence of so many groups results in a vicious sort of double-counting. Recall that the utility of an individual act receives an adjustment if it counts as participation in a group act. Consider Danny's taking the long way home from work in his gas-guzzling SUV. This act counts as participation in the group act performed by AE. But without some principled reason for thinking otherwise, it also counts as participation in the group acts performed by AE-i_1, AE-i_2, AE-i_3, and so on. Each of these group acts has a significant disutility. So the utility of Danny's commute is adjusted millions of times; he participates in group acts performed by millions of groups. Thus, the participation-adjusted utility of Danny's commute is astronomically and unacceptably negative. It is perhaps just as negative as the disutility associated with AE's group act.

The profusion of groups also reveals the importance of determining how the disutility of a group act should be divvied up among the group's members. Nolt calculates the harm done by an average American's lifetime of emissions and attributes it to members of the group: each member of the relevant group gets assigned an averaged-sized share of the disutility of the group act. But this focus on an average-sized assignment to the individual members of AE+b will result in an unjustifiable adjustment to the utilities of the acts of innocent bystanders. Recall that the group AE+b

contains all the members of AE plus some bystander member. The bystander acts in a way that doesn't have anything to do with what each of the members of AE does. Nonetheless, the resulting group act—the act performed by AE+b—has significant disutility. On the view that says that each member of the group gets assigned an average-sized share, the utility of the bystander's act gets adjusted down. He's unacceptably penalized for something he didn't take part in. This reveals that we cannot let participation-adjusted utility be an average of the disutility of a group act under the liberal approach. So if the liberal approach is to be developed, we need some principled way of assigning a zero-sized share to innocent bystanders. We may say that b gets a zero-sized share because he doesn't participate. But then we need a separate account of participation, which is not included in the liberal approach to group acts.

Thus, the liberal approach on its own fails to accommodate the data in the climate change case. While the liberal approach succeeds in identifying AE as an act-performing group, we cannot identify the single unique group act in which each individual member of AE participates. Thus, we have no way of identifying AE as the group of interest—the group that performs the group act with significant disutility that must be divided up. Accordingly, I cannot see how PAAU could be developed under the liberal approach without generating unacceptable normative implications.

2.4 Cohesive and Non-Cohesive Group Acts

If PAAU is to be adequately developed, it must be on some middle ground between the conservative and liberal approaches. We may accept that any collection of individual acts makes up a group act. But we cannot allow it to be the case that every group act corresponds to a group of the appropriate sort. Some group acts are *cohesive*: there is justification for thinking that the individuals have acted together, in concert, *as a group.* Other group acts are non-cohesive: the individual acts are totally unrelated to each other, like my last eye-blink and Nero's burning of Rome.

Assume that only cohesive group acts are such that their disutility must be divided up; it's only cohesive group acts in which the individual acts count as participation. To accommodate the data, we would need to establish that AE is capable of performing a cohesive group act—that AE can act in such a way that there is some feature of the individual acts that ties them all together. Intuitively, one might say that the individual members of AE have acted together in virtue of the fact that each participates. But we need an account of cohesive group acts that bears out this intuition. To this end, we may reflect on two intuitively plausible approaches to cohesive group acts.

2.4.1 Intention-based Account

According to a particularly popular approach, a group act is cohesive just in case all the individuals who perform acts contained in the group act share a particular sort of intention. There are different ways of thinking about the required intention.[12] Since many of the individual members of AE have minimal to no interactions with each other, I follow Christopher Kutz's account, which is apparently the least demanding of the intention-based accounts.[13] Say that a *shared goal* is some particular outcome that two or more individuals would like to see obtain. Say that each of these individuals acts with a *participatory intention* if he or she has an intention to do his or her part in a group act the performance of which is sufficient for producing the shared goal.

[12] Compare Michael Bratman's and J. David Velleman's views in Bratman (1993) and Velleman (1997).

[13] Kutz (2000)

We may see a group act as cohesive just in case all the component individual acts are accompanied by participatory intentions.[14]

It is fairly straightforward to see that the participatory intentions account cannot accommodate the data in the climate change case. The individuals who contribute to climate change do not all share a goal; there's no particular outcome that all the emitters would like to see obtain. AE is made up of individuals who go about their days in almost total independence from the others. If we could somehow gather the members together and ask 'What does each of you aim to accomplish by increasing the atmospheric concentration of CO_2?', it's not clear what sort of answer we'd get. Someone might suggest that he wants to bring about climate change. But the production of dangerous climate change is almost universally *unwanted*. Someone else might suggest that she wants to enjoy a certain level of convenience. But this wouldn't establish that the emitters have a shared goal. Suppose you and I both drive to work instead of biking. It's not as if you want to see that *I* secure a bit of extra convenience, and I don't want to see that *you* secure a bit of extra convenience.

But even if there were some shared goal in the anthropogenic climate change case, it would be implausible to think that each individual intends to do his or her part in a group act that would be sufficient for producing the shared goal. Danny doesn't intend to do his part in anything; he simply intends to take pleasure in a

[14] To see how this account of cohesive group acts works, it may be helpful to reflect on one of Kutz's examples: "Suppose that while we are having a picnic, it begins to rain. I jump up, grab the sandwiches and head for the car. I intend to do my part of our saving the picnic, hoping you will simultaneously grab the drinks and the blanket. If you do, then it is reasonable to say we will have jointly saved the picnic. We might not have acted jointly, if, say, you had been dozing when the rain hit. But if we do both act with participatory intentions, then we will have jointly intentionally saved the picnic..."(Kutz, 2000, 18)

Under Kutz's account, there is justification for thinking that we have acted together, despite the fact that neither of us communicates with the other. It is simply that you and I have a shared goal that the picnic is saved, I intend to do my part in a group act that's sufficient for bringing about this state of affairs, and you also intend to do your part in a group act that's sufficient for bringing it about that the picnic is saved. I grab the sandwiches, you grab the drinks and the blanket, and our group act is cohesive.

longer commute in his SUV. Meanwhile, suppose that Franny the Flier routinely takes plane flights from San Francisco to Maui on vacation. Whether Danny wants to see the same outcome obtain that Franny does, he doesn't intend to do anything with Franny in order to produce it. It is implausible to characterize the individuals as acting with participatory intentions. So the intention-based account of cohesive group acts implies that the group act performed by AE is not cohesive. Accordingly, the account cannot accommodate the data about participation.[15]

2.4.2 Act-type Account of Participation

Consider a different approach to cohesive group acts. Instead of identifying a common type of intention, we may instead identify a common type of act. We might say that a group act is cohesive just in case all the component individual acts are of the same type.

One problem with this account is that it incorrectly characterizes situations in which individual members of the same group are trying to compete with the others. Consider a tiny group composed of two 'interrupters'. Each performs an act of the same type: trying to frustrate the other from performing his act. Chaos ensues; the individuals jolt each other around. Intuitively, the group act is not cohesive—in this case, there's justification for thinking that the interrupters are *not* acting together. But the interrupters perform acts of the same type.

[15] Some might see the explicit focus on intentions as misguided. According to Margaret Gilbert, a collection of individual acts is cohesive whenever the individuals have mutual obligations to each other around some shared goal. See Gilbert (1990). These mutual obligations need not require intentions (though often they do). But even if AE does have a shared goal, the individual members apparently do not satisfy the condition of mutual obligations to one another. Suppose that Danny and Franny live on opposite coasts, living totally independent lives. Intuitively, there's nothing about what either does that generates any sort of mutual obligation to the other. Danny doesn't owe Franny anything. Were Danny to stop driving, Franny would have no grounds for complaint. And were Franny to stop flying, Danny wouldn't feel slighted in the least. So the Gilbert-inspired account also fails to classify AE's group act as cohesive.

Another problem with this account is that it fails to deliver the correct verdict in connection with certain paradigmatic act-performing groups. Consider the German football team that won the 2014 World Cup. During the last seven minutes of the championship game, the players acted together—the group act was cohesive—but not every player acted in the same way. Some players sat on the bench. Others ran around on the field. Some defended; others were on the offensive. One of the players, Gotze, scored. At any moment up until the game-winning goal, each player performed a different type of act.

For a similar reason, the account cannot accommodate the data in the climate change case. There are many different ways to contribute to the atmospheric concentration of CO_2. Danny travels by road. Franny travels by air. These are very different types of act. Someone might say that each of Danny and Franny performs an act of the type 'traveling', but it is possible to travel without emitting CO_2—sailing, for example. Someone else might suggest that each of Danny and Franny performs an act of the type 'emitting CO_2', but in the case of Franny, it's complicated. When Franny rides aboard a plane, she's not producing CO_2; the plane is. A given plane seats hundreds of passengers. We may imagine that Franny has never caused a plane to fly: she has never purchased a deciding ticket—a ticket such that the plane wouldn't have flown had she not purchased it. Accordingly, it's not clear that we should describe her individual acts of flying as emitting CO_2. Her individual acts do not strictly speaking put any CO_2 into the atmosphere. When we speak about Franny as an 'emitter', it's in an extended sense. And yet, intuitively, Franny contributes to climate change; her act shares some feature with Danny's act that ties them together.

Of course, Danny and Franny both perform acts of the type 'contributing to climate change'. When Franny purchases a ticket and rides on planes, she increases the demand for air travel. An increase in flights causes an increase in atmospheric CO_2. It is for a reason of this sort that Franny is thought to be a contributor.

But then we may as well say that AE performs a cohesive group act because each individual act counts as contribution—and what is it to contribute to climate change other than to participate in AE's group act?

2.5 Problems for the Causal Account of Participation

We need an account of cohesive group acts that establishes that the group act performed by AE is such that each individual act token contained in it counts as participation. Otherwise, there is no basis for making an adjustment to the utilities of the individual acts. Furthermore, it must be established that each individual who performs an act token contained in AE's group act has some alternative under which he or she wouldn't have participated—or, as I will suggest in this section, an alternative under which he or she wouldn't have participated *to the same extent*. Otherwise, all of every individual's alternatives will receive the same utility adjustments, and there will be no basis for concluding that any act token is morally wrong under PAAU.

In an attempt to meet these requirements, start by noticing that each individual act of contribution to anthropogenic climate change alters the climate system in some way: each individual American emitter causes a change in the atmospheric concentration of CO_2. The change may be caused directly, as in the case of Danny's drives, or it may be caused more indirectly, as in the case of Franny's flights. Either way, each individual act manipulates the same environmental variable as is manipulated by AE's group act: the group act causes a large change to the atmospheric concentration of CO_2; each individual act causes a tiny change.

Thus, we may say that each individual American emitter participates in the group act performed by AE in virtue of causing some change to a common environmental variable—which we may call the *underlying dimension*. AE has alternatives under which certain value states differ, though no individual member of the group has an alternative under which these value states differ. But each individual has alternatives under which the underlying dimension differs. Enough changes to the underlying dimension cause changes to the value states. We may say that AE's group act is cohesive—and that each individual act counts as participation in the group act—in virtue of each of the individual act's causing a change to the underlying dimension.[16]

In order to evaluate whether this account of participation can accommodate the data in the case of anthropogenic climate change, we must identify an underlying dimension in the case. An intuitive starting point is atmospheric concentration of CO_2. But the causal features of the anthropogenic climate change case are quite complex: it's not exclusively atmospheric CO_2 that causes changes to the climate system. When it comes to global warming, other so-called 'greenhouse gases' have an effect. These compounds include methane, ozone, water vapor, and CFCs. The problem is that certain greenhouses gases interact with others. CFCs, for example, destroy ozone. Suppose someone uses up some old cans of hairspray, sending CFCs up into the atmosphere. Has this person caused a change to the atmospheric concentration of greenhouse gases? The answer is not straightforward. Causal interactions *within* the underlying dimension increase the complexity of the solution. Perhaps we should instead take the underlying dimension to be something like the 'heat-trapping potential' of the atmosphere. It not clear how this environmental variable would be specified.

Furthermore, it's important to notice that the unifying feature of individual acts in the case of anthropogenic climate change is not simply that each causes a change to the underlying dimension. Certain acts of mitigation cause changes in the atmospheric concentration of CO_2. Some entities—so-called 'carbon sinks'—operate in such a way as to remove CO_2 from the atmosphere. Trees are carbon sinks because they absorb CO_2 as they grow. Suppose you plant hundreds of trees. This is a way for you to offset

[16] I borrow the concept of an underlying dimension from Shelly Kagan and Julia Nefsky; see Kagan (2011) and Nefsky (2012a).

your emissions precisely because it causes a change in the atmospheric concentration of CO_2. But your act of mitigation is not to count as participation in the group act performed by AE.[17]

Another issue is that the members of AE cause changes to the underlying dimension in myriad ways. Danny drives an SUV. The drive causes some CO_2 to go into the atmosphere. Franny purchases an airline ticket. The ticket purchase causes Delta Airlines to introduce a new flight. The new flight causes some CO_2 to go into the atmosphere. Intuitively, Franny's purchasing a ticket is at a greater causal distance from the underlying dimension than Danny's taking a drive. We may wonder whether both Danny's drive and Franny's ticket purchase are to count equally as acts of participation.

More generally, suppose events of type E cause events of type E'. Suppose events of type E' cause events of type U. Suppose events of type U cause changes to value states V. Suppose we identify U as the underlying dimension. Then E is at a *greater causal distance* from the underlying dimension than E'. A sufficiently general account of participation will need to indicate whether causal distance matters in qualifying an individual act as participation in a group act.

Perhaps a deeper problem with the causal approach is that members of AE have no alternatives under which they don't cause a change to the underlying dimension. You expel CO_2 when you breathe, and—excepting suicide and holding your breath—all of your alternatives are accompanied by breathing. So under the causal account of participation, whatever you would do, you would participate in the group act.

[17] Perhaps we could enhance the account of participation under the causal account in order to address this concern. We could say that an individual act counts as participation in a group act in some situation just in case the individual act causes a change to the underlying dimension *in the same direction* as the change to the underlying dimension caused by the group act. In connection with the anthropogenic climate change case, each of the individual American emitters acts in such a way as to cause an *increase* in the atmospheric concentration of CO_2, as does the group. More would need to be said about the notion of directionality in order to evaluate this suggestion.

That's apparently a failure to accommodate the data in the climate change case. We apparently need an account of participation under which each member of AE has an alternative under which he or she wouldn't have participated. Otherwise all of the individual's alternatives will receive a utility adjustment, and there will be no basis for concluding that any individual act token is morally wrong.

It seems to me that the advocate of PAAU would need to embrace some idea about levels of participation. This would have to be explained by appeal to the extent to which an individual act affects the underlying dimension. The greater your contribution of CO_2, the higher your level of participation; the smaller your contribution of CO_2, the lower your level of participation. And perhaps the greater the causal distance from the contribution of CO_2, the lower your level of participation as well. For many contributors, it cannot be established that each has an alternative under which he or she wouldn't have caused a change to the underlying dimension. But perhaps each has an alternative under which he or she would have either (i) caused a smaller change to the atmospheric concentration of CO_2, or (ii) retreated to an act at a greater causal distance from the contribution of CO_2.[18]

Even if all the preceding concerns can be addressed, a 'big picture' worry remains for the causal account. Any advocate of PAAU who is inclined to attach special significance to the disutility of AE's group act will be unable to do so under the causal

[18] A simple illustration may serve to demonstrate how an appeal to levels of participation would work. Suppose that Danny could either drive his gas-guzzling SUV the long way home from work or take a more direct route home. Each of Danny's alternatives—the long drive and the direct drive—would cause an increase in the atmospheric concentration of CO_2. Thus, whichever he would perform, it would count as participation in a group act performed by AE. And whichever of his alternatives he would perform, the group act would have the same disutility; after all, Danny doesn't make a difference. But were Danny to take the long drive, he would cause a greater increase in the atmospheric concentration of CO_2 than were he to take the direct drive. Under the idea about levels of participation, he would participate more by taking the long drive than by taking the direct drive. Accordingly, the utility of the long drive must get adjusted by a larger fraction of the disutility of the group act than the utility of the direct drive. Supposing that Danny takes the long way home from work, he will have an alternative with a greater participation-adjusted utility. There will be some basis for concluding that his individual act token is morally wrong according to PAAU.

account of participation. The cohesive group act of concern is much larger: anyone who has ever existed has caused a change in the atmospheric concentration of CO_2. This enormous group—which we may call the 'Global Historical Emitters Collective' (GHEC)—has performed a colossal group act spanning thousands of years and taking place within all geopolitical regions across the entire planet. And so it is GHEC's group act in which each American's individual act token will count as participation under the causal approach. An advocate of PAAU would make a mistake in identifying AE's group act as the group act with significant disutility that must be divided up. Instead, the participation-adjusted utility of any individual act of emission will be a share of the disutility of GHEC's group act. And any individual act that has caused a change in the atmospheric concentration of CO_2 counts as participation—even those acts of emission performed in countries that have done very little to impact climate change. So the causal account of participation identifies a group act that's much larger than our intuitive target, and it distributes wrongdoing more widely than is intended.

2.6 Conclusion

The proposal involving participation-adjusted utility faces some serious difficulties in connection with the case of anthropogenic climate change. There are assumptions about groups and participation that the proposal must be able to accommodate if PAAU is to resolve the problem for consequentialism: it must be established that the individual contributors to anthropogenic climate change compose a group that is capable of performing group acts, and it must be established that each individual contributor participates in the group act performed by AE. I have highlighted challenges associated with accommodating this data. While something like the causal account of participation appears to be the best way forward, it would take some serious work to develop the account in light of the many problems I have raised. Unless these difficulties can be adequately addressed, I conclude that the solution embodied by PAAU cannot be made to work. Consequentialists concerned about the problem posed by anthropogenic climate change either should not pursue the group-based approach, or they should not attempt to resolve the mismatch problem by appealing to participation-adjusted utility.

CHAPTER 3

EVALUATING THE MOVE TO EXPECTED UTILITY

3.1 Introduction

The mismatch problem for Act Consequentialism arises in connection with several different cases. Some philosophers have been attracted to the idea that an appeal to expected utility might offer a solution in some of these cases. In this chapter, I explain why appealing to expected utility does not work as a general solution to the mismatch problem.

Let's start by considering a standard example of the mismatch problem for Act Consequentialism.

> *Two Shooters*: You and I are sharpshooters. We shoot at an innocent victim simultaneously, our bullets striking the same fatal location in the victim's chest. Neither of us could have prevented the other from shooting, and either shot is sufficient for the victim's immediate death. The probablity that the victim dies given that at least one of us shoots is 1. The probability that the victim dies given that neither one of us shoots is 0. Before we shoot, the probability that I will shoot is .99. Similarly, the probability that you will shoot is also .99. What actually happens is that we both shoot.

According to *Act Consequentialism*, an act is morally permissible just in case there's no alternative with a better outcome. In Two Shooters, neither of us could have prevented the other from shooting. Thus, when we think about what would have happened had you holstered your weapon, we are to imagine a counterfactual world in which I shoot the victim just as I do in the actual world. Similarly, when we think about what would have happened had I holstered my weapon, we are to imagine a counterfactual world in which you shoot the victim just as you do in the actual world.

Accordingly, Act Consequentialism fails to condemn either one of our individual acts. I don't have an alternative with a better outcome: you would shoot and instantly kill the victim even were I to holster my weapon instead. And you don't have an alternative with a better outcome either: I would shoot and instantly kill the victim even were you to holster your weapon instead. The victim dies no matter what either of us, individually, does. On the other hand, we could have both holstered our weapons, which would have resulted in a better outcome. We *together* have acted wrongly according to Act Consequentialism. Paradoxically, two individual rights make a collective wrong. Act Consequentialism delivers mismatched verdicts between how we act as individuals and how we act together.

It's natural to think that the mismatch problem will not arise under expected utility versions of consequentialism. Before either of us shoots the victim, the probability that the other will not shoot is .01. Expected utility versions of consequentialism make this probability relevant to the moral appraisal of our acts. Imagine that you are deciding whether to shoot the victim. The probability that the victim lives given that you don't shoot is .01. On the other hand, the probability that the victim lives given that you do shoot is 0. Thus, given some natural assumptions about the value of the consequence in which the victim dies as opposed to the value of the consequence in which he lives, the expected utility of holstering your weapon is greater than the expected utility of your shooting the victim. Similarly, the expected utility of my holstering my weapon is greater than the expected utility of my shooting the victim. According to *Expected Utility Act Consequentialism* (EUAC), an act is morally permissible just in case there's no alternative with a higher expected utility. Each of our individual acts is morally wrong according to EUAC. So it's not the case that two rights make a wrong under EUAC. The mismatch problem seems to be averted.

Frank Jackson mentions this approach explicitly in connection with Two Shooters.[1] Many others haved defended this same approach in structurally similar cases. Like Two Shooters, some voting cases involve an overdetermined bad outcome. Suppose that an inferior candidate wins an election by many votes. Derek Parfit argues that consequentialists should appeal to expected utility to explain why individuals act wrongly in voting for the inferior candidate.[2] Factory farming also involves an overdetermined bad outcome. Suppose the same number of animals will be factory-farmed for meat whether I purchase factory-farmed chicken today. Peter Singer, Alastair Norcross, and Shelly Kagan argue that consequentialists should appeal to expected utility to explain why I act wrongly in purchasing factory-farmed chicken.[3]

I think that the exact nature of the problem that Two Shooters poses for consequentialism hasn't been fully appreciated. To see whether EUAC avoids the problem, we must examine more carefully how an appeal to expected utility is supposed to head off the possibility of mismatched verdicts in Two Shooters. Once we understand exactly how the appeal to expected utility is supposed to work, we will discover that the problem of mismatched verdicts persists under EUAC. Furthermore, as I will point out, the problem of mismatched verdicts enjoys a particularly troubling variant under EUAC that cannot arise under Act Consequentialism.

In section 3.2, I explain how I'm understanding the problem in Two Shooters.[4] In section 3.3, I explain what it takes for EUAC to solve the problem. In section 3.4, I offer a slightly modified version of Two Shooters to illustrate why EUAC fails to solve the problem. In section 3.5, I explore an objection related to expected

[1] See Jackson (1997).

[2] See (Parfit, 1984, 73-74).

[3] See Singer (1980), (Norcross, 2004, 232-233), and Kagan (2011). For other arguments that make an appeal to expected utility in connection with bad outcomes brought about by groups, see (Gibbard, 1990, 26-27), and (Regan, 2000, 69-70).

[4] Readers familiar with Chapter 1 may skip this section.

utility calculations and I offer some replies. And in section 3.6, I discuss a variation of the problem of mismatched verdicts that arises under EUAC but not under Act Consequentialism.

3.2 Background: The Mismatch Problem in Two Shooters

As I have explained in more detail in Chapter 1, a clear understanding of the problem that Two Shooters poses for consequentialism requires that we think about the normative appraisal of group acts in a certain way. The problem arises because a natural extension of Act Consequentialism to group acts condemns what we together do, yet fails to condemn what each of us does individually. Thus, we should grant that groups act, at least in the following minimal sense: a group act is simply a set of individual acts. For our purposes here, we may assume that any set of individual acts composes a group act.[5] In Two Shooters, the group act is [I shoot, you shoot]. For our present purposes, this is not to be taken as a metaphysical claim about the existence of group acts. Rather, it will be a convenient way of thinking about group acts for the sake of clarity in exposition.

An act is wrong under Act Consequentialism only if there's an alternative with a better outcome. Accordingly, if the theory condemns what we together do in Two Shooters, then we must grant that group acts have alternatives. We may assume that a group act has an alternative for any compossible combination of individual alternatives.[6] Under this way of thinking about group alternatives, we have three

[5] Jackson accepts this framework in Jackson (1987). So does Wlodzimierz Rabinowicz in Rabinowicz (1988).

[6] Why the appeal to compossibility? Suppose you and I are at the salon. I can fit into the tanning booth, and you can fit into the tanning booth, but we cannot both fit into the tanning booth. So if we both opt out of using the tanning booth, it doesn't follow that [I use the tanning booth, you use the tanning booth] is one of our alternatives. It is not a compossible combination of your and my individual alternatives. In this paper, all the examples will be ones for which all sets of individual alternatives are compossible.

alternatives in Two Shooters. These are [I shoot, you holster], [I holster, you shoot], [I holster, you holster].

Finally, it is important to emphasize that the problem of mismatched verdicts cannot arise unless we are willing to grant that group acts are subject to moral evaluation.[7] Without allowing for the normative appraisal of group acts, it's hard to see why Two Shooters poses a problem. The paradoxical implication of Act Consequentialism is that two rights make a wrong. This cannot happen unless the group act is wrong.

The *mismatch problem* arises for a moral theory whenever the theory delivers mismatched verdicts between a group act and the individual acts that compose it. In Two Shooters, we have an alternative with a better outcome, namely [I holster, you holster]. So our group act is morally impermissible according to Act Consequentialism. But I don't have an alternative with a better outcome. And neither do you. So you act permissibly according to Act Consequentialism, and so do I. This is paradoxical: the theory says one thing about the group act and a different thing about each contributing individual act. Thus, Two Shooters gives rise to the mismatch problem for Act Consequentialism.

The mismatch problem arises under EUAC if there's some case in which the group act has an alternative with a greater expected utility though no individual contributing act has an alternative with a greater expected utility. As I will argue in what immediately follows, Two Shooters is not such a case. But it would be a mistake to conclude on these grounds that EUAC is immune to the mismatch problem. Indeed, as I will argue in Section 3.4, a slight modification to Two Shooters is sufficient to establish that moving to EUAC doesn't solve the mismatch problem for consequentialism.

[7] For arguments that some group acts are morally wrong, see Jackson (1987) and Killoren and Williams (2013).

3.3 Explaining the Expected Utility Solution

3.3.1 Individual Acts and Expected Utility

To see why Two Shooters does not give rise to the mismatch problem for EUAC, we must understand how to calculate expected utilities. Suppose some possible act, a, may lead to several possible outcomes, $O_1, O_2, \ldots, O_n$. Suppose that for each outcome, O_i, there is an amount of value, V, associated with O_i. Suppose also that for each of these outcomes, there is a probability, $Pr(O_i \mid a)$. This is the conditional probability of outcome O_i on action a. The expected utility of a is the sum, for all these possible outcomes, of $V(O_i) * Pr(O_i \mid a)$.[8]

There are several different interpretations of $Pr(O_i \mid a)$. Intuitively, it is the likelihood that outcome O_i will occur, if a is performed. But there are at least three ways to interpret "likelihood" in the context of an expected utility calculation. It may derive from the agent's actual degree of belief in the proposition O_i *occurs given that I do a*—the credential interpretation. It may instead derive from the degree of belief in this proposition that's justified by the agent's evidence—the evidential interpretation. As a third option, it may derive from the objective chance that O_i occurs given that the agent does a—the objective interpretation. For now, I'll conduct the discussion in a neutral way. I'll assume that the probabilities specified in Two Shooters can be understood under the credential, evidential, and objective interpretations. For simplicity of exposition, I won't defend this assumption until Section 3.5.

There's also a temporal dimension to $Pr(O_i \mid a)$ that's important to keep in mind. Suppose we operate under the credential interpretation of expected utility. Then $Pr(O_i \mid a)$ may vary through time. I might believe firmly in the proposition O_i *occurs given that I do a* one moment, but become doubtful about this same thing just a moment later. In Two Shooters, I might have a credence of .99 in the proposition

[8] I take this formulation of expected utility from Fred Feldman; see Feldman (2006).

the victim dies given that I holster my weapon before you shoot the victim, and a credence of 1 in this same proposition after you in fact shoot him. For this reason, I will assume that the calculation of expected utilities in Two Shooters is to occur in the moments before we fire our weapons.

Calculating the expected utility of my shooting the victim and my not shooting the victim in Two Shooters will serve as an illustration of the concept of expected utility. There are two possible outcomes in the case: either 1 DIES, or 0 DIE. Each of these gets assigned a value. Say that

$$V(1\ \textsc{dies}) = -1$$
$$V(0\ \textsc{die}) = 0.$$

The conditional probabilities are specified by the case. The probability that the victim dies given that at least one of us shoots him is 1. So

$$Pr(1\ \textsc{dies} \mid \text{I}\ \textsc{shoot}) = 1$$
$$Pr(0\ \textsc{die} \mid \text{I}\ \textsc{shoot}) = 0$$

Thus, the expected utility of my shooting the victim is

$$\begin{aligned} eu(\text{I}\ \textsc{shoot}) &= (-1)(1) + (0)(0) \\ &= -1 \end{aligned}$$

On the other hand, the probability that the victim dies given that I don't shoot is given by the probability that you shoot, .99. So

$$Pr(1\ \textsc{dies} \mid \text{I}\ \textsc{holster}) = .99$$
$$Pr(0\ \textsc{die} \mid \text{I}\ \textsc{holster}) = .01$$

Thus, the expected utility of my not shooting the victim is

$$\begin{aligned} eu(\text{I HOLSTER}) &= (-1)(.99) + (0)(.01) \\ &= -.99 \end{aligned}$$

These calculations allow us to see that $eu(\text{I SHOOT}) < eu(\text{I HOLSTER})$. That means that I have an alternative to shooting with a higher expected utility. According to EUAC, I act wrongly if I shoot. Mutatis mutandis, EUAC condemns your shooting the victim as well.

3.3.2 Group Acts and Expected Utility

Recall that Act Consequentialism delivers a verdict of moral permissibility on each of our individual acts in Two Shooters. EUAC, on the other hand, delivers verdicts of individual wrongdoing. This, I take it, is what leads people to think that EUAC is equipped to solve the mismatch problem. It's important to notice, however, that demonstrating that EUAC locates individual wrongdoing in Two Shooters does not establish that EUAC resolves the mismatch problem. In order to establish that, we would need to calculate the expected utility of our group act and each of our alternatives. We would need to see that [I shoot, you shoot] has a lower expected utility than at least one of our alternatives (presumably [I holster, you holster]). For only then would we have established that EUAC delivers the same verdict at the group and individual levels.

Does it really make sense to suppose that group acts have expected utilities? If group acts do not have expected utilities, then EUAC cannot resolve the mismatch problem. [I shoot, you shoot] is impermissible under EUAC only if at least one of our group alternatives has a higher expected utility. And [I shoot, you shoot] is permissible under EUAC only if none of our group alternatives has a higher expected utility. So if our group alternatives lack expected utilities altogether, then EUAC does not deliver a verdict on our act. Assume that some acts are neither permissible nor impermissible. Call such acts *morally undefined*. If group acts do not have expected utilities, then [I shoot, you shoot] is morally undefined under EUAC. But then Two Shooters produces a mismatch problem for EUAC. The theory says one thing about the group (morally undefined behavior) while saying a different thing about each individual (morally wrong action). Thus, if EUAC is to resolve the mismatch problem, we must assume that group acts have expected utilities.

In a later section, I'll discuss some complications that arise under the credential and evidential interpretations of expected utility. These interpretations seem to require that groups have credences in propositions, and furthermore that groups have evidence that justifies credences. I'll return to this issue in Section 3.5. For now, we can read off the conditional probabilities from the case. As stipulated in Two Shooters, the probablity that the victim dies given that at least one of us shoots is 1. So

$$\begin{aligned} Pr(1 \text{ DIES} \mid [\text{I SHOOT, YOU SHOOT}]) &= 1 \\ Pr(0 \text{ DIE} \mid [\text{I SHOOT, YOU SHOOT}]) &= 0 \end{aligned}$$

Thus, the expected utility of [I shoot, you shoot] is

$$\begin{aligned} eu([\text{I SHOOT, YOU SHOOT}]) &= (-1)(1) + (0)(0) \\ &= -1 \end{aligned}$$

The probability that the victim dies given that neither one of us shoots is 0. So

$$\begin{aligned} Pr(1 \text{ DIES} \mid [\text{I HOLSTER, YOU HOLSTER}]) &= 0 \\ Pr(0 \text{ DIE} \mid [\text{I HOLSTER, YOU HOLSTER}]) &= 1 \end{aligned}$$

Thus, the expected utility of [I holster, you holster] is

$$\begin{aligned} eu([\text{I HOLSTER, YOU HOLSTER}]) &= (-1)(0) + (0)(1) \\ &= 0 \end{aligned}$$

which shows that $eu([\text{I SHOOT, YOU SHOOT}]) < eu([\text{I HOLSTER, YOU HOLSTER}])$. That means that EUAC condemns what we together do in Two Shooters. So there's no mismatch of verdicts. EUAC says that our group act is impermissible and it says that each individual act is impermissible as well. Thus, we see that EUAC resolves the mismatch problem in Two Shooters provided that group acts have expected utilities.

3.4 The Mismatch Problem for EUAC

EUAC doesn't resolve the mismatch problem for a simple modification of Two Shooters. Let's imagine that there's an even worse outcome than the death of the victim that would come about were only one of us to shoot.

> *Two Shooters+*: Same as Two Shooters, but our guns are connected to an explosive device that lies beneath us. The device will explode just in case only one of us shoots the victim; if both of us shoot the victim, or if neither of us shoots the victim, the device will not explode. Since there is a probability of .99 that you will shoot the victim, the probability that the device explodes given that I holster my weapon is .99. Accordingly, the probability that the device explodes given that I shoot the victim is .01. So the conditional probabilities are these:
>
> $$\begin{aligned} Pr(3 \text{ DIE} \mid \text{I HOLSTER}) &= .99 \\ Pr(0 \text{ DIE} \mid \text{I HOLSTER}) &= .01 \\ Pr(1 \text{ DIES} \mid \text{I SHOOT}) &= .99 \\ Pr(3 \text{ DIE} \mid \text{I SHOOT}) &= .01 \end{aligned}$$
>
> Since there is a probability of .99 that I will shoot the victim, the probabilities are similar if we consider your options:
>
> $$\begin{aligned} Pr(3 \text{ DIE} \mid \text{YOU HOLSTER}) &= .99 \\ Pr(0 \text{ DIE} \mid \text{YOU HOLSTER}) &= .01 \\ Pr(1 \text{ DIES} \mid \text{YOU SHOOT}) &= .99 \\ Pr(3 \text{ DIE} \mid \text{YOU SHOOT}) &= .01 \end{aligned}$$

> We together have four options corresponding to the four possible group acts: [I shoot, you shoot], [I shoot, you holster], [I holster, you shoot], and [I holster, you holster]. And given the aforementioned features of the case, the probabilities are these:
>
> $$\begin{aligned} Pr(1 \text{ \textsc{dies}} \mid [\text{I \textsc{shoot}, \textsc{you shoot}}]) &= 1 \\ Pr(3 \text{ \textsc{die}} \mid [\text{I \textsc{shoot}, \textsc{you holster}}]) &= 1 \\ Pr(3 \text{ \textsc{die}} \mid [\text{I \textsc{holster}, \textsc{you shoot}}]) &= 1 \\ Pr(0 \text{ \textsc{die}} \mid [\text{I \textsc{holster}, \textsc{you holster}}]) &= 1 \end{aligned}$$

It's important to emphasize that the only difference between Two Shooters and Two Shooters+ is the addition of a worse possible outcome in the latter case. Our options remain the same: each of us can either shoot or holster. The probabilities on our individual acts remain the same: there's a probability of .99 that you shoot the victim and a probability of .99 that I shoot him. The best that we could together do remains the same: no one would die if we were both not to shoot. The only difference is that if you don't shoot the victim and yet I do (or if I don't shoot the victim and yet you do), then we die along with the victim.

Accordingly, I would take a huge risk were I not to shoot the victim in Two Shooters+. I would risk the likely death of three people at the prospect of a tiny chance at saving everyone. Similarly, you would take a huge risk were you not to shoot the victim. In trying to secure the best that we could together bring about, you would likely bring about the worst instead. On the other hand, we together would take no risks were we both to holster; this group act is guaranteed to bring about the best possible outcome.

Thus, it is fairly straightforward to see that EUAC delivers mismatched verdicts in Two Shooters+. First, we have to calculate the expected utility of I shoot and compare it with the expected utility of I holster.[9] Second, we have to calculate the

[9] Since the conditional probabilities are the same on you shoot and you holster respectively, the expected utilities will be the same for you shoot and you holster, respectively. So we have to run the calculation only once.

expected utility of [I SHOOT, YOU SHOOT] and compare it with the expected utilities of the alternative group acts.

First, the moral assessment of my act. There are two possible outcomes on I SHOOT. These are 1 DIES and 3 DIE. Let's say that $V(3 \text{ DIE}) = -3$. There are two possible outcomes on I HOLSTER. These are 3 DIE and 0 DIE. Thus, $eu(\text{I SHOOT})$

$$= V(1 \text{ DIES}) * Pr(1 \text{ DIES} \mid \text{I SHOOT}) + V(3 \text{ DIE}) * Pr(3 \text{ DIE} \mid \text{I SHOOT})$$

$$= (-1)(.99) + (-3)(.01)$$

$$= -1.02$$

And $eu(\text{I HOLSTER})$

$$= V(3 \text{ DIE}) * Pr(3 \text{ DIE} \mid \text{I HOLSTER}) + V(0 \text{ DIE}) * Pr(0 \text{ DIE} \mid \text{I HOLSTER})$$

$$= (-3)(.99) + (0)(.01)$$

$$= -2.97$$

This shows that $eu(\text{I SHOOT}) > eu(\text{I HOLSTER})$. So, according to EUAC, I act morally permissibily in Two Shooters+. Since the conditional probabilities are the same on your acts, we could run a similar set of calculations to show that you act morally permissibly in Two Shooters+ as well.

But *we* do not act permissibly according to EUAC. There is one possible outcome on [I SHOOT, YOU SHOOT]. This is 1 DIES. Similarly, there is one possible outcome on each of our alternatives. On [I SHOOT, YOU HOLSTER], it's 3 DIE. On [I HOLSTER, YOU SHOOT], it's 3 DIE. And on [I HOLSTER, YOU HOLSTER], it's 0 DIE. So, we may calculate the expected utilities of the group acts as follows:

$eu([\text{I SHOOT, YOU SHOOT}])$

$$= V(1 \text{ DIES}) * Pr(1 \text{ DIES} \mid [\text{I SHOOT, YOU SHOOT}])$$

$$= (-1)(1)$$

$$= -1$$

$eu([\text{I } \textsc{shoot}, \textsc{ you holster}])$

$$\begin{aligned} &= V(3 \textsc{ die}) * Pr(3 \textsc{ die} \mid [\text{I } \textsc{shoot}, \textsc{ you holster}]) \\ &= (-3)(1) \\ &= -3 \end{aligned}$$

$eu([\text{I } \textsc{holster}, \textsc{ you shoot}])$

$$\begin{aligned} &= V(3 \textsc{ die}) * Pr(3 \textsc{ die} \mid [\text{I } \textsc{holster}, \textsc{ you shoot}]) \\ &= (-3)(1) \\ &= -3 \end{aligned}$$

$eu([\text{I } \textsc{holster}, \textsc{ you holster}])$

$$\begin{aligned} &= V(0 \textsc{ die}) * Pr(0 \textsc{ die} \mid [\text{I } \textsc{holster}, \textsc{ you holster}]) \\ &= (0)(1) \\ &= 0 \end{aligned}$$

This shows that $eu([\text{I } \textsc{shoot}, \textsc{ you shoot}])$ is less than the expected utility of one of its alternatives, namely [I HOLSTER, YOU HOLSTER]. This shouldn't be surprising. According to the conditional probabilities stipulated in the case, the survival of everyone is guaranteed if we both don't shoot. And the death of at least one person is guaranteed given the way we act. So, according to EUAC, the group act is morally wrong in Two Shooters+.

Therefore, Two Shooters+ gives rise to the mismatch problem for EUAC. The theory says one thing about the group act (that it's morally impermissible) and a different thing about each individual act (that it's morally permissible). This shows that moving to expected utility does not dissolve the mismatch problem for consequentialism.[10]

3.4.1 Explaining the Mismatch in Two Shooters+

It takes only a moment's reflection on a simple fact about expected utility to see why it's natural to expect mismatches for EUAC. Recall that in calculating the expected utility of an act, we calculate the likelihoods of various possible outcomes occuring if that act is performed. But one and the same outcome may have a different likelihood of occuring on different acts. Suppose I bend my left leg while standing. The likelihood that my hat gets closer to the ground is very low. Suppose I bend my right leg while standing. Again, the likelihood that my hat gets closer to the ground is very low. And yet suppose that I bend both my left leg and my right leg while standing. Now the likelihood that my hat gets closer to the ground is very high.

We can see that the likelihood of one and the same outcome may be different under a set of acts than it is under any one of the individual acts. Since a group act is a set of acts, the likelihood of an outcome occuring if a group act is performed may be very different from the likelihood of that same outcome occuring if any one of the component individual acts is performed. Indeed, this is exactly what happens in Two Shooters+. In the case, 0 DIE is guaranteed on the assumption that [I HOLSTER, YOU HOLSTER] is performed and yet it's very unlikely on the assumption that either one of the component individual acts is performed. Instead, on the assumption that I HOLSTER is performed, 3 DIE is very likely. Given the difference in value between

[10] There's another version of Two Shooters that gives rise to the mismatch problem under the credential interpretation of expected utility: imagine that each of us is *certain* that the other will shoot. See Nefsky (2012a) and Pinkert (2015). In response to such a case, Kagan concedes that the move to expected utility is supposed to work only under the condition of individual uncertainty in Kagan (2011). Two Shooters+ meets the condition of individual uncertainty and yet still gives rise to the mismatch problem for EUAC.

0 DIE and 3 DIE, we see a divergence between the expected utility of the group act and the expected utilities of the individual acts.

Once we understand the basic mechanism for the divergence between the likelihood of an outcome on a group act and on its individual component acts, we should realize that mismatches for EUAC are quite common. Imagine that a good outcome is guaranteed only if we both cooperate. Imagine that it's very unlikely that you will cooperate and very unlikely that I will cooperate. And imagine that each of us would make things worse overall by being the sole cooperator. Then supposing that I cooperate, the likelihood of a good outcome is very low. But supposing that both of us cooperate, the likelihood of that same good outcome is guaranteed. In any case of this form, EUAC says that the group act is wrong when we both defect, but it says that each individual defection is morally permissible.

What is it, precisely, that explains why the mismatch problem arises for EUAC in connection with Two Shooters+ but not in connection with the original Two Shooters case? In Two Shooters, the same bad outcome arises whether one of us fails to holster or both of us do. In Two Shooters+, on the other hand, a worse possible outcome results if only one of us holsters. Since it's very likely that the other will not holster, this significantly lowers the expected utility of holstering in Two Shooters+. That's why shooting the victim has a greater expected utility than not shooting the victim in Two Shooters+, which results in the mismatch problem for EUAC.

3.5 More About the Expected Utility of Group Acts

I've put off until this section a discussion of the different interpretations of expected utility. A potential worry for the Two Shooters+ case is that the conditional probability assignments in the case cannot be sustained under some ways of thinking about how to extend the concept of expected utility to group acts. For example, suppose we use the credential interpretation of expected utility. Then, in Two Shooters+, we are supposed to imagine that the group has a degree of belief of 1 in the proposition *only the victim dies given that we perform* [I SHOOT, YOU SHOOT]. But, we may worry that groups do not have degrees of belief. If groups do not have degrees of belief, then the conditional probability attribution to the group act cannot be sustained.

It's important to recognize what lies at the heart of this worry. Suppose you and I have different credences in a conditional on b. Then we need to adopt some aggregation procedure to determine *our* credence in a conditional on b. It can be difficult, maybe even impossible to settle on an aggregation procedure. This, I believe, represents the resistence to attributing credences to groups of people.

But there are some cases in which the difficulty associated with non-uniform credences need not arise. These are cases in which you and I have the same credences in all the relevant propositions. In such cases, *our* credence in some proposition is simply either of our individual credences in that proposition.[11]

There's a version of Two Shooters+ in which you and I have the same credences in all the relevant propositions. Let's imagine that each of us has a credence of 1 in *author's shot will kill victim*, *reader's shot will kill victim*, *three will die if author holsters and reader shoots*, *three will die if reader holsters and author shoots*, and *no one dies if both holster*. Then it's straightforward to attribute credences to the group that sustain the assignment of conditional probabilities in Two Shooters+.

On the other hand, consider a version of Two Shooters+ in which you and I do not have uniform credences. If there's no aggregation procedure to settle our group credences, then there is no way to determine what EUAU implies about the moral obligations of the group. Thus, EUAC fails to deliver a verdict on [I SHOOT, YOU SHOOT], producing a mismatch problem of a different but related sort. The theory

[11] See (Hylland and Zeckhauser, 1979, 1323).

says one thing about the group (morally undefined behavior) while saying a different thing about each individual (morally wrong action).

Mismatches of this sort may be even more plausible under the evidential interpretation of expected utility. In order to assign a conditional probability of 1 to [I SHOOT, YOU SHOOT], we are to imagine that the group's evidence justifies a degree of belief 1 of in the proposition *only the victim dies given that we perform* [I SHOOT, YOU SHOOT]. But we may worry that groups do not have evidence. If you and I do not have uniform bodies of evidence, it can be difficult, maybe even impossible to settle on an aggregation procedure. EUAC will fail to deliver a verdict on [I SHOOT, YOU SHOOT].

Of course, there are some situations in which a group is made up of people with uniform evidence. Imagine a group of people in a jury who all form exactly the same beliefs on the basis of the evidence presented during the trial. Imagine that the individual jury members start with the same background beliefs. Then the jury is made up of individuals all of whom have the same credences, and all of whom have the same evidence. In such a situation, it's reasonable to assume that the group's evidence is the same as any individual's evidence.

We can see that there's a version of Two Shooters+ that sustains the attributions of conditional probabilities under the evidential interpretation. Imagine that, before either of us shoots, each is presented with the same conclusive evidence about both the author's and the reader's sharpshooting ability and the workings and effects of the explosive device. In this case, *our* evidence is the same as my and your evidence. We can then safely assume that our evidence justifies a credence of 1 in the proposition *only the victim dies given that we perform* [I SHOOT, YOU SHOOT] and a credence of 1 in the proposition *no one dies given that we perform* [I HOLSTER, YOU HOLSTER].

In this version of Two Shooters+, the mismatch problem arises for EUAC under the evidential interpretation of expected utility.[12]

3.6 A Variation of the Mismatch Problem for EUAC

Say that an *ought-not/ought* mismatch occurs when a moral theory deems a group act morally impermissible while the same theory requires every individual act of participation. (An act is *morally required*, or *morally obligatory*, just in case it is morally permissible and none of its alternatives is.) Say that an *ought/ought-not* mismatch occurs when a moral theory deems a group act morally obligatory while the same theory condemns every individual act of participation. The version of the mismatch problem that arises for EUAC in connection with Two Shooters+ is an ought-not/ought mismatch. EUAC says that we act wrongly. But since I must either shoot or holster, and since $eu(\text{I SHOOT}) > eu(\text{I HOLSTER})$, EUAC says that I am morally obligated to shoot. Similarly, EUAC says that you are morally obligated to shoot as well.

In this section, I will argue first that ought/ought-not mismatches are possible under EUAC. I will then argue that ought/ought-not mismatches are impossible under Act Consequentialism. This is sufficient to show that, between these two theories, there's a variation of the problem of mismatched verdicts that arises only under EUAC.

It is relatively straightforward to see that ought/ought-not mismatches arise under EUAC. We need only consider a small modification to Two Shooters+.

[12] I must note here that I cannot see how any special difficulties will arise in connection with the assignment of objective probabilities to group acts. You and I might have different credences in *a* conditional on *b*, and you and I might have different bodies of evidence justifying *a* conditional on *b*, but the objective probability of *a* conditional on *b* cannot differ between us. Thus, I see no reason to worry about the attributions of conditional probabilities in Two Shooters+ under the objective intepretation of expected utility.

> *Two Shooters++*: Same as Two Shooters+, but instead of each of us shooting the victim, we actually perform [I holster, you holster]. Neither of us could have prevented the other from not shooting.

The group act [I holster, you holster] is morally obligatory under EUAC—as we saw above, the expected utility of [I holster, you holster] is greater than the expected utility of all the group's alternatives. On the other hand, since $eu(\text{I HOLSTER}) < eu(\text{I SHOOT})$, it is morally wrong for me not to shoot. Similarly, it is morally wrong for you not to shoot. So in Two Shooters++, two wrong individual acts make for an obligatory group act. The case gives rise to an ought/ought-not mismatch for EUAC.

It's natural to expect that Two Shooters++ involves an ought/ought-not mismatch under Act Consequentialism as well. But it doesn't. Notice that I HOLSTER is morally permissible according to Act Consequentialism. That's because, in Two Shooters++, we are assuming that you also holster, and we are also assuming that I could not have prevented you from doing this. Thus, when we think about what would have happened had I shot the victim instead, we are to imagine a counterfactual world in which you holster your weapon just as you do in the actual world. In this counterfactual world, the explosive device explodes, and we all die. Accordingly, I don't have an alternative with a better outcome. For parallel reasons, you don't have an alternative with a better outcome either. Thus, Act Consequentialism says that each of us acts morally permissibly in Two Shooters++, and so the case does not give rise to an ought/ought-not mismatch for Act Consequentialism.

There's a quick argument that ought/ought-not mismatches are impossible under Act Consequentialism. Suppose, for reductio, that

> (a) A group act is morally obligatory according to Act Consequentialism.
> (b) At least one of the individual acts is morally impermissible according to Act Consequentialism.

If (a) is true, then the group has no alternative way of acting with a better outcome. So were any individual member to act differently, it would make things no better.

But if (b) is true, then there is at least one individual member who does have a way of acting that would make things better. Contradiction.[13]

One might wonder why ought/ought-not mismatches are possible under EUAC and not under Act Consequentialism. The answer, I think, is this. If an outcome is accessible to an individual, then it is accessible to every group of which that individual is a part.[14] So if an individual could do better, then so could his or her group. However, it is not true that the conditional probability of an outcome on an individual act is the same as the conditional probabilities of that outcome on every group act containing it. So if an individual act has an alternative with a higher expected utility, it doesn't follow that it's a part of a group act with an alternative with a higher expected utility.

This makes EUAC vulnerable to a version of the mismatch problem that's not a worry for regular old Act Consequentialism. This is yet another reason why philosophers shouldn't make the move to expected utility if they are concerned about the mismatch problem. Thus, when it comes to resolving the mismatch problem for consequentialism in a general way, an appeal to expected utility does not hold the answer.

[13] For more rigorous versions of this argument, see Chapter 7 of Feldman (1986) and Chapter 9 of Zimmerman (1996).

[14] An outcome is accessible to an individual if one of the individual's alternatives has that outcome. An outcome is accessible to a group if one of the group's alternatives has that outcome. To see why any outcome accessible to an individual is also accessible to every one of that individual's groups, suppose some outcome, y, is accessible to me. I in fact bring about x by doing ϕ, but I could have brought about y instead, had I pursued my alternative ψ. Then y is compatible with the way that others would have acted had I done ψ. But that just means that any of my groups could have brought about y. And so y is accessible to every group to which I belong.

CHAPTER 4

KAGAN'S EXPECTED UTILITY SOLUTION

4.1 Introduction

As we have seen in previous chapters, the mismatch problem for Act Consequentialism arises in connection with a wide variety of cases. In a recent article, Shelly Kagan tries to resolve the mismatch problem for a specific type of case.[1] 'Kagan-style' cases involve large numbers of people; there is some type of act such that each of these people has the ability to perform an act of that type; each individual act does not make a difference for the worse, but when enough such acts are performed, the result is bad overall. Kagan argues that Expected Utility Act Consequentialism (EUAC) does not deliver mismatched verdicts in connection with such cases. In this chapter, I first identify the essential features of Kagan-style cases. In section 4.3, I explain why Kagan thinks that EUAC solves the mismatch problem in connection with these cases. In section 4.4, I show why EUAC fails to solve the mismatch problem for all Kagan-style cases.

4.2 What are Kagan-Style Cases?

Before identifying the central features of Kagan-style cases, it is important to be clear about the problem that Kagan thinks these cases pose for consequentialism. I believe that, although Kagan does not frame the problem as one of 'mismatched

[1] See Kagan (2011).

verdicts', it is plausible to interpret Kagan as focused on the mismatch problem for Act Consequentialism. To see this, consider the following distinction.

Some challenges to Act Consequentialism are *external* to the theory: a certain pretheoretic idea about what determines whether an act is right or wrong cannot be captured by the consequentialistic framework. For an example, consider the familiar organ harvest example. In the service of producing the overall best outcome available to her, a doctor cuts up a healthy patient in order to save five patients in need of organ transplants. Intuitively, the doctor acts wrongly: she violates the healthy patient's right to autonomy, and she murders an innocent patient who is seeking her medical help. According to pretheoretic common sense, an act like this is never permissible. The external challenge that arises in connection with the organ harvest example is to see whether Act Consequentialism can be modified to accommodate this pretheoretic idea.

But other challenges to Act Consequentialism are *internal* to the theory: a certain inconsistency is discovered in the theory's formulation, or the theory delivers contradictory advice, or the theory is self-defeating. In connection with internal challenges, Act Consequentialism appears to fail by its own lights. The mismatch problem is one such internal challenge.

In his article, Kagan says he is interested in resolving an internal challenge to Act Consequentialism. The cases he discusses are ones for which "consequentialism appears to fail even in its own favored terrain, where we are concerned with consequences and nothing but consequences."[2] What precisely leads consequentialism to fail? In Kagan-style cases, a group brings about a suboptimal outcome. But each individual member of the group could not have done better by acting differently. As Kagan sees it, the problem is that "the acts in question need to be condemned be-

[2] (Kagan, 2011, 107)

cause of the results that eventuate from everyone's performing them... [y]et despite this, 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."[3] Thus, (whether he would frame it in precisely this way) the problem that Kagan is dealing with concerns an inconsistency within Act Consequentialism that arises when the theory condemns a group act, but apparently cannot condemn the individual acts. This problem is equivalent to the mismatch problem.

Kagan wants to resolve the mismatch problem for Act Consequentialism in connection with cases that have the following structure:

> A certain number of people—perhaps a large number of people—have the ability to perform an act of a given kind. And if a large enough group of people do perform the act in question then the results will be bad overall. However—and this is the crucial point—in the relevant cases it seems that it makes no difference to the outcome what any given individual does. And this is true regardless of whether others are doing the act or not. Thus, if enough people do perform the act the results are bad overall; but for all that, it remains true of each individual agent that it makes no difference to the overall results whether or not they perform the act in question.[4]

It will be helpful to highlight a few of the defining features of Kagan-style cases as described in this passage. First, let's make a distinction between act tokens and act types. An *act token* is a concrete action taken by a specific agent at a specific time and place. Act tokens are not repeatable. Distinct act tokens may fall under the same *act type*, or general way of acting. In many familiar situations, more than one person performs an act token of the same act type, and though a single tokening of this act type would not be bad, the result of some large number of people tokening this act type is bad overall. Some situations of this sort include the lawn crossing

[3] (Kagan, 2011, 107)

[4] (Kagan, 2011, 107)

problem (many people take a shortcut across a lawn on campus, killing the grass and damaging the overall appearance of the landscape) and the vegetarianism problem (many people purchase factory farmed chicken, thereby creating a market demand for cheap meat, and many new chickens suffer in factory farms as a result).[5] Say that an act type is *repeatedly bad* if the performance of one act token of that type doesn't have any bad results, but enough repeated performances of act tokens of that type have bad results overall. Each of the just-mentioned problems involves a repeatedly bad act type: *crossing the lawn* and *purchasing factory farmed chicken*.

A second defining feature of Kagan-style cases is that individual tokens of the repeatedly bad act type apparently have innocuous effects. The "crucial point" that applies to such cases is that "it seems that it makes no difference to the outcome what any given individual does. And this is true regardless of whether others are doing the act or not." No matter how many people perform acts of the repeatedly bad act type, it seems that no individual act token of that type ever makes a difference for the worse. To understand this feature of Kagan-style cases, it will be important to distinguish a few different cases from each other.

First, consider

> *Votes*$_{90}$: Ninety senators of a 100-member senate vote 'yes' on a bill with disastrous consequences. Each of these senators could have voted 'no'. Had at least forty of these senators voted 'no', the bill would not have passed. The results would have been much better. No senator had the power to influence any of the others.

I use the subscript '90' to indicate that ninety people have participated in the relevant group act. Notice that Votes$_{90}$ gives rise to the mismatch problem for Act Consequentialism. The group of 'yes' voters could have done much better, and yet no individual senator in this group could have done any better by voting 'no'. Fur-

[5] Perhaps another situation of this sort is found in the case of anthropogenic climate change discussed in Chapter 2. Though, as I suggest, it is difficult to specify the act type in connection with that case.

thermore, Votes_{90} is such that each individual senator performs an act of the same repeatedly bad act type. Were only one senator to vote 'yes', it wouldn't have any bad results, but enough repeated performances of this act type results in the passage of the disastrous bill. And yet Votes_{90} is not a Kagan-style case; it doesn't meet the condition that no individual act token seemingly ever makes a difference for the worse. We will see this in just a moment through comparing Votes_{90} with another case.

Next, consider

> *Chickens*$_{51}$: Every time 25 customers purchase factory farmed chicken carcasses at the grocery store, the butcher places an order for 25 more chicken carcasses. Each order results in the lifetime confinement and suffering of 25 chickens who would otherwise not have been born. 51 customers purchase chicken, so the butcher places two orders, and 50 chickens suffer as a result. No individual customer could have prevented the others from purchasing chicken.

Chickens_{51} also gives rise to the mismatch problem for Act Consequentialism. In Chickens_{51}, the chicken consumers together could have avoided producing chicken suffering entirely. But no individual consumer could have done better by foregoing his purchase. Furthermore, Chickens_{51} is such that each individual chicken consumer performs an act of the same repeatedly bad act type. Were only one person to have purchased a chicken carcass, the butcher would not have placed an order, and no new chickens would have suffered. But when enough people purchase chicken carcasses, the butcher places at least one order, and the result is overall bad.

Notice that in both of the foregoing cases, a group brings about an avoidable bad outcome. In Votes_{90}, the bad outcome is the passage of a disastrous bill. In Chickens_{51}, the bad outcome is the suffering of fifty chickens. But while both cases contain a bad outcome, there is a distinction to be drawn among these outcomes. The passage of the bill is bad, but it would have been no worse had more senators voted 'yes'. The bill would not have 'passed more' had it received more votes. On the other hand, 25 more chickens would have suffered had 24 more people purchased chicken. The overall chicken suffering could have been even worse. We may say that the outcome in Votes$_{90}$ is *on/off*. This means that, as a function of the number of people who token the repeatedly bad act type, it either happens, or it doesn't. Let's say that the bad outcome in Chickens$_{51}$ is *degreed*. This means that, as a function of the number of people who token the repeatedly bad act type, it can get worse.

As a third example of a case that involves a repeatedly bad act type, recall the situation involving the bean-stealing bandits from Chapter 1, which I formulate this time as follows.

> *Beans*$_{100}$: 100 people in a village are about to eat lunch. Each has a bowl containing 100 beans. Suddenly, a group of 100 bandits swoops down on the village. Each bandit takes one bean from each of the bowls of the 100 villagers. The result is 100 empty bowls, and 100 hungry villagers. Each bandit enjoys a tasty snack, but the total pleasure the bandits enjoy is outweighed by the total hunger the villagers suffer. Had any one bandit not participated in the raid, the other 99 still would have. The loss of one bean does not make a perceptible difference to any villager.

Beans$_{100}$ gives rise to the mismatch problem for Act Consequentialism. The bandit collective could have done better, though we are to imagine that no individual bandit could have done any better. We are supposing that the loss of one bean cannot make a perceptible difference, and so apparently one bean more or less cannot affect the well-being of any of the villagers. Had any bandit not participated, the other ninety-nine still would have—the single defecting bandit would have missed out on a tasty snack and apparently none of the villagers would have benefited. Note that each bandit performs an act of the same repeatedly bad act type, *stealing one bean from each villager*. Enough repeated performances of act tokens of this type has overall bad results. When all one hundred bandits steal in this way, all of the villagers go hungry.

Given the distinction between on/off and degreed outcomes presented in Votes$_{90}$ and Chickens$_{51}$, it is natural to reflect on the nature of the outcome in Beans$_{100}$. It seems to me that the outcome is degreed. Hunger is not an on/off sort of thing. And yet, it appears that the villagers would not suffer hunger to any degree were each to miss out on just one bean: the loss of one bean cannot make a perceptible difference, and hunger must be perceptible to be bad. Perhaps at around ten missing beans or so each villager would be a little bit hungry. Each villager is very hungry as a result of every bandit's participating in the raid.

In all three of the foregoing cases, the bad outcome brought about by the group act is a function of the number of individuals who will token the repeatedly bad act type. In Votes_{90}, the bad outcome will arise just in case at least 51 senators vote 'yes'. In Chickens_{51}, the bad outcome will become worse just in case some multiple of 25 people purchases a chicken carcass. In Beans_{100}, however, it is more difficult to specify the relevant function. It's not clear under what levels of participation a bad outcome will become worse.

We may describe the various functions from participation to bad outcome more perspicuously as follows. In Votes_{90}, since the outcome is on/off, the case very clearly has precisely *one threshold*. The bad outcome arises when and only when at least 51 senators vote 'yes'. In Chickens_{51}, since the bad outcome is degreed, the case has *multiple thresholds*. A bad outcome arises when at least 25 people buy chicken, and gets worse whenever an additional 25 people buy chicken. But in Beans_{100}, it's not clear whether the case has any thresholds. Since hunger is not an on/off matter, we must assume that the bad outcome in Beans_{100} is worse than what would have resulted had ten fewer bandits participated, which is worse than what would have resulted had twenty fewer bandits participated, and so on. But we are imagining that a single missing bean never makes a perceptible difference to hunger. Accordingly, there are apparently no sets of neighboring outcomes in which the bad outcome gets worse between them. We may say that Beans_{100} *seems to lack thresholds*.

We are now in position to clarify Kagan's "crucial point". Kagan-style cases are those in which it seemingly makes no difference to the outcome what any given individual does, regardless of whether others are performing the repeatedly bad act type or not. In other words, Kagan-style cases are those that seem to lack thresholds. So of the preceding three cases, only Beans_{100} satisfies the conditions of a Kagan-style case.

However, there is an argument to be made that Chickens_{51} also counts as a Kagan-style case. It has multiple thresholds, but does it *seem* to have no thresholds? *Seeming* is an epistemic relation. It takes a proposition—the thing that seems to be a certain way—and a subject—the person or people to whom it seems this way. In describing the problem cases, Kagan makes use of this epistemic relation. Kagan makes it clear that the cases he's interested in are those that seem to have no thresholds. But there's a certain matter of unclarity regarding the intended subject. To whom does it seem this way?

Here's the natural interpretation that figures in Kagan's presentation. We are first presented with a general characterization of a problem case according to which it seems to lack thresholds to *us*, the ethicists evaluating the case. Beans_{100} is like this. It's natural to imagine that a single bean cannot make a difference to well-being. It's natural to expect that no neighboring outcomes involve the bad outcome becoming any worse. But perhaps after we (the ethicists) think more carefully about the case, we will have something more exact to say about whether the appearance of no thresholds is just that—*an appearance*—or if the case really does have a no threshold structure.

Chickens_{51} does not count as a Kagan-style case under this interpretation. As the case is presented, we (the ethicists) are told that the bad outcome becomes worse at every multiple of 25 chicken purchases. It doesn't seem to us (the ethicists) that the case has no thresholds. Rather, it's quite obvious that the case has multiple thresholds.

But there's a second interpretation of the intended subject of the seeming relation according to which Chickens$_{51}$ counts as a Kagan-style case. Perhaps Kagan intends for us to focus on versions of the problem case that seem to *the participants* to have no thresholds. Chickens$_{51}$ is plausibly like this. When each chicken consumer deliberates about whether to token the repeatedly bad act type, each is confronted with the following epistemic situation: it seems to him or her that an individual act cannot make a difference to the bad outcome, and it seems this way no matter how many others are thought to be participating. Let's suppose that Ordinary Larry is one of the consumers in Chickens$_{51}$. He doesn't know the butcher. He doesn't know whether the store makes purchasing decisions based on multiples of some exact number of chicken carcasses sold. Instead, Larry quite reasonably believes that the chicken market is simply not sensitive to individual purchases. Perhaps Larry hasn't thought things through philosophically. Larry just reflects on the sheer size of the poultry industry and reasonably concludes that there are no thresholds. Larry is mistaken, of course. And we (the ethicists) recognize this because of the way the case has been presented to us. But Larry is a participant in the case, and what he understands about it may differ from what we understand about it.

We are now in position to precisely state the essential conditions on Kagan-style cases. First, the cases involve repeatedly bad act types. Second, Kagan-style cases seem to have a no-threshold structure, and they seem this way either to the participants, or to the ethicists conducting an evaluation of the case.

4.3 Explaining Kagan's Solution

Now that we have a characterization of Kagan-style cases, we may consider what Kagan says about them. Since Beans$_{100}$ meets both the conditions of a Kagan-style case, I will center our discussion on this case. In rough outline, Kagan's idea proceeds as follows. In Kagan-style cases—indeed, in any case that gives rise to the mismatch problem for Act Consequentialism—each individual *in fact* has no alternative with a better outcome; each individual *in fact* makes no difference for the worse. For this reason, Act Consequentialism fails to condemn any of the individual acts. But under EUAC, an act may be wrong even though it in fact makes no difference for the worse: in some situations, EUAC condemns an act because there is *a chance* that the act makes a difference for the worse. Kagan thinks that every Kagan-style case involves a situation of this sort: the probability that the relevant bad outcome occurs is greater under the individual tokening of a repeatedly bad act type than under some alternative. If Kagan is right, then EUAC condemns each individual act. On this basis, Kagan thinks that EUAC is equipped to deliver concordant verdicts: wrong group act, but wrong individual acts as well.

To apply Kagan's idea to Beans$_{100}$, we need to see why Kagan thinks each individual theft carries a chance of making things worse, even though each individual bandit would not in fact have done any better by not stealing. Roughly, Kagan believes that all cases involving repeatedly bad act types have thresholds. According to Kagan, the appearance of no thresholds in Beans$_{100}$ is just that: an appearance. Here's what Kagan says:

> It is never the case that a large enough number of acts make a morally relevant difference, but each individual act makes no perceptible difference at all. In effect, I want to concede that were there cases of this sort, they would indeed pose a serious challenge to the adequacy of consequentialism, but insist nonetheless that this fact needn't concern the consequentialist at all, for cases of this sort simply do not exist.[6]

In this passage, Kagan explicitly targets cases in which the impact of each additional performance of the repeatedly bad act type apparently can never be perceived. Beans$_{100}$ is a case of this sort: it doesn't seem like an additional missing bean can ever make a perceptible difference to the villagers. Kagan's claim is that, contrary

[6] (Kagan, 2011, 130)

to appearances, there must be some points at which the villagers end up hungrier than they would have been had each eaten one more bean. According to Kagan, at some point, for each villager the answer to the question "Are you hungry?" must go from "no" to "yes". (Otherwise, Kagan thinks we'd have to accept the absurd conclusion that the villagers are no hungrier in Beans_{100} than they would have been had no bandits participated.) Kagan claims that the apparent no threshold structure in Beans_{100} is impossible.[7]

But if there are thresholds in Beans_{100}, then the case is structurally like Chickens_{51}. In connection with that case, Kagan says that

> . . . I have a 1 in 25 chance of being part of a cohort where I can correctly say that had I acted differently, the results would have been better, and a 24 in 25 chance of being part of a cohort where I can correctly say that had I acted differently, the results would have been the same. This means, of course, that it is most likely that my act of purchasing a chicken made no difference. But I cannot actually know whether this is the case. Instead, I can only know the expected results of my act. If I am part of a cohort that makes up an exact multiple of 25, then it will be true of me that had I acted differently 25 chickens would have been spared their suffering. And I have a 1/25th chance of being a member of such a cohort. Thus the expected disutility of my act is still equivalent to one chicken's suffering. Even when we offset this by the pleasure I get from eating the chicken, the net expected utility of my purchasing a chicken remains negative. And so [expected utility act] consequentialism still condemns my act of purchasing a chicken.[8]

If each chicken purchaser doesn't know how many others will purchase chicken, then the probability that more chickens suffer is greater under an individual act of purchasing chicken than under refraining from purchasing chicken. So EUAC condemns each individual act in Chickens_{51}.

Similar reasoning applies to Beans_{100} provided that there are indeed thresholds. Suppose you are one of the bandits. If you are thinking about stealing one bean from

[7] For our purposes here, we need not engage with Kagan's arguments for this claim, but see Nefsky (2012a) for an insightful and critical discussion.

[8] (Kagan, 2011, 126-127)

each villager, you are sure to have an alternative with a higher expected utility. For some numbers of participating bandits, were one fewer bandit to participate, the bad outcome would not be as bad. You do not know for certain how many people will participate, or, if you do, you don't know exactly where the thresholds are: you do not know if there will be more than enough participants (so that a threshold will be reached without your participation), or if not enough others will participate (so that a threshold will not be reached when you participate), or if there will be almost enough participants (so that a threshold will be reached just in case you participate). So there is some non-zero probability that your theft will get the group to exactly reach some threshold. The result of exactly reaching a threshold on your act of theft is worse than the result of not reaching a threshold. And so the product of the negative value of this worse result with the small chance of its obtaining under your act is negative. This negative value is assumed to be enough to guarantee that your not stealing has a higher expected utility.

Admittedly, this presentation of Kagan's solution contains several unclarities. Is it really the case that every Kagan-style case has thresholds? And will the math always work out as is suggested in the illustrations involving Chickens_{51} and Beans_{100}? In particular, what about cases in which each individual is *certain* that a threshold will not be breached on his or her act?

In what follows, I brush these unclarities aside.[9] We may assume that the foregoing reasoning is sound; we may assume that EUAC delivers a verdict of moral impermissibility on each act of theft in Beans_{100}. As I demonstrate in the next section, even under this assumption, Kagan's solution will not avoid delivering mismatched verdicts in some Kagan-style cases.

[9] But see Budolfson (2014) and Nefsky (2012a) for criticisms of Kagan's solution that involve pressing against these unclarities.

4.4 Beans$_0$ and the Mismatch Problem for EUAC

In this section, I will demonstrate that the mismatch problem arises for EUAC in connection with some Kagan-style cases, namely certain cases in which each individual has an incentive to perform an act of the repeatedly bad act type, but no one performs the repeatedly bad act type. I will focus on a variant of Beans$_{100}$, and I will show that, for this modified case, the bandit collective acts wrongly under EUAC, but according to Kagan's own reasoning, each individual bandit acts permissibly under EUAC.

Before proceeding, it will be important to highlight three features of Kagan's solution in Beans$_{100}$. First, note that the verdict of wrongdoing on each individual act of theft remains the same no matter how many other bandits will actually participate. Suppose, as it turns out, no bandits other than you will steal. Then even though we are assuming that the bad outcome does not arise, it will still be morally wrong under EUAC for you to steal. That is, Kagan's solution is *participation-level-invariant* in Beans$_{100}$: it doesn't matter how many people in fact will token the repeatedly bad act type, EUAC says that it would be wrong to participate—there is always a chance that enough others will participate to tip a threshold.

Second, note that we may accept Kagan's claims about the existence of thresholds in Beans$_{100}$ without thinking that there's a threshold between every set of neighboring outcomes. As long as there are thresholds *somewhere*, Kagan thinks that EUAC delivers a verdict of moral impermissibility on the individual acts. In particular, Kagan's solution is supposed to work even under the plausible assumption that the villagers would not suffer any noticeable amount of hunger were each to miss out on just one bean. Kagan's solution applies to Beans$_{100}$ even though it lacks an *early threshold*: in Beans$_{100}$, EUAC says that it is wrong to participate despite the fact that exactly one act of participation won't cross the first threshold.

Third, note that Kagan's solution applies even though each bandit has an incentive to participate. In Beans$_{100}$, each bandit gets a tasty snack when he participates in the raid. This comes as a benefit to him. It also comes as a benefit to the bandit collective since, after all, each bandit is a member of the group. The important point is that the total pleasure the bandits enjoy after the raid is outweighed by the total hunger the villagers suffer. So each bandit's theft carries with it a chance that it will tip a threshold resulting in an outcome that's bad overall. For this reason, Kagan's solution is meant to work in Beans$_{100}$ even though each participant has an incentive to token the repeatedly bad act type.

In summary, Kagan's solution is participation-level-invariant, it applies to some cases lacking early thresholds, and it works in some cases in which the individuals have an incentive to participate. We may exploit these three highlighted features of Kagan's solution to identify a mismatch problem for EUAC. Start by considering the following simple variation of the original Beans$_{100}$ case:

> *Beans*$_0$: 100 people in a village are about to eat lunch. Each has a bowl containing 100 beans. 100 bandits sit on a nearby hill. Not a single of these bandits decides to steal beans from the villagers. Each bandit is such that, had he stolen, he would have taken one bean from each of the bowls of the 100 villagers. But no bandit in fact steals, and so each bandit forgoes a tasty snack. Had any one bandit decided to steal, the other 99 still wouldn't have. The loss of one bean does not make a perceptible difference to any villager.

Unlike the original Beans$_{100}$ case, this version of the case involves no bandits participating in the raid. Accordingly, I use the subscript '0' to indicate that zero people steal beans in the relevant group act. In what follows, I will argue that (1) Beans$_0$ is a Kagan-style case, (2) EUAC says that each bandit acts permissibly in Beans$_0$, (3) EUAC says that the group acts wrongly in Beans$_0$. This will establish that EUAC does not avoid the mismatch problem for all Kagan-style cases.

To see that Beans$_0$ is a Kagan-style case, it is sufficient to note that it meets the two essential features identified in section 4.2. It involves a repeatedly bad act type—*stealing one bean from each villager*—and it seems to lack thresholds just as Beans$_{100}$ does. Of course, no bandit actually tokens the repeatedly bad act type in Beans$_0$. But Kagan-style cases don't require that the repeatedly bad act type is actually tokened, only that "[a] certain number of people—perhaps a large number of people—have the *ability* to perform an act of a given kind" (emphasis added).

To see that each individual acts permissibly in Beans$_0$ according to EUAC, we simply reflect on each bandit's alternative. Each bandit could have stolen a bean from each villager. But had a bandit done so, he would have acted wrongly according to Kagan's application of EUAC. As we just saw, Kagan's solution is participation-level-invariant. It would condemn a single act of participation. Each act of theft carries a nonzero probability that the act gets the group to exactly reach some threshold. This makes each act of theft wrong under EUAC, even if no others will in fact participate. But if each bandit's only relevant alternative would have been wrong according to EUAC, then each bandit must act permissibly under EUAC by not stealing in Beans$_0$. Since the theory would condemn each individual act of theft, it must advise each bandit not to steal.

And yet, the group acts wrongly in Beans$_0$ according to EUAC. To see this, we have to examine precisely how EUAC is to deliver verdicts on group acts. In Chapter 3, I discuss some issues that arise when attempting to calculate the expected utility of a group act. One isssue is that if group acts do not have expected utilities, then EUAC encounters the mismatch problem. In Beans$_0$, we may represent the group act as '[NONE STEALS]'. [NONE STEALS] is impermissible under EUAC only if at least one of the group's alternatives has a higher expected utility. And [NONE STEALS] is permissible under EUAC only if none of the group alternatives has a higher expected utility. So if [NONE STEALS] lacks an expected utility altogether, then EUAC does not deliver a verdict on the group act. Assume that some acts are neither permissible nor impermissible. Call such acts *morally undefined.* If group acts do not have expected utilities, then [NONE STEALS] is morally undefined under EUAC. But then Beans$_0$ produces a mismatch problem for EUAC. The theory says one thing about what the bandit collective does (that it's morally undefined behavior) while saying a different thing about each individual does (that he acts permissibly). Thus, if EUAC is to resolve the mismatch problem, we must assume that group acts have expected utilities.

As I suggest in Chapter 3, this can be done if we assume that the bandits in Beans$_0$ have uniform credences and evidence. For the sake of argument, we may stipulate the credences and evidence as follows. Suppose that every bandit is certain that none of the villagers will go hungry if no bandits steal. Suppose that every bandit is also certain that none of the villagers will go hungry if just one of the bandits steals. Suppose that every bandit is certain that no bandits will get a tasty snack if none steals, and suppose that every bandit is certain that exactly one will get a tasty snack if exactly one of them steals. Suppose that each is certain of these things because he has the same conclusive evidence for these facts as all the other bandits have. Then, under these assumptions and stipulations, the group act is wrong in Beans$_0$ according to EUAC. Without getting too deep into the mathematics, a simple line of reasoning will establish that [NONE STEALS] has an alternative with a higher expected utility. We need only compare the expected utility of [NONE STEALS] with the expected utility of [ONE STEALS]. Start by calculating the expected utility of the former. Under the assumption that no bandit steals, it is guaranteed that none of the villagers goes hungry. This is the result of aggregating the bandits' uniform justified certainty that no villager goes hungry if no bandits steal. Thus, there's one possible outcome on [NONE STEALS]: all bandits miss out on a tasty snack, and none of the villagers misses out on a single bean. So the expected utility of [NONE STEALS] is the same as the value of this outcome. For the sake of simplicity, let's assign a value of $-0 + 0$ to the outcome. The -0 comes from the fact that no villagers experience hunger. The $+0$ comes from the fact that no bandits get a tasty snack. So the expected utility of [NONE STEALS] is simply 0.

Consider instead the expected utility of [ONE STEALS]. It is guaranteed under the assumption that exactly one bandit steals that none of the villagers goes hungry. Again, this is the result of aggregating the bandits' uniform justified certainty that no villager goes hungry if exactly one bandit steals. There's one possible outcome on [ONE STEALS]: exactly one bandit gets a tasty snack, and each of the villagers misses out on just one bean. We assume that the loss of one bean does not make a perceptible difference to any villager. In other words, we assume that Beans_0 lacks an early threshold: the point at which the villagers would notice their getting hungry is perhaps at ten thefts. So under the assumption that exactly one bandit steals, it is guaranteed that no threshold has been crossed. The expected utility of [ONE STEALS] is the same as the value of no villagers going hungry while one bandit gets a tasty snack. Let's assign a value of $-0 + 1$ to the outcome. The -0 comes from the fact that no villagers experience hunger. The $+1$ comes from the fact that a single bandit gets a tasty snack. So we may assign a value of 1 as the expected utility of [ONE STEALS].[10]

Here's why the group acts wrongly in Beans_0. The group performs [NONE STEALS], but has [ONE STEALS] as an alternative. Given the previous calculations, the group has an alternative with a higher expected utility. So the group acts wrongly. This is at the same time that each individual acts permissibly according to EUAC. The result is that EUAC delivers mismatched verdicts for a Kagan-style case.

An insight from Chapter 3 explains the possibility of this result. The likelihood of one and the same outcome may be different under a set of acts than it is under any one of the individual acts. Since a group act is a set of acts, the likelihood of an outcome occurring if a group act is performed may be very different from the

[10] In Chapter 6, I suggest that it may be indeterminate how to assign value to the part of the outcome in which each villager misses out on a single bean. For our purposes here, we may ignore this complication. It does not affect the essential point of my argument against EUAC.

likelihood of that same outcome occurring if any one of the component individual acts is performed. In Beans$_0$, a better outcome is guaranteed only if exactly one bandit steals. This is due to the fact that the villagers would be unaffected, but one bandit would be slightly better off. And yet, just as in Beans$_{100}$, it's unlikely that only one bandit will steal. So on the supposition that an individual bandit steals, the likelihood of ending up in a situation in which exactly one bandit steals is very low. The likelihood of tipping a threshold and ending up in a worse situation is very high. So each individual bandit acts permissibly under EUAC by not stealing. But the group acts impermissibly under EUAC because it could have guaranteed a better outcome had it acted differently.

It may be helpful to note that regular old Act Consequentialism *does not* encounter a mismatch problem in Beans$_0$. The bandit collective acts wrongly according to Act Consequentialism since it has an alternative with a better outcome. But each individual acts wrongly according to Act Consequentialism as well. Had he stolen, it is stipulated by the case that none of the others in fact would have joined him. So each bandit has an alternative with a better outcome: it would have been better for him to steal. This reveals that EUAC sometimes gives rise to a mismatch problem that Act Consequentialism can easily handle.

As I have demonstrated, EUAC delivers mismatched verdicts in connection with some Kagan-style cases. Beans$_0$ is a Kagan-style case, each individual bandit acts permissibly according to EUAC, but the group acts impermissibly according to EUAC. We may draw a more general lesson from reflection on this case. Consider any Kagan-style case that lacks early thresholds and in which a tiny incentive accompanies each tokening of the repeatedly bad act type. In any such case, suppose that no one tokens the repeatedly bad act type. Since Kagan's solution is participation-level-invariant, each person acts permissibly. But the group misses out on securing a tiny benefit for one of its members without bringing about a bad outcome. So EUAC finds fault with the group. In any Kagan-style case of this sort, the group has an alternative with a higher expected utility, and mismatched verdicts result.

It will perhaps be helpful to close by explaining why ‘Chickens_0’ is a case of this sort. In Chickens_0—note the subscript—no one purchases a chicken at the store. So each person acts permissibly according to Kagan’s solution. But one purchase, and one purchase only, is not enough to lead the butcher to place another order. And we may naturally imagine that purchasing a chicken carcass brings a small benefit to the family of the purchaser: a tasty dinner. Accordingly, the group act ‘[NO ONE BUYS]’ has an alternative with a higher expected utility: on the group act ‘[ONE PERSON BUYS]’, a slightly better outcome is guaranteed. So the group act is wrong according to EUAC while each individual member acts morally permissibly.

I conclude that EUAC cannot resolve all Kagan-style cases. In section 4.2, I identified the essential conditions of such cases. And in section 4.3, I explained why Kagan thinks that EUAC doesn’t deliver mismatched verdicts in Kagan-style cases. But as I have demonstrated, there are Kagan-style cases in which (a) each individual has an incentive to perform an act of the repeatedly bad act type, but (b) no one performs the repeatedly bad act type. In some cases of this sort, such as Beans_0 and Chickens_0, EUAC delivers mismatched verdicts. The mismatch problem persists.

CHAPTER 5

EVALUATING THE UNCOOPERATIVENESS SOLUTION

5.1 Introduction

As we have seen in previous chapters, there are several cases in which a group acts in such a way as to bring about a bad outcome, and though the group could have done something better, no individual member of the group could have done any better than he or she actually did. In connection with such cases, consequentialism delivers mismatched verdicts; the theory paradoxically condemns what the group does without being able to say anything against what the individuals have done.

It is commonly assumed that an essential feature of such cases is the failure of individuals to cooperate. And, recognizing this, several consequentialists have suggested that uncooperativeness constitutes a moral failing; they have suggested that consequentialism can be revised so that it condemns the acts that these uncooperative individuals perform.[1] Unfortunately, as I explain in this chapter, this strategy does not work for all versions of the problem case. In section 5.2, I present the uncooperativeness solution. Then, in sections 5.3 and 5.4, I present and discuss a slightly modified version of the problem case to illustrate why the appeal to uncooperativeness fails as a general solution to the problem.

[1] See, for example, Zimmerman (1996), Kierland (2006), Jackson (1997), and Pinkert (2015).

5.2 The Uncooperativeness Solution

It will be helpful to center our discussion on a particular case of the sort under consideration.

> *Two Voters*: Dr. Mediocre and Professor Beneficent are the two candidates up for public election, and Beneficent is by far the superior candidate. It's best if Beneficent wins, second-best if Mediocre wins, and worst if the vote results in a tie, in which case no one wins. Vincent and Virgil are the only two voters in the election. Vincent is determined to see Mediocre elected, and so he votes for Mediocre. Furthermore, Vincent is uncooperative: he would vote for Mediocre even were Virgil to vote for Beneficent. Similarly, Virgil is determined to see Mediocre elected, and he's uncooperative as well; he too votes for Mediocre, and he would do so even were Vincent to vote for Beneficent. Accordingly, the inferior candidate, Mediocre, receives two votes and wins the election.

In Two Voters, Vincent and Virgil could have together elected the better candidate. But neither of the individuals would have done his part in the best pattern of collective behavior had the other done his part in it. Had either one of them cast his vote for the better candidate, it would have resulted in the worst possible outcome, a split vote. That's because the other still would have voted for Mediocre. So Two Voters is a problem case of the sort under consideration: a group of people acts together to bring about a bad outcome, and though the group could have done something much better, no individual member of the group could have done any better.

In Two Voters, neither Vincent nor Virgil has an alternative with a better outcome. Had Vincent voted Beneficent, Virgil still would have voted Mediocre, and the result would have been worse. Had Virgil voted Beneficent, Vincent still would have voted Mediocre, and the result would have been worse. Thus, given what the other would do, casting a vote for Mediocre is the best that either can do. So each individual act is morally permissible according to Act Consequentialism. On the other hand, the two voters together could have both voted for Beneficent, which would have resulted in a better outcome. Thus, the group has acted wrongly according to Act Consequentialism. Paradoxically, two individual rights make a collective wrong in Two Voters.

Act Consequentialism delivers mismatched verdicts between how Vincent and Virgil act as individuals and how their group acts.

Cases like Two Voters have lead some philosophers to conclude that Act Consequentialism is indeterminate: the theory sometimes fails to direct individuals toward the best outcome that they could collectively bring about.[2] Toward clarifying this idea, notice that there are two possible configurations of individual acts under which each of Vincent and Virgil acts permissibly under Act Consequentialism. The first is the actual configuration under which each votes for Mediocre. The second is the configuration under which each votes for Beneficent. Since either configuration has both individuals satisfying Act Consequentialism, the theory doesn't direct the individuals away from the suboptimal collective pattern of behavior. I assume that indeterminacy is a problem because it means that in some cases Act Consequentialism delivers mismatched verdicts between a group-level evaluation and the individual-level evaluations. In cases like Two Voters, the group act is morally impermissible, but each individual act is morally permissible.[3]

[2] Derek Parfit, for example, in section 21 of Parfit (1984).

[3] I thank an anonymous referee for encouraging me to elaborate on this point. I discuss some of the other approaches to Two Voter-type cases in Chapter 1. Note also that there's a different—but equivalent—way of presenting the problem. Many philosophers believe that morality has a social function: if every person in some group does all that is required of him or her by morality, then the result will be the morally best outcome attainable by the group. Apparently, Baier (1958) and Castaneda (1974) are proponents of this idea. It's natural to think that this so-called 'principle of moral harmony' is a basic requirement on all moral theories. See Feldman (1980), Kierland (2006), Pinkert (2015), and Portmore (2016) for discussion. But Two Voters reveals that Act Consequentialism violates the principle of moral harmony. Vincent and Virgil together have the option of producing a world in which Beneficent wins the election. From the perspective of consequentialism, this is the morally best world attainable by the group. But Vincent and Virgil together are not guaranteed to actualize this world when each of Vincent and Virgil acts permissibly according to Act Consequentialism.

It's important to see that Act Consequentialism violates the principle of moral harmony just in case the group could have done better though no individual member of the group could have done any better. But in precisely these cases, Act Consequentialism will deliver mismatched verdicts between how the individuals act and how the group acts. So it will suffice for our purposes here to assume that the problem of mismatched verdicts that besets Act Consequentialism in connection with Two Voters is equivalent to the problem concerning Act Consequentialism's violation of the principle of moral harmony.

It is perhaps well-known to those familiar with the mismatch problem that Two Voters and structurally similar cases have resisted a tidy and satisfactory solution. Starting with Donald Regan's and Derek Parfit's influential discussions, a number of philosophers have tried to deal with the problem by adding some additional elements to their consequentialist theories.[4] Regan suggested that an individual may act wrongly if he or she fails to engage in a specific procedure meant to identify potential cooperators and pursue optimal outcomes with them. Parfit suggested that an individual may act wrongly because he or she belongs to a group that acts wrongly. Other philosophers have suggested adopting some expected utility formulation of Act Consequentialism.[5] Each attempted solution involves abandoning plain old Act Consequentialism. A modified version of the theory takes its place. Under the modified theory, each of Vincent and Virgil acts wrongly. I have discussed deficiencies with some of these approaches in previous chapters.[6] This chapter is concerned with a (relatively recent) form of attempted solution that has not received much critical attention.[7]

The attempted solution involves condemning each of Vincent and Virgil for being uncooperative. In his 1996 book, Michael Zimmerman faults each of Vincent and Virgil for being 'intransigent'—that is, having the disposition to act in one and the same way no matter how others will act. Each of Vincent and Virgil votes for Mediocre with intransigence; each would vote for Mediocre even were the other to perform the act that would be required to elect Beneficent. Thus, according to Zimmerman, the solution "is in outline simply (and unsurprisingly) this: one's moral obligation is to

[4] The influential discussions are found in Regan (1980) and Parfit (1984).

[5] See Chapters 3 and 4.

[6] See also Pinkert (2015) for an excellent survey of some of the problems that beset the approaches outlined in this paragraph.

[7] As far as I am aware, criticism of the approach is limited to Forcehimes and Semrau (2015). I advance a different line of criticism here.

do the best one can while avoiding intransigence; that is (to coin a term), one ought *transigently* to do the best one can." Doing the best one can must be accompanied by the adoption of a certain *attitude*."[8] In a 2006 article, Brian Kierland also faults each of Vincent and Virgil for failing to possess the attitude in question: "Lacking such an attitude, there will be circumstances in which an agent will not be disposed to maximally promote deontic value; these will be circumstances in which the agent has the opportunity for cooperating with others in the promotion of deontic value. . . So Vincent and Virgil are each subject to negative agent evaluation in virtue of each failing to possess an attitude of cooperativeness."[9]

Similar suggestions have been made in connection with structurally equivalent cases. In a 1997 essay, Frank Jackson offers the following suggestion for how to fault two intransigent sharp-shooters X and Y who overdetermine my death: "although consequentialists should say that X and Y do nothing wrong, they can and should say that X and Y are people of bad character in that in a certain case they would have done wrong."[10] And most recently (2015), Felix Pinkert has defended the same sort of approach in connection with a case in which two factory owners, Ann and Ben, overdetermine the pollution of a river. Pinkert notes that "by being agents who would pollute even if the other agent produced cleanly, Ann and Ben make it impossible for each other to achieve better outcomes by acting differently. A moral principle that condemns such uncooperativeness then finds fault in cases where Act Consequentialism cannot do so."[11]

Each of these writers seems to be proposing essentially the same solution to the mismatch problem for Act Consequentialism. A clause may be added to plain old

[8] (Zimmerman, 1996, 263)

[9] (Kierland, 2006, 400-401)

[10] (Jackson, 1997, 50)

[11] (Pinkert, 2015, 981-982)

Act Consequentialism so that an individual faces negative moral evaluation in some choice situation if he or she is disposed to fail to do his or her part in a better group act even if others were doing their parts in it. In Two Voters and the related cases, the individuals are uncooperative in the specified way, so each faces negative moral evaluation after all.

Some clarifications are in order. First, it is important to note that ‘uncooperative’ here is a somewhat technical notion. There are two ways in which the idea has been presented in the literature, only one of which I find plausible. Some writers try to capture the notion like this: an agent is uncooperative just in case the agent would bring about a suboptimal outcome were others to act differently.[12] But this characterization is too strong—it makes it too difficult to be cooperative. There are perhaps infinitely many ways in which others might act in some situation. For you to be cooperative, you would need to be disposed to act optimally in response to all of these possible configurations of acts. This is too stringent a requirement. Furthermore, it’s a much stronger requirement than is necessary to identify a morally objectionable character trait in each of Vincent and Virgil.

Under a more plausible characterization—and the one I adopt here—an agent is uncooperative in some choice situation just in case the agent is disposed to not play his or her part in the optimal collective pattern of behavior in that situation. We first identify the optimal group act in some situation, and we then consider what each individual would have to do to perform his or her part in the group act. An individual is uncooperative just in case the agent would fail to do his or her part in the optimal collective pattern of behavior were all others to play their parts in it. This characterization makes the requirement to be cooperative more straightforward:

[12] This, for example, is how the idea is officially stated in (Pinkert, 2015, 982): “for all possible combinations of the actions of other agents, if that combination were instantiated, [a cooperative agent] would act optimally in these circumstances”.

simply be such that you would help to bring about the best collective outcome were it possible for you to do so.

To see precisely why Vincent is uncooperative in Two Voters, consider the counterfactual situation in which Virgil casts his vote for Beneficent. In this situation, Virgil does his part in the optimal group act. But as is stipulated in Two Voters, Vincent still votes for Mediocre. Accordingly, though Vincent does his best in the actual world, he fails to do his best in the relevant nearby worlds. Since Virgil similarly brings about a suboptimal outcome in those counterfactual situations in which Vincent votes Beneficent, Virgil is uncooperative as well.

It is also important to be clear about the nature of the negative moral evaluation that would accompany an agent's being uncooperative under the proposed solution. Apparently, only Zimmerman makes it morally obligatory to act with cooperativeness. Kierland, Jackson, and Pinkert on the other hand are talking about agent evaluation as opposed to act evaluation. They make cooperativeness a requirement of good character, not of morally permissible behavior.[13] Accordingly, the latter three commentators pursue a 'solution' that does not address the problem with which we are presently concerned. The mismatch problem arises in connection with Two Voters because Act Consequentialism says that the group *acts wrongly*, but it cannot say that either of Vincent or Virgil *acts wrongly*. Thus, a satisfactory solution to the mismatch problem must be in terms of act evaluation: Act Consequentialism must be modified in such a way that at least one of Vincent and Virgil is said to act

[13] Kierland makes this a central feature of his discussion. He believes that the focus on agent evaluation distinguishes his approach from that found in Zimmerman (1996). Jackson explicitly states of the overdetermined sharp-shooters that "the case displays immorality, but immorality of character rather than immorality of action" (Jackson, 1997, 50). Pinkert also makes it clear that he is focused on agent evaluation: "[Ann's and Ben's] uncooperativeness shows that there is something wrong with them as moral agents: They do not satisfy the demands of Act Consequentialism modally robustly, and this shows that they do not appropriately and effectively care about the livelihoods of the workers and fishermen. . . . Ann and Ben thus each individually show a morally problematic character trait" (Pinkert, 2015, 987).

wrongly. Saddling the voters with bad character doesn't go far enough; it still leaves an unsatisfying mismatch between wrong group behavior and permissible individual behavior.

However, for the purpose of giving a general criticism of the uncooperativeness approach, we may blur the distinction between act evaluation and agent evaluation in what follows. Let *Disposition Consequentialism* be the view that an agent has a good disposition just in case the agent would do his or her part in the optimal collective pattern of behavior were all others to do their parts in it. According to *Act Consequentialism + Disposition Consequentialism* (ACD), an agent escapes negative moral evaluation in some situation just in case the agent satisfies both Act Consequentialism and Disposition Consequentialism. According to Zimmerman, we may understand the negative moral evaluation squarely in terms of act evaluation. The Zimmerman-inspired approach would resolve the mismatch problem. According to Kierland, Jackson, and Pinkert, on the other hand, we are to understand the negative moral evaluation as more complex—as encompassing act evaluation at the group level and agent evaluation at the level of the individuals. In the next section, I demonstrate that, however we understand the negative moral evaluation, there is a simple variation of Two Voters under which neither Vincent nor Virgil is subject to it.

5.3 The Mismatch Problem for ACD

The mismatch problem for Act Consequentialism arises in Two Voters because each individual act lacks an alternative with a better outcome. It is commonly assumed that this happens only when the individuals are mutually uncooperative. But we need not make this assumption. To see this, consider a variation of the Two Voters case:

> *Two Voters+*: Mediocre and Beneficent are the two candidates up for public election, and Beneficent is by far the superior candidate. Vincent and Virgil are the only two voters in the election, and it takes two votes

> for the same candidate to get that candidate elected. Each voter will cast a vote for Mediocre. A mechanical defect in the voting machines means that the options available to each voter are restricted based upon how the other will in fact vote: the machines accept either two Mediocre votes or two Beneficent votes, but no split votes. Since Virgil will in fact cast a vote for Mediocre, it turns out that Vincent *cannot* cast a vote for Beneficent. Similarly, since Vincent will in fact cast a vote for Mediocre, it turns out that Virgil *cannot* cast a vote for Beneficent either. Each would have cast a vote for Beneficent were the other to have cast a vote for Beneficent. But given how the other will in fact vote, neither can cast a vote for Beneficent.

We may represent the situation in Two Voters+ by Table 5.1. We are not to think

	Vincent votes Benef.	**Vincent votes Med.**
Virgil votes Benef.	*best*	IMPOSSIBLE
Virgil votes Med.	IMPOSSIBLE	***worst***

Table 5.1. The possible outcomes in Two Voters+

that some third actor has sabotaged the machines. This would potentially introduce a morally wrong individual act into the case. Instead, the machines come to have the defect through some natural cause. Accordingly, the only relevant acts and outcomes are those represented in Table 5.1. Notice that there's no possibility of a split vote in Two Voters+. What Vincent actually does is represented in bold in the very top row. Given that Virgil actually votes for Mediocre, Vincent could not have voted for the other candidate. Similarly, given that Vincent actually votes Mediocre, Virgil could not have voted otherwise. There are nonetheless two possible outcomes. In those worlds in which Vincent and Virgil both will vote for Beneficent, the *best* candidate wins. In those worlds in which both will vote for Mediocre—and this includes the actual world—the *worst* candidate wins. Each of these two outcomes is accessible to the group; the group could have brought about either of them. The group actually brings about the *worst* outcome in the bottom right box (in bold), but it could have brought about the better outcome in the top left box had both Vincent and Virgil voted differently.

It may be helpful to think about the situation in Two Voters+ in terms of two actors mutually restricting each other's options. By his actually voting for Mediocre, Vincent makes it so that Virgil is unable to vote for a different candidate. And Virgil returns the favor. By his actually voting for Mediocre, Virgil makes it so that Vincent is unable to vote for a different candidate. Of course, each individual voter has perhaps thousands of alternatives in Two Voters+. Perhaps Vincent votes for Mediocre with his right hand. He could instead vote for Mediocre with his left hand. He could vote with either a frown on his face or with a beaming smile. But the crucial point is that none of his alternatives incorporates his voting for Beneficent. Similarly, none of Virgil's alternatives incorporates his voting for Beneficent either.

Notice that Two Voters+ gives rise to the problem of mismatched verdicts for Act Consequentialism. The group act, casting two votes for Mediocre, has an alternative with a better outcome. Accordingly, the group act is morally impermissible. But no individual has an alternative with a better outcome: in fact, no individual has an alternative with an outcome that differs in an axiologically relevant way from the outcome that actually results—each is forced to cast a vote for Mediocre, in tandem with the other. Thus, each individual act is morally permissible in Two Voters+. Two rights make a wrong under Act Consequentialism. So Two Voters+ is a version of the problem case that we should expect ACD to resolve.

And yet each individual voter escapes negative moral evaluation under ACD. The optimal group act has both voters casting votes for Beneficent. Each voter would do his part in this act were the other to do his part in it. Consider the counterfactual situation in which Vincent casts a vote for Beneficent. In that situation, as stipulated by the case, Virgil also votes for Beneficent; since the machines are rigged together, there are no worlds containing a split vote. (Mutatis mutandis for how Vincent would act in the relevant counterfactual situation.) So each voter is cooperative in the required sense. Thus, ACD cannot resolve the mismatched verdicts that Act Consequentialism delivers in Two Voters+.

5.4 What's Going on in Two Voters+?

Advocates of the uncooperativeness solution will want to resist my characterization of Two Voters+. In particular, it may be objected that the case rests on shaky metaphysical assumptions about groups, acts, and alternatives. The case requires that Vincent and Virgil together have an alternative that neither can perform his part in. This may strike some as implausible, but I will elaborate on a certain temporal aspect of the case that reveals how such a relationship between the individual and group alternatives is possible. After exploring this temporal dimension of Two Voters+, I will discuss an objection: if the group can cast two votes for Beneficent, then it may seem that there must be *some point* in time during the voting process during which one of the voters is guilty of being uncooperative. I will explain why this insight does not help ACD resolve the mismatch problem as described in Two Voters+.

Before proceeding, note that almost all presentations of the mismatch problem involving two individuals have them acting independently of each other: each could act in one of two significantly different ways, regardless of how the other would act.[14] This generates an array of four possible group acts. But the stipulation that the actions available to the individuals are entirely independent from each other is an unfortunate artifact of traditional presentations of the problem, and it is not essential. The essential feature is that a group pursues a suboptimal outcome, but each individual member could not have performed an act that would have resulted in a better

[14] This feature of such cases apparently goes at least as far back as Allan Gibbard's presentation of the problem in Gibbard (1965). Gibbard's example involves two actors who are "placed in separate isolation booths, so that the actions of one can have no influence at all on the actions of the other."(Gibbard, 1965, 214)

outcome than the one he or she in fact brings about. We may capture this essential feature with situations involving an array of two group acts, such as in Two Voters+. In these situations, the individuals' alternatives are dependent on each other; the individuals must perform their acts in tandem along one of two possible courses.

To begin laying down the metaphysical foundation for such situations, it may help to make a few observations. First, notice that facts about what will in fact happen often serve to limit the things a person can do. If the power will go out, then you cannot keep the lights on. If your Internet connection will stop working, then you cannot finish sending your emails. And, in particular, if the voting machine will not accept a vote for Beneficent, then Virgil cannot cast a vote for Beneficent.

Second, notice that the actions of others may determine such limiting future events. We may imagine that the power will go out because the neighbor will blow a fuse. We may imagine that the Internet connection will stop working because the cable company will not provide a quality Internet connection. And as is stipulated in Two Voters+, we may imagine that the voting machine will not accept a vote for Beneficent because Vincent will cast a vote for Mediocre.

Third, notice that limiting future events may be overdetermined. We may imagine that you will blow a fuse too. Or that two cable companies each will fail to provide you with a quality Internet connection. Or, in particular, that Vincent and Virgil each will vote in such a way that the voting machine will not accept votes for Beneficent. It is in this way in Two Voters+ that each of Vincent and Virgil restricts the other's options.

We may elaborate on this further by having the individual acts in Two Voters+ take place over two stages in time. Let's imagine that the Beneficent buttons work like this. First, each voter must depress his Beneficent button halfway. If both buttons are pressed in this way, then a mechanism in the machine unlocks, allowing the two buttons to be depressed completely. Only when both buttons are completely depressed do the machines register votes for Beneficent. Suppose that the Mediocre buttons work in a similar manner. This is why split votes are impossible. If Vincent depresses his Mediocre button halfway while Virgil depresses his Beneficent button halfway, there is no way for either voter to completely depress his button. Accordingly, the machines cannot register one vote for Mediocre and one vote for Beneficent.

Intuitively, the group can completely depress both Beneficent buttons, bringing the buttons first down to the halfway level, and then from halfway down to flush with the console. But suppose that neither voter is going to press his Beneficent button halfway. If we attend carefully to the temporal elements of the case, we see that the group has an alternative such that no individual member of the group has as an alternative his part in it. The group's alternative spans two stages. During the 'preparation stage', which would take place at t_1, both members of the group depress their Beneficent buttons halfway. During the 'fruition stage', which would take place at t_2, both members depress their Beneficent buttons from halfway down to flush with the console. We may describe the whole thing, which would take place from t_1 to t_2, as the group's 'casting two votes for Beneficent'.

Since neither voter is going to participate in the preparation stage of the group's casting two votes for Beneficent, neither Vincent nor Virgil has as an alternative his part in the fruition stage of the group's alternative. To fill this in, imagine that each voter is quick to press his Mediocre button if he finds his Beneficent button stuck at t_1, and suppose that this is a reflex outside of his control. Suppose that each would depress his Mediocre button halfway at t_1 even if the other were to depress his Beneficent button halfway at t_1. Then if just one of them depresses his Beneficent button halfway at t_1, it will turn out that each will ultimately cast a vote for Mediocre. To cast a vote for Beneficent, Vincent would have to perform an act like this: at t_1, he presses his Beneficent button halfway, and at t_2 he continues depressing the button until it is flush with the console. Since Virgil is not going to press his Beneficent button halfway at t_1, Vincent would not be able to continue depressing his Beneficent button at t_2. That is, Vincent does not have an alternative in which he casts a vote for Beneficent. Neither does Virgil. Thus, Two Voters+ has the following profile of individual and group alternatives: neither Vincent nor Virgil has casting a vote for Beneficent as an alternative, but the group has casting two votes for Beneficent as an alternative.

Under these conditions, notice that in all of the nearest worlds in which either casts a vote for Beneficent, the other does too. Remember that the performance of casting a vote for Beneficent comprises two stages: at t_1 Virgil presses his Beneficent button halfway, and at t_2 he continues depressing the button until it is flush with the console. In all the nearby worlds in which Virgil does this, Vincent presses his Mediocre button halfway at t_1, but then realizes that the button is locked. Since it is stipulated that Virgil continues depressing his Beneficent button at t_2 in all these worlds, it must be that in the time between t_1 and t_2 Vincent eventually switches to pressing his Beneficent button to the halfway point, thereby unlocking both buttons. We are evaluating the counterfactual by holding fixed that Virgil casts a vote for Beneficent. So in all nearby worlds in which Virgil casts a vote for Beneficent, so too does Vincent. This counterfactual relationship among the group's alternatives and the individual alternatives is what stymies ACD from delivering the desired condemnation of each individual. Each voter would end up participating in the optimal group act were it possible for him to do so.

In particular, by reflecting on the temporal features of the case, we see that the following two claims are consistent:

> (a) Each of the two voters has no choice but to cast a vote for Mediocre; each of his alternatives incorporates his voting for Mediocre.
> (b) That being said, were either of the two voters to cast a vote for Beneficent, the other would cast a vote for Beneficent as well.

It is in virtue of (a) and (b) that neither Vincent nor Virgil is subject to negative moral evaluation under ACD in Two Voters+.

But now, by reflecting on this newly revealed temporal aspect of Two Voters+, an advocate of ACD may suggest at least one way in which the theory avoids giving mismatched verdicts: each voter is uncooperative during the preparation stage of the group's casting two votes for Mediocre. Each voter depresses his Mediocre button halfway at t_1, and he would do so even were the other to depress his Beneficent button halfway at t_1. But this means that each voter actively constrains the other's voting options, which results in a suboptimal group act being performed. So if we focus on the temporal slice of Two Voters+ that takes place just at t_1, we see that ACD does not encounter a mismatch problem there.[15]

I believe this insight does not resolve the mismatch problem for Two Voters+. To see this, it will be important to identify precisely *when* each individual is uncooperative. If we clarify exactly how ACD avoids the problem of mismatched verdicts at t_1, we see why it doesn't follow that ACD steers clear of the problem in Two Voters+.

Consider the possible configurations of individual behavior at t_1, which we may represent in Table 5.2. For each combination of halfway button depressing at t_1, we see what the group would end up doing during the time interval from t_1 through t_2. Vincent and Virgil actually perform the bolded pieces of behavior at t_1. The result

	Vincent depresses his B button halfway at t_1	**Vincent depresses his M button halfway at t_1**
Virgil depresses his B button halfway at t_1	the group casts two votes for Beneficent	the group casts two votes for Mediocre
Virgil depresses his M button halfway at t_1	the group casts two votes for Mediocre	*the group casts two votes for Mediocre*

Table 5.2. Configurations of individual behavior during t_1 of Two Voters+

[15] I want to thank an anonymous referee for raising this important objection.

is that their group in fact proceeds to perform the italicized act in the bottom right box. Each voter could have depressed his Beneficent button halfway at t_1 instead. But had he done so, the other still would have depressed his Mediocre button halfway at t_1. Accordingly, neither can get the group to perform the optimal group act in the top left box.

Thus, there is a version of the mismatch problem for Act Consequentialism that's represented in Table 5.2. We may see the little piece of group behavior at t_1 as morally wrong according to Act Consequentialism. It brings about the group act in the bottom right box (which results in a suboptimal outcome), though it could have brought about the group act in the top left box (which results in the best outcome). But we may see each little bit of individual behavior at t_1 as permissible according to Act Consequentialism. Had Vincent behaved differently at t_1, Virgil still would have depressed his Mediocre button halfway at t_1. The result would have been the same group act (resulting in the same suboptimal outcome) as in the actual world. The same goes for Virgil.

ACD resolves this Table 5.2 version of the mismatch problem. The optimal piece of group behavior at t_1 is both voters depressing their Beneficent buttons halfway at t_1. Had Virgil done his part in this piece of group behavior, Vincent still would have depressed his Mediocre button halfway at t_1. So Vincent is uncooperative. Similarly, Virgil is uncooperative. By focusing only on the preparation stage of Two Voters+, we see that each voter behaves in such a way as to be subject to negative moral evaluation under ACD at t_1.

Notice, however, that we may distinguish between Two Voters+ and the case just described. Two Voters+ centers on individual acts performed in the time spanning t_1 and t_2: each voter's casting a vote for Mediocre begins at t_1 with his pressing his Mediocre button halfway and ends at t_2 with his completely depressing his Mediocre button. Each component individual act is performed over two stages in Two Voters+, as opposed to the smaller temporal pieces of individual behavior represented in the choice situation in Table 5.2. Thus, the case in Table 5.2 is merely a temporal slice of Two Voters+. And simply because ACD avoids mismatched verdicts for a temporal slice of a case, it doesn't follow that ACD steers clear of a mismatch for the parent case.

The key point is that an individual may be cooperative with respect to a group act and yet uncooperative with respect to a smaller temporal chunk of the group act. In Two Voters+, the optimal group act spans a certain length of time. During the performance of that temporally extended group act, each voter would do his part were the other to do his part. So each voter is cooperative in Two Voters+. On the other hand, each voter is uncooperative with respect to the situation in Table 5.2 in which the group act in Two Voters+ has not yet been performed.

This shows that for some mismatch problems that arise for ACD, the theory does not encounter a mismatch at an earlier time. But this does not mean that ACD works as a general solution to the mismatch problem for Act Consequentialism. We want a modification to Act Consequentialism that delivers concordant verdicts in Two Voters+. If ACD delivers concordant verdicts during a small temporal piece of the Two Voters+ case, then this gives us only a piece of the solution that we are after.

CHAPTER 6

HARMLESS TORTURERS, ACCUMULATION, AND HOW TO RESOLVE THE MISMATCH PROBLEM FOR CLIMATE CHANGE

6.1 Introduction

In many familiar cases, a group of people acts together to bring about a bad outcome, and though the group could have done something much better, no individual member of the group could have done any better. Act Consequentialism encounters a challenging problem in connection with such cases. The theory delivers mismatched verdicts; the theory condemns what the group does while being unable to condemn any of the individual acts.

This chapter is concerned with a particularly tricky sort of case of this form. In a case of accumulation, a bad outcome arises under the collective behavior of a large group of people, but each individual member of the group produces an apparently negligible effect: each individual actor nudges things closer to ruin, but without its being true that his or her act has made things worse. Under increasing levels of participation, disvalue gradually accumulates but without any sharp boundaries between the value of the outcome under one level of participation and the value of the outcome under the next.

For an example of such a case, consider Derek Parfit's *Harmless Torturers* (HT).[1] A victim is strapped to a torture machine with one hundred and one settings. If

[1] The case appears in (Parfit, 1984, 80). In Parfit's original example, there are one thousand victims, and one thousand torturers. I have modified the example slightly, but not in ways that would affect the fundamental points I wish to make about the case.

the machine remains at the zeroth setting, zero volts of electric current will course through the victim's body; if the machine is brought up to the one hundreth setting, it will be one hundred volts. An incremental change of one volt is too tiny for the victim to perceive.[2] For all n, the victim's experience under the nth setting of the machine feels the same as what the victim would experience under the $n-1$st setting. The situation is similar to a series of color patches going from red to orange with neighboring patches having the same appearance: though the final state of the machine feels terrible and the initial state feels entirely painless, neighboring settings of the torture machine feel the same way to the victim. One hundred eager torturers stand facing one hundred switches. However many switches are flipped, the machine will be brought up to the corresponding setting. No torturer is able to influence any of the others, and all at the same time the one hundred torturers flip their switches. As a result, the victim suffers an intense episode of pain; he is made to experience the sensory feeling associated with one hundred volts of current coursing through his body. But it is not true of any individual act that it makes things worse than they would have been were that act not performed. Why not? Because we are to assume that pain is the only bad-making feature in this situation, and the victim would have felt the same had one fewer torturer flipped his switch.

HT gives rise to a version of the mismatch problem for Act Consequentialism. According to *Act Consequentialism*, an act is morally permissible just in case there's no alternative with a better outcome. The group could have done better, so Act Consequentialism condemns the act performed by the group. But each individual act appears to be permissible under Act Consequentialism since it is not true of any torturer that he has an alternative with a better outcome. In this chapter, I offer a

[2] If this is too difficult to imagine, then increase the number of settings. According to pretheoretic common sense, human perceptual abilities are limited; there is some discriminability threshold for electric current. Imagine that the difference in voltage between adjacent settings of the machine falls below the discriminability threshold of the victim.

solution to this problem. I first identify and explain the conditions under which the mismatch problem arises in connection with HT. I diagnose the peculiar axiological feature that separates HT from related versions of the mismatch problem: unlike other standard versions of the problem, HT involves indeterminate value comparisons. As a result, the mismatch problem in connection with HT involves a verdict of moral wrongdoing at the group level and yet indeterminate individual-level verdicts. I then show that by making a small modification to Act Consequentialism, we may exploit the axiological indeterminacy in such a way as to avoid mismatched verdicts. I close by extending the solution to cover a variation of the mismatch problem that arises in connection with anthropogenic climate change.

6.2 Some Preliminary Clarifications

Before proceeding, it is important to make a few clarifications about how I am approaching HT. The case is especially complex and puzzling; it embodies a tricky axiological situation. To see this, imagine what would have happened had the torturers flipped their switches sequentially. The bad outcome would have arisen gradually through the performances of several individual acts, but each individual act along the way would have had seemingly trivial effects. Each additional flipped switch goes unnoticed by the victim: he always feels the same as he had previously felt the moment before. So supposing the torturers had flipped their switches one after the other, it's not clear how any individual act within the resulting series would have made things worse than before the act was performed. All the while, the individuals involved would have inched toward the production of a disaster. The insignificant effects would have 'added up'; a terrible situation would have slowly developed. Eventually, the result of the entire set of individual acts would have been bad—as is the outcome when the switches are all flipped simultaneously. Just as a man goes from hirsute to bald gradually and apparently without a sharp boundary, the set of possible outcomes accessible to the torturer collective gets worse only gradually—the disvalue accumulates, but without a sharp boundary between any neighboring outcomes.[3]

This axiological situation may strike some as implausible for the following reason. The end result is bad, and the initial state is not bad at all: certainly there must be some point in between when the situation is made worse. How could it be that no individual act makes things worse, no matter how many other acts are performed?

In order to begin to see how this may work, it will be helpful to address two preliminary issues with HT. One issue concerns how it could be that each setting of the machine feels the same as the one previous, while the last setting of the machine feels much different from the initial setting. Another issue concerns how we might give a more precise formulation of the lack of sharp boundaries among the values of neighboring outcomes in HT.

First, let's introduce some terminology. In HT, the group of torturers could have brought the machine to any of the one hundred and one settings. To capture this idea, let a *group act* be any set of individual acts. Under this minimal conception, a group has an alternative for any compossible combination of individual alternatives. The group of torturers has many alternatives. In the actual world, the group act is composed of every torturer's flipping his switch. The relevant outcome of this group act is V_{100}: the state of affairs of the victim's suffering an episode of pain under the one hundreth setting of the machine. Were exactly one of the torturers to abstain from flipping his switch, the resulting group act would have brought about V_{99}: the state of affairs of the victim's suffering an episode of pain under the ninety-ninth setting.[4] The group also could have brought about V_0: the state of affairs of the

[3] There are several discussions of the mismatch problem for Act Consequentialism in which philosophers apparently mean to characterize axiological situations of this sort. See (Glover, 1975, 173), the discussion of Kagan-style cases in Chapter 4, (Andreou, 2014, 210), (Rabinowicz, 1988, 40), and most recently Nefsky (2016) and Barnett (2017).

[4] A certain complication arises. There are one hundred torturers. Accordingly, there are one hundred different ways for one fewer torturer to flip his switch. It could have been that torturer

victim's suffering no pain whatsoever under the zeroth setting of the machine. And the group could have brought about every outcome in between.

In HT, the episode of pain in V_{100} feels the same as the episode of pain in V_{99}, the episode of pain in V_{99} feels the same as the episode of pain in V_{98}, and so on. But this may seem to conflict with the description of the case. For the episode of pain in V_0 does not feel the same as the episode of pain in V_{100}, and perhaps it will be thought that 'feels the same as' is a transitive relation.

To see why the description of the case is consistent, we must disambiguate between two relations: 'being phenomenologically identical to' and 'being indistinguishable from'. Each of these could stand in for the notion of feeling the same. The first is a transitive relation. Suppose that the episode of pain in V_{100} has exactly the same set of phenomenological properties as the episode of pain in V_{99}. And suppose that V_{98} contains exactly the same phenomenological properties as are contained in V_{99}. Then it must follow that the episode of pain in V_{100} has exactly the same set of phenomenological properties as the episode of pain in V_{98}. This reasoning will lead us to conclude that V_{100} is phenomenologically identical to V_0. To avoid this contradiction, when it is stipulated that the episode of pain in V_{100} feels the same as the episode of pain in V_{99}, this is not to be understood in terms of phenomenological identity.

On the other hand, 'being indistinguishable from' is not a transitive relation. Perhaps you have had experiences to confirm this: you cannot perceive a difference between color patch A and color patch B, nor can you perceive a difference between color patch B and color patch C, but you can perceive the slightest difference between

number one doesn't flip his switch. It could instead have been that torturer number two doesn't flip. And so on. So it's not really accurate to say that V_{99} is *the* state of affairs of the victim's suffering an episode of pain involving his experiencing ninety-nine volts of current. Rather, it is one such state of affairs out of one hundred suitably related states of affairs. For our purposes here, we may ignore this complication.

color patch A and color patch C. I am imagining that the states of the torture machine are like this for the victim. Each is indistinguishable from the one previous. Since 'being indistinguishable from' is not a transitive relation, there is no contradiction in this way of understanding the case.[5]

The next issue concerns how to understand the lack of sharp boundaries among the outcomes in HT. I assume that the series of outcomes from V_0 through V_{100} forms a value continuum: any difference in value that obtains between one pair of adjacent outcomes will obtain between all pairs of adjacent outcomes. Whatever difference there is in the victim's pain between V_0 and V_1, a difference of the same sort would be found between any two adjacent outcomes. Pain is the only axiologically relevant feature in HT. So whatever basis would be cited for making a determinate value comparison between V_0 and V_1, this same basis would be found between all other pairs of adjacent outcomes. There cannot be a difference in value relation without a corresponding difference in the axiological basis for it.[6] We may state this feature of HT as follows:

> *Value Continuity*: If some comparative value claim is true of V_n and V_{n+1}, it is true of all pairs (V_i, V_{i+1}) for $0 \leq i \leq 99$.

According to Value Continuity, if some philosopher affirms that V_{99} is equally bad as V_{100}, then she will also have to affirm that V_{98} is equally bad as V_{99}, that V_{97} is equally bad as V_{98}, and so on.

HT gives rise to the mismatch problem for Act Consequentialism only if it is not true that V_{99} is better than V_{100}. By Value Continuity it follows that if HT is

[5] See Spiekermann (2014) for a similar point.

[6] Suppose you are comparing two states of affairs A and B. Imagine you write up a pro and con list. You catalogue all the features that count in favor of A over B. You catalogue all the features that count in favor of B over A. Suppose that you examine the list, and you determine the value relation that obtains between A and B. Now suppose C and D compare to each other with respect to all the same features—they have the same pro and con list. It would be unacceptable to assign a different value relation.

indeed a mismatch problem case, there are no neighboring states V_n and V_{n+1} for which it is true that V_n is better than V_{n+1}. It is for this reason that there are no sharp boundaries among the outcomes in HT. There are no neighboring states where a sudden ‘jump’ in value occurs.

In later sections, these two features of HT—the indistinguishability of neighboring states and Value Continuity—will help us to see how accumulation cases arise. First, however, it is important to see how HT differs from other versions of the mismatch problem.

6.3 Distinguishing HT from Other Problem Cases

As I see it, HT is importantly different from other standard versions of the mismatch problem. The axiological situation is unique. We see this if we contrast HT with certain familiar cases.

Start by considering *Two Voters*.[7] Dr. Mediocre and Professor Beneficent are the two candidates up for public election, and Beneficent is by far the superior candidate. It’s best if Beneficent wins, second-best if Mediocre wins, and worst if the vote results in a tie, in which case no one wins. Vincent and Virgil are the only two voters in the election. Vincent is determined to see Mediocre elected, and so he votes for Mediocre. Furthermore, Vincent is uncooperative: he would vote for Mediocre even were Virgil to vote for Beneficent. Similarly, Virgil is determined to see Mediocre elected, and he’s uncooperative as well; he too votes for Mediocre, and he would do so even were Vincent to vote for Beneficent. Accordingly, the inferior candidate, Mediocre, receives two votes and wins the election. That being said, both voters could have voted for Beneficent.

[7] This case is discussed in Zimmerman (1996) and Kierland (2006). See Regan (1980), Feldman (1980), and Pinkert (2015) for discussions of closely related cases. I discuss this case in Chapter 5.

We may identify two core features of Two Voters. First, the group does something bad but could have done something much less bad. In other words

> *F1:* Some outcome accessible to the group is better than the outcome the group actually brings about.

And second, each voter has an excuse: "I could have voted for Beneficent, but had I done it, the other guy still would have voted for Mediocre, and a split vote would have resulted. So I could have done something else, but if I had done it, the result would have been worse." In other words

> *F2:* For any individual, the outcome he or she actually brings about is better than each of his or her other accessible outcomes.

It is in virtue of feature F2 that we may say that neither voter has an alternative with a better outcome. Act Consequentialism then delivers a verdict of moral permissibility on each of the individual acts. Thus, F1 and F2 together generate a mismatch problem for Act Consequentialism.

But while HT has the first of the two features, the case pretty clearly does not instantiate F2. Had an individual torturer not flipped his switch, the resulting situation would not have been worse. Imagine that we progress through the outcomes accessible to the group of torturers, from V_0 through V_{100}. We are to imagine that the situation gradually becomes worse, and that there's no point at which it becomes better. The value decreases monotonically. Were an individual torturer to abstain from flipping, the resulting state of affairs, V_{99}, would not have been worse than V_{100}. So HT is not an F1 + F2 type case.

Not all standard versions of the mismatch problem have feature F2. Consider *Two Shooters*.[8] You and I are sharpshooters. We shoot at an innocent victim simultaneously, our bullets striking the same fatal location in the victim's chest. Either shot

[8] This case is discussed in Parfit (1984), Jackson (1997), Zamir (2001) and in many other places. See also Chapters 1 and 3.

is sufficient for the victim's immediate death, and neither of us could have prevented the other from shooting: you still would have shot the victim even if I had holstered my weapon; I still would have shot the victim even if you had holstered your weapon. That being said, we both could have refrained from shooting, in which case the victim would have lived. So while our group has an alternative with a better outcome, neither of us has an alternative with a better outcome.

In Two Shooters, the mismatch problem arises through a different set of core features. F1 remains intact: some outcome accessible to the group is better than the outcome the group actually brings about. But each of us has a slightly different excuse than the excuses offered by the voters. The outcome wouldn't have been any worse had either of us not shot. Instead, each of us may say, "I could have done something else, but even if I had done it, *the result would have been exactly as bad*." So instead of F2, the case instantiates

> *F3:* For each individual, all of the outcomes accessible to him or her have the same value.

It is in virtue of feature F3 that we may say that neither individual has an alternative with a better outcome in Two Shooters.

Based on a natural understanding of Act Consequentialism, the theory delivers a verdict of moral permissibility on each of the individual acts in F3-type cases. An act is morally permissible just in case there's no alternative with a better outcome. Suppose that an act a is performed. Suppose that there is some alternative b such that the outcome of b is better than the outcome of a. Then Act Consequentialism says that a is morally impermissible. Suppose, instead, that there is no such alternative. Then Act Consequentialism says that a is morally permissible. In Two Shooters, I do not have an alternative such that its outcome is better than the outcome of my shooting the victim. All of my alternatives would produce outcomes with the same values. Accordingly, Act Consequentialism says that my act is permissible.

Several other problem cases exemplify features F1 and F3. In any case that instantiates F3, we may say that the outcomes accessible to any given individual exhibit a certain uniformity in value: each individual could act differently, but whatever he or she would do, the result would be exactly as bad. Thus, we may call F1 + F3 type cases the 'uniformity' cases.

But HT does not appear to be a uniformity case. It's not plausible that the result would have been *exactly as bad* had an individual torturer not flipped his switch. Each flipped switch brings the victim *closer* to the intense episode of pain he experiences in V_{100}. Do we really want to say that V_{99} is equally as bad as V_{100}?

A simple argument may be given in support of the claim that V_{99} and V_{100} are not equal in value. Recall the assumption of Value Continuity. Whatever evaluative relation obtains between V_{99} and V_{100} also holds between every set of neighboring outcomes. If some philosopher affirms that V_{99} is equally bad as V_{100}, then she will also have to affirm that V_{98} is equally bad as V_{99}, that V_{97} is equally bad as V_{98}, and so on. But I assume that the standard value relations 'better than', 'equally good as', and 'worse than' are transitive. In particular, this means that for any three states of affairs s_1, s_2, and s_3, if s_1 is equally bad as s_2 and s_2 is equally bad as s_3, then it follows that s_1 is equally bad as s_3. I acknowledge that this is a controversial assumption.[9] But for our purposes here, I am prepared to accept the assumption of transitivity as a basic requirement on consequentialist reasoning. Accordingly, by Value Continuity and the transitivity of the standard value relations, it follows that HT does not instantiate F3. For if it is true that V_{100} is equally bad as V_{99}, then it will follow that V_{100} is equally bad as V_0. This is absurd.

So I conclude that HT is axiologically distinct from certain standard versions of the mismatch problem. Unlike Two Voters, HT doesn't instantiate F2; for each torturer

[9] See the vast literature inspired by Temkin (1996).

it is not the case that the outcome he or she actually brings about is better than each of his or her other accessible outcomes. And unlike Two Shooters, HT doesn't instantiate F3; for each torturer, it is not the case that all of his or her accessible outcomes have the same value.[10]

If HT is to give rise to the mismatch problem for Act Consequentialism, then the case must instantiate *some* axiological feature under which it is not true of any individual torturer that he has an alternative with a better outcome. There are several possibilities that would be worth exploring. One possibility is that V_{100} and V_{99} are not evaluatively comparable. This would happen in the case of value incommensurability: two items differ from each other in such a way that it doesn't make sense to compare them on a common scale. But this idea does not seem to apply in HT. V_{100} and V_0 are plainly evaluatively comparable. The two states of affairs differ along just one evaluative dimension: the degree to which each is painful. And as we move through the series, comparing V_{100} with V_1, V_2, and so on up to V_{99}, nothing fundamental changes about the items to be compared. If V_{100} and V_0 are evaluatively comparable, then it is natural to expect that V_{100} and V_{99} will be evaluatively comparable as well.

Another possibility is that a non-standard value relation obtains between them. An example might be Ruth Chang's relation of *parity*.[11] According to Chang, the parity relation applies in certain 'hard cases' of comparison. In such cases, "...although we agree what considerations are relevant to the comparison, it seems all we can say is that the one alternative is better with respect to some of those considerations while the other is better in others, but it seems there is no truth about how they compare all things considered."[12] Parity would apply between V_{100} and V_{99} if the two states

[10] Of course, there are also 'mixed cases'. In such cases, each individual is such that some alternatives to his act have worse outcomes and other alternatives have outcomes equal in value. But it should be clear from the foregoing discussion that HT is not a mixed case either.

[11] See Chang (2002).

[12] (Chang, 2002, 659)

of affairs present us with a hard case of comparison—if V_{100} is better than V_{99} with respect to some considerations and V_{99} is better than V_{100} with respect to others. But this doesn't seem to be the right approach. It's difficult to see what, if anything, would count in favor of V_{100} over V_{99}.

A third possibility is that it is indeterminate which value relation obtains between V_{100} and V_{99}. Under this possibility, it will *not be true* that a torturer has an alternative with a better outcome (though it will not be false either), and so a mismatch will result. I pursue this third possibility: the value comparisons among the outcomes in HT involve indeterminacy. It is not determinately true that V_{99} is better than V_{100}, but it is not determinately false either. In section 6.4, I explain how this works. For present purposes, a sketch will suffice. As we progress through the outcomes accessible to group of torturers, from V_0 through V_{100}, we may recognize a variant of the sorites paradox. There is apparently no adjacent pair of outcomes for which the victim goes from an episode of pain with one intensity to an episode of pain with a greater intensity. Were the victim to inhabit each of the outcomes, one after the other, the transitions between the adjacent sensory feelings in the series would be imperceptible. The victim would begin feeling worse, but he wouldn't be able to identify any particular transition point. So we are to imagine that each torturer brings the victim closer to suffering a worse episode of pain, but no individual flip of a switch determinately makes things worse.[13]

On this basis, I believe that HT is a case in which no individual has an alternative with a better outcome because of feature

[13] I am inspired by Julia Nefsky's understanding of the case. As Nefsky sees it, "...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 he did at the setting immediately prior. But HT seems not to have any such sharp boundaries." (Nefsky, 2012a, 386).

> *F4:* For any individual, the outcomes accessible to him or her are such that the comparative value claims about them are indeterminate.

In HT, it's not that all alternatives have the same value, nor is it that the alternatives are all worse. Rather, the case instantiates F4. Since the disvalue accumulates without any sharp boundaries, I will call any such F4-type case an 'accumulation' case. So we may say that HT is unique; it is different from the other standard versions of the mismatch problem because it is an accumulation case.

It is important to reflect on the exact nature of the mismatch problem for HT given its status as an accumulation case. Each torturer flips his switch, and each torturer could have abstained from flipping. It is indeterminate whether the outcome of switch flipping is worse than the outcome of not flipping. Accordingly, it is *indeterminate* under Act Consequentialism whether it is permissible for a torturer to have flipped his switch. So the mismatch problem that arises for Act Consequentialism in connection with HT is of a different sort than the mismatch problem that arises in connection with the other standard cases. In Two Voters, the group act is determinately wrong, and each individual act is determinately permissible. The mismatch in deontic status is between determinate wrongness and determinate permissibility. A mismatch of the same sort arises in Two Shooters. But in HT, the mismatch results because the group act is determinately wrong, but the deontic status of each individual act is indeterminate between being wrong and being permissible.

Some may not view mismatches of this sort as problematic. I disagree. If it is indeterminate whether an act is wrong, then it is appropriate to have a different moral reaction to the performance of the act than we would have if the act were determinately wrong. A thought experiment illustrates this. Imagine for a moment that a deontological theory is true on which there is an absolute prohibition against killing people. Suppose than an act $A1$ is such that it is indeterminate whether it is the killing of a person—perhaps because it is vague whether the victim is a person.

Suppose that an alternative act $A2$ is such that there is no indeterminacy about whether it counts as the killing of a person: it is determinately a killing. Suppose that an agent has only $A1$ and $A2$ as alternatives. Suppose there are no other relevant moral considerations that are applicable to the case. Perhaps one such case involves the decision about whether to kill a mother in order to avoid killing her premature baby. If you are a firm believer in the absolute prohibition against killing, what will you counsel the agent to do in such a case?[14]

One reaction is that the case as described embodies a moral dilemma for the deontological theory. Nothing speaks in favor of one alternative over the other. But I don't have that reaction. I think that there's a morally relevant difference between the deontic status of A1 and the deontic status of A2 under the deontological theory. I think advocates of the absolute prohibition against killing should prefer A1 over A2. I take this thought experiment to indicate that if it is indeterminate whether an act is wrong, then we should see the act as having a different deontic status from acts that are just plain wrong.

But then HT gives rise to a variation of the mismatch problem for Act Consequentialism. The group act gets assigned one deontic status by Act Consequentialism, but no individual act has that same deontic status. In F4-type cases, we want to find a way to modify Act Consequentialism so that it doesn't deliver mismatched verdicts of this sort. We want to modify the theory so that it says that each torturer acts wrongly, *determinately.*

6.4 Cautious Act Consequentialism

In connection with some accumulation cases, we may develop a relatively straightforward solution to the mismatch problem. We need only think about how Act Conse-

[14] I thank Chris Heathwood for suggesting a thought experiment much like this one.

quentialism should be modified to handle choice situations that involve indeterminate value comparisons. What if it were made determinately morally impermissible to bring about an outcome for which it is indeterminate how that outcome evaluatively compares to the outcome of an alternative? We would then have a moral theory that determinately condemns each of the individual acts in HT. It is indeterminate how the outcome that each torturer in fact brings about evaluatively compares with V_{99}. So a natural proposal is to modify Act Consequentialism in order to advocate *caution* in connection with indeterminacy: not only should you avoid bringing about suboptimal outcomes, but you should also avoid bringing about some outcome if it would be indeterminate how that outcome evaluatively compares to the outcome of an alternative.

Notice that this solution prevents indeterminacy from 'trickling through': indeterminacy exists at the axiological level, but it does not infect the deontic level. This is a natural strategy. We expect a moral theory to tell us what to do when we are confronted with axiological indeterminacy. It's unsatisfying to be told that it's indeterminate what we should do in such situations.

The problem with this initial sketch of a solution is that it's *too cautious.* To see this, notice that the indeterminacy in value relation between V_{99} and V_{100} 'goes both ways'. Reflect for a moment on the nature of the indeterminacy in HT. Each torturer has two accessible outcomes: V_{100} under the condition that he or she flips, and V_{99} under the condition that he or she doesn't flip. As I argued in section 6.3, it is indeterminate what value comparison holds between V_{99} and V_{100}: it is not true that V_{99} is better than V_{100}, but it is not false either.

This indeterminacy also infects whether these two states of affairs are equally bad. Could it be *true* that V_{99} is equally bad as V_{100}? No, because then there would be a value relation that holds determinately between the two states, which would would make it false that V_{99} is better than V_{100}. I assume that if any of the standard value relations holds determinately between these two states of affairs, then it's determinately false that either of the others holds between them.

Could it instead be *false* that V_{99} is equally bad as V_{100}? No, because it is false that V_{99} is worse than V_{100} (see the argument for why HT is not an F2-type case), and I assume that at least one of the standard value relations holds between V_{99} and V_{100}: if it's both false that V_{99} is worse than V_{100} and false that V_{99} is equally bad as V_{100}, then it must be true that V_{99} is better than V_{100}; but it is indeterminate whether V_{99} is better than V_{100}. So it's neither true nor false that V_{99} is equally bad as V_{100}.

In other words, the indeterminacy in HT comes from an indeterminacy between two candidate value relations: *either* V_{99} is better than V_{100} *or* V_{99} is equally bad as V_{100}, but it is indeterminate which. This is a direct result of the profile of truth values concerning the various possible evaluative comparisons between V_{99} and V_{100}, as represented in Table 6.1. Accordingly, if a torturer were to abstain from flipping, it would be indeterminate how the resulting outcome evaluatively compares to the outcome of his alternative. It's not true that his alternative would have resulted in an outcome that's equally bad (though it's not false either). And it's not true that his alternative would have resulted in an outcome that's worse (though it's not false either).

evaluative claim	**truth value**
"V_{99} is better than V_{100}"	neither true nor false
"V_{99} is equally bad as V_{100}"	neither true nor false
"V_{99} is worse than V_{100}"	false

Table 6.1. The nature of the indeterminacy in HT

So the proposal that it is always impermissible to bring about an indeterminately valued outcome makes it so that each torturer would act wrongly no matter what he or she would do. There are two problems with this implication. First, HT doesn't have the appearance of a moral dilemma. Intuitively, there is some act that each torturer should have performed: not flipping his or her switch. Second, consequentialists tend to prefer moral theories that respect ought-implies-can. But the proposal that it is always impermissible to bring about an indeterminately valued outcome violates ought-implies-can in connection with accumulation cases like HT in which the indeterminacy is unavoidable.

We need a less cautious way of formulating the view. We cannot have a blanket prescription against bringing about an indeterminately valued outcome; we need a more qualified approach. Toward this end, reflect further on the nature of the evaluative indeterminacy in HT. The indeterminate value comparisons between V_{99} and V_{100} contain a certain asymmetry. In the series of outcomes V_0 through V_{100}, no outcome is determinately worse than its preceding neighbor, but each outcome "inches closer" to being worse.[15] If this intuitive idea about closeness can be worked out, we may then reformulate Act Consequentialism so as to make 'closeness to being worse' relevant to the moral evaluation of an act. We could express the requisite caution like so: if evaluative indeterminacy is unavoidable, you shouldn't bring about some outcome that would be closer to being worse than some other accessible outcome. Since each torturer brings about some outcome that's closer to being worse than the outcome of not flipping, each torturer would act wrongly under this proposal.

In what follows, I explain one possible way in which this idea could be developed. In section 6.4.1, I offer a diagnosis of the indeterminacy in HT. How could it be that there's no determinate answer to the question of whether V_{99} is better than V_{100}? I believe that one explanation may be found by reflecting more closely on the axiology of pain. In section 6.4.2, I show that the idea about 'closeness to being worse' makes sense under the proposed account of the evaluative indeterminacy in HT. Finally, in

[15] This notion of closeness is easier to grasp in the sequential case. In the counterfactual case, we still have a notion of closeness: it's a relation between counterfactual situations.

section 6.4.3, I present an official formulation of Cautious Act Consequentialism that does not deliver mismatched verdicts in connection with HT.

6.4.1 An Explanation of the Indeterminacy in HT

The key to understanding the evaluative indeterminacy in HT is to look more closely into the nature of pain. Under some ways of thinking about pain, we accord axiological significance to certain extra-sensory aspects of episodes of pain. On a desire-based approach to the axiology of pain, it may be indeterminate whether the pain experienced in V_{100} is worse than the pain experienced in V_{99}. So under the desire-based approach, a bit of indeterminacy creeps in. Or so I suggest in what follows.

Start by considering a certain common sense idea about the axiology of pain. In his discussion of HT, Shelly Kagan states the idea clearly and succinctly:

> Even if there is more to well-being than hedonism posits, it is difficult to take the idea of an imperceptible pain, or an imperceptible increase in the level of pain, seriously, or to believe that it could be bad to impose such a thing. When the relevant bad outcome is pain, what matters morally are differences in how people feel.[16]

As Kagan notes, it is natural to assume that two episodes of pain that are indistinguishable to the victim will have the same value. According to Kagan, it is difficult to see how there could be an alternative account.

But compare the Kagan account with attitudinal conceptions of pain.[17] According to attitudinal conceptions, an episode of sensory pain is a complex state of affairs. Every episode of pain contains a sensory feeling experienced by some subject at some time. But it also contains a distinctive attitude of displeasure that the subject has toward the fact that he or she is experiencing the sensory feeling at that time. In his

[16] (Kagan, 2011, 166)

[17] Perhaps the best treatment of the attitudinal conception of pain is offered by Fred Feldman in Feldman (2004).

discussion of HT, Parfit advocated for an attitudinal conception of the axiology of pain:

> . . . someone's pain can become less painful, or less bad, by an amount too small to be noticed. Someone's pain is worse, in the sense that has moral relevance, if this person minds the pain more, or has a stronger desire that the pain cease. . . [S]omeone can mind his pain slightly less, or have a slightly weaker desire that his pain cease, even though he cannot notice any difference.[18]

According to Parfit's suggestion, one of two indistinguishable sensory feelings may factor into a worse episode of pain if the person has a stronger desire that the one sensory feeling cease than that the other sensory feeling cease.

HT gives us a reason to favor Parfit's axiology of pain over Kagan's axiology. Notice that HT is a uniformity case under Kagan's account: since V_{99} is indistinguishable from V_{100}, these states have the same value. This quickly leads to contradiction under Value Continuity and the transitivity of the standard value relations—recall the argument against taking HT as a uniformity case in section 6.3. Under a desire-based conception of pain, on the other hand, we have a way of avoiding a contradiction. The victim might have a stronger desire that his pain cease under V_{100} than he would have had under V_{99}, despite the fact that these states are indistinguishable to him. So the desire-based account does not force us to see V_{99} and V_{100} as having the same value.

And yet, under a desire-based account, we are not forced to see V_{99} as determinately better than V_{100}, either. Instead, we can explain the possibility of evaluative indeterminacy in HT. We can appeal to the possibility of the victim's having vague desires. Were the victim to experience the sensory feeling associated with V_{99}, he would have a strong desire not to be experiencing it. Under V_{100}, he similarly has a strong desire not to be experiencing the associated sensory feeling. But suppose

[18] (Parfit, 1984, 79)

that were the victim in a position to compare the two sensory feelings he would think to himself "Maybe I desire more strongly that I don't experience the V_{100} feeling, or maybe I don't; I can't really tell."[19]

Under this picture, the desire-based account gives us a natural diagnosis of the indeterminacy in HT. The victim's desires are vague, and this trickles into the axiological relations between neighboring states. Accordingly, there's no determinate answer as to whether V_{99} is better than V_{100}.

6.4.2 Why V_{100} is Closer to Being Worse Than V_{99}

Importantly, note that however the indeterminacy is resolved in HT, it will not be the case that V_{99} is worse than V_{100}. There is a certain plausible constraint on the victim's desires: he does not desire more strongly that his pain cease in V_{99} than he desires his pain cease in V_{100}. One way to motivate this constraint is to reflect on the preponderance of reasons for the victim's desiring more strongly that he doesn't experience the one hundred volt sensory feeling ('sf_{100}') than that he doesn't experience the ninety-nine volt sensory feeling ('sf_{99}'). Though these feelings are indistinguishable to him, sf_{100} puts the victim closer to feeling increased tingliness, sharpness, and throbbingness than experiencing sf_{99} does. The probability of increased discomfort is greater under the experience of sf_{100} than it is under sf_{99}. Supposing that the victim is mistaken about the phenomenal characteristics of sf_{100} and sf_{99}, it is likely that sf_{100} is in fact more uncomfortable than sf_{99}. Supposing that the victim anticipates more tingliness, sharpness, and throbbingness yet to come, his experience of sf_{100} will contain more such anticipation than his experience of sf_{99}.[20]

[19] I thank Chris Heathwood for this suggestion.

[20] Some of these observations are inspired by discussions in Arntzenius and McCarthy (1997), Regan (2000), Voorhoeve and Binmore (2006), and Spiekermann (2014). In each discussion, the author (or authors) point out that the victim has reason to be more concerned about experiencing sf_{100} than about experiencing sf_{99}.

But reasons like these don't necessarily have a direct impact on the strength of the victim's desires. Even if it is *reasonable* to more strongly desire to not experience sf_{100}, the desire to not experience sf_{100} may be of indeterminate strength. This is the key to identifying the indeterminacy in value relations as represented in Table 6.1. When the strength of a desire is vague, we cannot assign a precise number to capture the desire's strength. Instead, we may assign a range of numbers, or a *blur*.[21] When the victim introspects on whether he more strongly desires to not experience sf_{100} than he desires to not experience sf_{99}, he can't tell. There are two possibilities open to the him: either the two desires have equal strength, or the desire to not experience sf_{100} is stronger. We may capture both of these possibilities at once by assigning ranges to the desires as follows. Represent the strength of his desire to not experience sf_{100} as $[-9.9, -10]$. Represent the strength of his desire to not experience sf_{99} as $[-9.8, -9.9]$. The overlap between these ranges explains why the victim has no answer as to whether the former desire is stronger. On a comparison of the lower bound of the former with the upper bound of the other, the strengths are the same. On a comparison of the upper bounds of both, the strengths are different. But on a comparison of the ranges, both possibilities remain open to the victim. This is to my mind a plausible way to try to represent the strengths of vague desires.

But under the assignment of ranges, we need an account of how to compare the strengths of the victim's desires. Consider two accounts. On the first account of comparison, two desires are equally strong if their ranges completely overlap; otherwise, the stronger desire is the one with the greater upper bound. On the second account, two desires are equally strong if there is any overlap in their ranges; otherwise, the stronger desire is the one whose lower bound is greater than or equal to the other's upper bound.

[21] See pages 63 - 64 of Regan (2000) for this idea, but as applied to the values of states of affairs rather than the strengths of desires.

With these two accounts of comparison, we may explain the intuitive way in which V_{100} is closer to being worse than V_{99}. Given the two ways of comparing the strengths of the victim's desires, there's at least one admissible resolution to the indeterminacy under which V_{100} is worse than V_{99}. But there's no resolution in the other direction. There are several ways of making the comparative value claims precise; under one of these precisifications V_{100} is worse than V_{99}; but under none of these precisifications is V_{99} worse than V_{100}.

It may be helpful to see how this account of closeness works within a supervaluationist treatment of vagueness. Consider two men, Shorty and Barry. Shorty is determinately not tall while Barry is on the borderline—it is indeterminate whether Barry is tall or not tall. Supervaluationists hold that the predicate 'is tall' is vague, and it has not been completely decided when the predicate applies. There are different acceptable precisifications of the predicate. On some precisifications, "Barry is tall" counts as true; on others, "Barry is tall" counts as false. Shorty, on the other hand, is determinately not tall because "Shorty is tall" does not come out as true on any of the acceptable precisifications. The supervaluationist framework gives us a way to explicate the intuitive way in which Barry is closer to being tall than Shorty is. Since there are many acceptable precisifications of 'is tall', there are many admissible resolutions to the indeterminacy. But however we resolve the indeterminacy in Barry's case, it is important to see that "Shorty is tall" will not count as true. On this basis, we may say that Barry is closer to being tall than Shorty is: there is a resolution to the indeterminacy under which "Barry is tall" counts as true; there is no resolution to the indeterminacy under which "Shorty is tall" counts as true.

To be clear, we need not assume a supervaluationist framework in order to make sense of the idea that V_{100} is closer to being worse than V_{99}. We need not assume that the indeterminacy in HT is a matter of linguistic vagueness. Instead, we need only assume that there is more than one admissible resolution to the indeterminacy in HT. On one admissible resolution V_{100} is worse than V_{99}; on another V_{100} is equally bad as V_{99}. On no admissible resolutions is V_{100} better than V_{99}.

The full account goes like this. The victim's desires are vague. We have two accounts of how to compare the strengths of vague desires. Each of these resolves the axiological indeterminacy. On one resolution, V_{100} is worse than V_{99}. On the other V_{100} and V_{99} are equally bad. But on no resolution is V_{99} worse than V_{100}. On this basis, V_{100} is closer to being worse than V_{99}.

6.4.3 An Official Formulation of Cautious Act Consequentialism

All that remains is to formulate a cautious version of Act Consequentialism so as to make 'closeness to being worse' relevant to the moral evaluation of an act. The basic idea is that when evaluative indeterminacy is unavoidable, you shouldn't bring about some outcome that would be closer to being worse than some other accessible outcome. To make this idea more precise, consider *Cautious Act Consequentialism* (CAC),

(i) In choice situations without indeterminacy, an act is permissible just in case there's no alternative with a better outcome.

(ii) In choice situations involving indeterminacy, an act a is impermissible if there's some alternative b such that the outcome of a is closer to being worse than the outcome of b. An act is permissible otherwise.

The second clause of CAC captures the requirement to act with caution in situations of indeterminacy—to avoid bringing about some outcome that's closer to being worse than the outcome of some alternative.

I do not offer a thorough defense of CAC here. My aim is modest. I aim only to show that CAC does not encounter the mismatch problem in connection with HT. I don't provide independent motivation for accepting CAC; I don't defend the theory from imagined objections. Exploration of the plausibility of CAC is left for future research. For our purposes here, I aim only to show that if we modify Act Consequentialism in the way I suggest, then the mismatch problem for HT can be solved.

And indeed CAC does not deliver mismatched verdicts in connection with HT. The group act is determinately wrong—there is an alternative group act (no torturers flipping their switches) with a determinately better outcome. And each component individual act of switch flipping is also morally wrong. Each torturer has an alternative under which V_{99} would result. The actual outcome, V_{100}, is closer to being worse than V_{99}. So each torturer acts determinately impermissibly according to clause (ii) of CAC. Determinately wrong group act; determinately wrong individual acts. Mismatch resolved.

Thus, CAC constitutes the promised partial solution to the mismatch problem for Act Consequentialism. The solution is meant to apply only to HT and to other axiologically similar cases. The solution relies on the contrast between the axiological situation in HT and the axiological situations in familiar versions of the mismatch problem. Unlike Two Voters, each individual in HT lacks a worse alternative. And unlike Two Shooters, it is not the case that each individual in HT has accessible outcomes all of the same value. Instead, HT is an accumulation case; it involves indeterminate value comparisons. So clause (ii) becomes relevant, and it condemns each of the individual acts.

To sum it up: I have offered a possible diagnosis of the indeterminacy based on the possibility of vague desires; I've used the diagnosis to make sense of the idea that each torturer makes the situation closer to being worse than it otherwise would be; and I have offered a modified consequentialist theory, CAC, in order to condemn each of the torturers for making things closer to being worse. In doing so, I hope to have offered a step forward for consequentialists: once we recognize the axiological nature of accumulation cases, one of the most perplexing of the mismatch problem cases may be solved.

6.5 Applying CAC to Climate Change

Some cases of anthropogenic climate change are plausibly understood as accumulation cases. In this section I will suggest that we can apply CAC to these cases to resolve the mismatch problem: each emitter acts wrongly not because he or she makes things worse, but because he or she makes things closer to being worse.

First, we must identify the relevant cases. Reflect on the axiology underlying certain climate disruptions. Most climate disruptions are overall bad. Droughts, heatwaves, floods: events like these impose uncompensated hardships on large numbers of people. In some cases, the badness of a climate disruption is 'on/off'. Imagine a town in which heavy rainfalls are usually harmless. Extreme rainfall would be bad only were water to breach the levee. Consider a series of high precipitation events, ranging from large amounts of rain to extreme amounts of rain. There is some point in this series at which the resulting volume of water breaches the levee. Before that point, nothing bad would occur. The badness is either 'off' or 'on'. In other cases, the badness of a climate disruption comes in degrees. The higher the temperature soars during a string of hot days, for example, the more hardship will be experienced. Consider a series of hot summer days, ranging from moderately hot to extreme. Every point in the series involves more discomfort and carries a greater risk of minor illness, hospitalization, and death. An extra bit of heat can make an existing medical condition slightly worse.

Under certain assumptions, some degreed cases are accumulation cases. For the purposes of illustration, let's focus on a particular degreed climate disruption: the unprecedented high temperatures throughout Europe during the summer of 2003. Scientists have attributed this particular heat wave to anthropogenic climate change.[22] Some cases of hardship suffered under the European heat wave bear a striking similarity to HT. First, consider our 'victim'. Say that Holly is working outside in Paris on a particular weekday during the heat wave. Though Holly is not at any increased risk of illness or mortality, Holly is very uncomfortable in the high heat: for Holly, the badness of this particular very hot summer day is a function of the degree to which she experiences discomfort.

Second, consider our 'torturer collective'. Prior to Holly's work day—let it be the day exactly one year before—each of millions of individuals throughout the world produced a *superfluous emission*: each of these people performed an act that resulted in a greater production of CO_2 than would have been produced by some alternative, and each performed the more highly emitting act without securing any additional benefit—each would have been no worse off had he or she pursued a lesser emitting alternative. These superfluous emitters composed a group, the *Global Superfluous Emitters Collective* (GSEC).

It is important to clarify the differences between three types of emission behavior. Does the emitter gain a benefit, and if so, what kind of benefit? In many situations, certain emissions are required in order to secure basic human needs. For example, I may need to run a gas-powered generator in order to pump fresh water to drink. The benefit I gain from emitting is substantial—I couldn't do without it. This is an example of a *subsistence emission.* In other situations (mostly in industrialized countries), acts that produce greenhouse gases are associated with securing more comfortable lifestyles. For example, I may prefer to use a gas powered lawn-mower rather than a manual one in order to save some effort. The benefit I gain from

[22] In particular, Mitchell et al. (2016) associates 70,000 premature deaths with the heat wave—deaths that would not have occurred had humans not contributed greenhouse gases into the climate system.

emitting in this case is significant—it makes my life better—but I could easily do without it. This is an example of a *luxury emission.* In yet other situations, acts that produce greenhouse gases are entirely pointless: they don't make life better for the emitter at all. For example, I may forget to switch off the light in my bathroom at night, or I may keep my car idling when I run inside to grab something. Such *superfluous emissions* don't contribute any positive value to the world. It would make no difference to the emitter to cut back in these cases.

When we consider the 'torturer collective' in the case of Holly and the European heat wave, we are to imagine that each emitter in the group performs a superfluous emission on the day in question. Each emitter has an alternative under which he or she emits less without it having any effect on his or her level of wellbeing. Accordingly, GSEC could have emitted much less without it making anyone worse off. Notice that a similar thing is true of the torturer collective in HT. It could have flipped fewer switches without it making anyone worse off. In HT and in the case of Holly, we are imagining that the only bad-making feature of any of the outcomes accessible to the group is the well-being of the victim.

Third, consider our 'torture machine'. In Holly's case, it's the climate system. It has a huge number of 'settings', each corresponding to some atmospheric concentration of CO_2. The higher the atmospheric concentration of CO_2, the greater the amount of heat trapped by the climate system. And the greater the degree to which this results in global warming, the higher the temperature on Holly's work day.

GSEC has several outcomes accessible to it, each of which may be specified by some amount of CO_2 in kilograms taken up by the atmosphere. Let the 'baseline' emission level, b kilograms, be equal to the amount of CO_2 that GSEC would contribute to the atmosphere were every member of GSEC to pursue a lesser emitting alternative. Let 'A_b' represent the atmosphere taking on this baseline amount of CO_2. And let's imagine that each individual superfluous emission contributes an additional 1 kg of

CO_2, so that if just one member of GSEC were to pursue a superfluous emission—while the rest refrained—the atmosphere would take on $b + 1$ kilograms of CO_2. The result of exactly one member of GSEC pursuing a superfluous emission would be A_{b+1}; the result of exactly two members of GSEC pursuing superfluous emissions would be A_{b+2}; and so on. For the sake of simplicity in exposition, let's imagine that GSEC has exactly one million members. Then there are one million relevant outcomes accessible to GSEC: A_b through $A_{b+1,000,000}$.

Under $A_{b+1,000,000}$ (the climatic situation in the actual world) let's say that the average temperature on Holly's work day was 90° Fahrenheit. Suppose that the average temperature on Holly's work day would have been 85° had every member of GSEC pursued a lesser emitting alternative. In other words, 85° is the average temperature on Holly's work day under A_b.

Suppose that the relationship between atmospheric concentration of CO_2 and the average temperature on Holly's work day is linear.[23] Then the average temperature under A_{b+n} is $85 + (\frac{5}{1{,}000{,}000})n°$. The basic idea is that, for every additional individual superfluous emission over the baseline, the temperature would go up $\frac{5}{1{,}000{,}000}$th of a degree. So when all one million members of GSEC produce superfluous emissions, the result is that the temperature goes up 5° over what it would have been had all of these people pursued a lesser emitting alternative.

At this point, we have all the required ingredients in order to identify a parallel between some instances of anthropogenic climate change and HT. We have a victim: Holly. We have a torturer collective: GSEC. We have a torture machine with multiple settings: the climate system under A_b through $A_{b+1,000,000}$. But do we have the distinctive axiological feature to complete the parallel? It would need to be the case

[23] Note that this is ultimately an empirical assumption. I am not prepared to engage in any sort of empirical investigation of the plausibility of this particular function. Thus, we may understand the conclusions I draw as conditional.

that each emitter brings about an outcome that's *closer to being worse* than another outcome accessible to him or her (though it is not in fact worse). That is, it would need to be the case that the evaluative comparison between $\mathrm{A}_{b+1{,}000{,}000}$ and $\mathrm{A}_{b+999{,}999}$ is indeterminate; and there would need to be some plausible explanation for the indeterminacy.

It is quite plausible that an incremental change of one $\frac{5}{1{,}000{,}000}$th of a degree is too tiny for Holly to perceive. Thus, for all i, the sensory feeling that Holly would experience under A_{b+i} is qualitatively identical with the sensory feeling that she would experience under A_{b+i+1}. The situation is just like that in HT. Just as neighboring settings of the torture machine have the same 'feel' for the victim, neighboring states of the climate system have the same 'feel' for Holly. Holly suffers an intense episode of discomfort in $\mathrm{A}_{b+1{,}000{,}000}$, but no individual act within the collection of individual acts makes things determinately worse than they would have been were that act not performed. Why not? Because in Holly's case, the badness of $\mathrm{A}_{b+1{,}000{,}000}$ is limited to the degree to which she experiences discomfort. How Holly feels is the only bad making feature, and she would have experienced a qualitatively indistinguishable sensory feeling under $\mathrm{A}_{b+999{,}999}$.

For similar reasons as those discussed in connection with HT, we certainly don't want to say that $\mathrm{A}_{b+999{,}999}$ is worse than $\mathrm{A}_{b+1{,}000{,}000}$. But we also don't want to say that $\mathrm{A}_{b+999{,}999}$ is equally bad as $\mathrm{A}_{b+1{,}000{,}000}$: under certain plausible axiological assumptions, this leads to absurdity. Instead, I believe we should say that it is indeterminate whether $\mathrm{A}_{b+999{,}999}$ is better than $\mathrm{A}_{b+1{,}000{,}000}$ or equally bad. I believe this can be explained by appeal to the possibility of Holly's having desires of indeterminate strength—just as was described in connection with HT.

Since it is not true that $\mathrm{A}_{b+999{,}999}$ is better than $\mathrm{A}_{b+1{,}000{,}000}$, no individual member of GSEC has an alternative with a better outcome. The case gives rise to a version of the mismatch problem for Act Consequentialism. But because there is indeterminacy in the evaluative comparisons of each individual's accessible outcomes—and because this indeterminacy is axiologically analogous to the indeterminacy found in HT—CAC will not encounter the mismatch problem. Each member of GSEC has an alternative such that the outcome he or she actually brings about is closer to being worse than the outcome of that alternative. According to CAC, each individual has acted wrongly by his or her superfluous emission.

Of course, Holly is not the only victim of anthropogenic climate change. And there are many groups of emitters other than GSEC. Thus, it's important to clarify that the solution here resolves the mismatch problem only under certain assumptions about the nature of some climatic harms brought about by some groups. I have argued that the case of Holly and the European heat wave is an accumulation case. Accordingly, CAC delivers concordant verdicts for this case. But other cases plausibly have F2 or F3-type structures, and CAC doesn't resolve the mismatch problem for these cases.

6.6 Where This Leaves Us

In this dissertation, I have presented the mismatch problem for Act Consequentialism and distinguished it from related problems. I have raised difficulties for some popular solutions to the mismatch problem, and I have proposed and defended a solution for accumulation cases. I have applied my solution to a version of the climate change case that originally got us thinking about the problem in Chapter 1.

One theme running through this dissertation is that there are a variety of mismatch problem cases. These include Kagan-style cases (discussed in Chapter 4 and mutually-restricted option cases (like Two Voters+, discussed in Chapter 5). In this chapter, I identified three axiologically distinct versions of the mismatch problem: F2-type cases, F3-type 'uniformity' cases, and F4-type 'accumulation' cases. The taxonomy suggests that even if a solution works for one type of case, we shouldn't expect it to work for the other types of case.

Where does this leave us? I believe this leaves us in a position that's both frustrating and promising. Frustrating because the solution I offer in this chapter does not cover all of the problem cases: it is at best a partial solution. We need to do more work if we want to modify Act Consequentialism to resolve the F2- and F3-type cases. But the position we are in is promising because we now have a more precise picture of how to tackle these cases. For example, it seems to me that it shouldn't be too difficult to formulate a version of Act Consequentialism that condemns the individual acts in F3-type cases. But that's a project for another time.

BIBLIOGRAPHY

Andreou, C. (2014). The good, the bad, and the trivial. *Philosophical Studies*, 169(2):209–225.

Arntzenius, F. and McCarthy, D. (1997). Self torture and group beneficence. *Erkenntnis*, 47(1):129–144.

Baier, K. (1958). *The Moral Point of View*. Cornell University Press.

Barnett, Z. (2017). No free lunch: The significance of tiny contributions. *Analysis*.

Bratman, M. E. (1993). Shared intention. *Ethics*, 104(1):97–113.

Broome, J. (2012). *Climate Matters: Ethics in a Warming World*. W W Norton and Company.

Budolfson, M. (2014). The ethics of the marketplace and a surprisingly deep question for normative theory: What are consumers required to do when products are produced in morally objectionable ways? retrieved from personal webpage.

Castaneda, H. N. (1974). *The Structure of Morality*. Charles Thomas Publisher.

Chang, R. (2002). The possibility of parity. *Ethics*, 112(4):659–688.

Christian Barry, G. O. (2015). Individual responsibility for carbon emissions: is there anything wrong with overdetermining harm? In Moss, J., editor, *Climate Change and Justice*, pages 165–183. Cambridge University Press.

Collins, S. (2013). Collectives' duties and collectivization duties. *Australiasian Journal of Philosophy*, 91(2):231–248.

Feldman, F. (1980). The principle of moral harmony. *The Journal of Philosophy*, 77(3):166–179.

Feldman, F. (1986). *Doing the Best We Can*. D. Reidel Publishing Company.

Feldman, F. (2004). *Pleasure and The Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. Clarendon Press.

Feldman, F. (2006). Actual utility, the objection from impracticality, and the move to expected utility. *Philosophical Studies*, 129:49–79.

Forcehimes, A. T. and Semrau, L. (2015). The difference we make: A reply to Pinkert. *Journal of Ethics and Social Philosophy.*

Gibbard, A. (1990). *Utilitarianism and Coordination.* Garland.

Gibbard, A. F. (1965). Rule-utilitarianism: Merely an illusory alternative? *Australiasian Journal of Philosophy*, 43(2):211–220.

Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies in Philosophy*, 40.

Glover, J. (1975). It makes no difference whether or not I do it. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 49:171–209.

Goldman, A. (1999). Why citizens should vote: A causal responsibility approach. *Social Philosophy and Policy*, 16(2):201–217.

Hiller, A. (2011). Climate change and individual responsibility. *The Monist*, 94(3):349–368.

Horwich, P. (1974). On calculating the utility of acts. *Philosophical Studies*, 25(1):21–31.

Hourdequin, M. (2010). Collective action and individual ethical obligations. *Environmental Values*, 19(443-464).

Hylland, A. and Zeckhauser, R. (1979). The impossibility of bayesian group decision making with separate aggregation of beliefs and values. *Econometrica*, 47(6):1321–1336.

Jackson, F. (1987). Group morality. In Smart, J., Pettit, P., Sylvan, R., and Norman, J., editors, *Metaphysics and Morality: Essays in Honour of J.J.C. Smart.* Blackwell.

Jackson, F. (1997). Which effects? In Dancy, J., editor, *Reading Parfit*, pages 42–53. Blackwell Publishers.

Jamieson, D. (2007). When utilitarians should be virtue theorists. *Utilitas*, 19(2):160–183.

Kagan, S. (2011). Do I make a difference? *Philosophy and Public Affairs*, 39(2):105–141.

Kernohan, A. (2000). Individual acts and accumulative consequences. *Philosophical Studies*, 97(3):343–366.

Kierland, B. (2006). Cooperation, ‘ought morally’, and principles of moral harmony. *Philosophical Studies*, 128(2):381–407.

Killoren, D. and Williams, B. (2013). Group agency and overdetermination. *Ethical Theory and Moral Practice*, 16:295–307.

Kuhn, S. (2017). Prisoner's dilemmas. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2017/entries/prisoner-dilemma/.

Kutz, C. (2000). Acting together. *Philosophy and Phenomenological Research*, 61(1):1–31.

Lawford-Smith, H. (2016). Difference-making and individuals' climate-related obligations. In Hayward, C. and Roser, D., editors, *Climate Justice in a Non-Ideal World*. Oxford University Press.

Lyons, D. (1965). *Forms and Limits of Utilitarianism.* Oxford Clarendon Press.

Mitchell, D., Heaviside, C., Vardoulakis, S., Huntingford, C., Masato, G., Guillod, B. P., Frumhoff, P., Bowery, A., Wallom, D., and Allen, M. (2016). Attributing human mortality during extreme heat waves to anthropogenic climate change. *Environmental Research Letters*, 11.

Nefsky, J. (2012a). Consequentialism and the problem of collective harm: A reply to Kagan. *Philosophy and Public Affairs*, 39(4):364–395.

Nefsky, J. (2012b). *The Morality of Collective Harm.* PhD thesis, University of California, Berkeley.

Nefsky, J. (2016). How you can help, without making a difference. *Philosophical Studies.*

Nolt, J. (2011). How harmful are the average american's greenhouse gas emissions? *Ethics, Policy and Environment*, 14(1).

Nolt, J. (2013). The individual's obligation to relinquish unnecessary greenhouse-gas-emitting devices. *Philosophy and Publich Issues*, 3(1):139–165.

Norcross, A. (2004). Puppies, pigs, and people: Eating meat and marginal cases. *Philosophical Perspectives*, 18.

Parfit, D. (1984). *Reasons and Persons.* Clarendon Press, Oxford.

Pinkert, F. (2015). What if I cannot make a difference (and know it). *Ethics*, 125(4):971–998.

Portmore, D. W. (2016). Maximalism and moral harmony. *Philosophy and Phenomenological Research*, (published online).

Rabinowicz, W. (1988). Act-utilitarian prisoner's dilemmas. *Theoria*, 55(1):1–44.

Regan, D. (1980). *Utilitarianism and Co-operation.* Clarendon Press.

Regan, D. H. (2000). Perceiving imperceptible harms (with other thoughts on transitivity, cumulative effects, and consequentialism). In Almeida, M. J., editor, *Imperceptible Harms and Benefits*, pages 49–73. Kluwer Academic Publishers.

Rentmeester, C. (2010). A Kantian look at climate change. *Climate Ethics*, 11(1).

Roberts, M. (2011). The problem of harm in the multiple agent context. *Ethical Perspectives*, 18(3):313–340.

Singer, P. (1972). Is act utilitarianism self-defeating? *The Philosophical Review*, 81(1):94–104.

Singer, P. (1980). Utilitarianism and vegetarianism. *Philosophy and Public Affairs*, 9(4):325–337.

Singer, P. (1998). A vegetarian philosophy. In Griffiths, S. and Wallace, J., editors, *Consuming Passions*, pages 66–72. Manchester.

Sinnott-Armstrong, W. (2005). It's not my fault: Global warming and invidual moral obligations. In Sinnott-Armstrong, W. and Howarth, R. B., editors, *Perspectives on Climate Change: Science, Economics, Politics, Ethics*, volume 5. Emerald.

Spiekermann, K. (2014). Small impacts and imperceptible effects: Causing harm with others. *Midwest Studies in Philosophy*, 38.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M. M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgely, P. M. (2013). *IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.

Temkin, L. (1996). A continuum argument for intransitivity. *Philosophy and Public Affairs*, 25(3).

Velleman, J. D. (1997). How to share an intention. *Philosophy and Phenomenological Research*, 57(1):29–50.

Voorhoeve, A. and Binmore, K. (2006). Transitivity, the sorites paradox, and similarity-based decision-making. *Erkenntnis*, 64(1):101–114.

Zamir, T. (2001). One consequence of consequentialism: Morality and overdetermination. *Erkenntnis*, 55(2):155–168.

Zimmerman, M. (1992). Cooperation and doing the best one can. *Philosophical Studies*, 65(3):283–304.

Zimmerman, M. (1996). *The Concept of Moral Obligation.* Cambridge University Press.