COLLECTIVE ACTION, CLIMATE CHANGE, AND THE ETHICAL SIGNIFICANCE OF FUTILITY

Mark Bryant Budolfson

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF PHILOSOPHY

Advisor: Gideon Rosen

April 2012

Table of Contents

Table of Contents…….……. iii

Abstract……..…….iv

Part One: What Should We Do?

Solving Seemingly Intractable Collective Action Problems……..2

Strategy for Introducing a Climate Treaty……21

Part Two: What Should You Do?

Climate Change and the Ethical Significance of Futility…….25

Consequentialism, the Inefficacy Objection, and the Ethics of the Marketplace……101

Why Morality and All Other Forms of Normativity are Sometimes Dramatically Directly Collectively Self-Defeating…….……130

The Equilibrium Objection…….……..…165

Conclusions…….169

Acknowledgements…….……..…….…….....173

Bibliography……176

Abstract

Some collective action problems are easy. For example, if a serious tragedy of the commons arises within a nation's borders, and if it is clear that the problem can be solved only by a particular kind of coercive governmental response, and if it is clear that every citizen would want the government to adopt such a response, then the government should adopt such a response, and such a response is justified. This is an easy collective action problem because it is obvious *what* course of action should be taken, and it is also clear enough *why* that course of action would be justified.[1]

In this dissertation I answer questions that arise from the most difficult collective action problems, such as our situation with respect to climate change, with respect to which neither economists nor philosophers have made any real progress. For example, in the first part of the dissertation I answer the following question:

> *What should we do when we face a collective action problem that is impervious to our standard tools for achieving cooperation and avoiding a*

[1] For example, consider taxation to provide a public good that is more valuable to each taxpayer than the tax required to provide it, where that good cannot plausibly be provided in any other way. In many nations, law enforcement, national defense, and national parks are approximate examples of such goods.

tragic outcome? In other words, what should we do when we are unable to rely on benevolence, natural 'market-based' coordination, bargaining, governmental coercion, and other familiar mechanisms to avoid an undesirable outcome?

My answer is that even when our standard tools are useless, we can often escape a tragic dilemma by turning the force of such a dilemma against itself. The trick is to intentionally place ourselves into a new dilemma of our own design, the predictable outcome of which is a desirable solution to the initial, otherwise insoluble dilemma. For example, with respect to anthropogenic climate change, where any effective response will be contrary to the interests of the current citizens of many nations, the trick is to engineer and introduce an international climate treaty that creates something like a multi-player prisoner's dilemma at the level of nations, the predictable outcome of which is that all nations will agree to the treaty, thereby making the current citizens of some of those nations worse off relative to the no-treaty status quo as a side-effect of a mechanism within the treaty that effectively combats climate change. I show that such a treaty is the only attractive option if – as seems true – we must secure the cooperation of many influential nations whose current citizens will be made worse off by any effective response, and if such nations are disposed to reject or pull out of treaties whenever they are contrary to the interests of their current citizens. I also explain why such a treaty is the key to engineering the most *ethical* climate treaty possible, and to making genuine progress beyond the unpromising *incrementalist, idealist*, and *hard-headed realist* approaches to climate treaties that

dominate the literature in philosophy, politics, economics, and law. More generally, the result is a straightforward, realistic, and non-coercive framework for constructing economically and ethically optimal solutions to many previously intractable collective action problems.

In the second part of the dissertation, I show that nations can be justified in acting in ways that make their citizens worse off, are unfair to their citizens, and that would be reasonably rejected by their citizens – which provides a counterexample to many political theories. Such situations arise most clearly when the citizens of a nation are, through no fault of their own, morally required to favor a national course of action that is unfair to them, contrary to their interests, and contrary to the interests of future generations of their descendants and fellow citizens. Arguably, the United States is in such a situation with respect to climate change, given that inaction on the part of the US may well ensure a catastrophic outcome for untold billions of future people, and given that any real solution to the problem arguably requires the US to agree to a climate treaty that is unfair to its citizens and contrary to their interests.[2]

At the same time, even given worst-case-scenario assumptions about the effects of climate change, I argue that *individual citizens* are not required to reduce their emissions by a significant amount, because without an effective global response such reductions are both costly and futile in a way that makes such reductions not required. In fact, I argue

[2] My own view is that an effective and ethical response need not be unfair to the United States or contrary to its long-run interests. Such a view of the empirical facts does not undermine my argument that a nation *could* be justified in acting in a way that is unfair to its citizens and contrary to their long-run interests if the empirical facts were different.

that such futility and lack of requirement at the individual level explains why individuals are required *to favor* a coercive intergovernmental response to climate change, and why nations are justified in adopting such a response even if it is unfair to their citizens and contrary to their interests. (Here it might be useful to recall that in the first part I explain why an effective intergovernmental solution is realistic even given pessimistic assumptions about the collective action problems that arise at the level of nations.)

In the course of arguing for these conclusions, I make progress on a number of issues at the foundations of normative ethics and political philosophy. Among other things, I analyze the ethical significance of futility, distinguishing between, on the one hand, cases in which actions are required despite being futile in some intuitive sense, and, on the other hand, cases in which actions are not required because they are futile in a more complete sense that makes actions genuinely not required. Futility has not been adequately investigated by philosophers and other theorists, and is of great practical and theoretical importance, because worries about futility arise in almost every area of practical concern, and threaten to undermine most normative theories. For example, I show that existent normative theories are unable to offer a plausible account of what individuals are required to do in the kind of collective action situations that are common in a market-based society, and are therefore unable to explain many of the most important facts about modern moral life. In particular, straightforward mathematical and empirical considerations show that an appeal to expected consequences cannot possibly deliver the verdicts on such cases that consequentialists themselves assume, and I show that alternative theories that appeal to 'universalizability', 'direct harm', and other notions also cannot deliver plausible verdicts on such cases. For these and other reasons, I argue

that a plausible account of what individuals are required to do in a large market-based society must invoke a distinction between activities that are *essential* to a product or to the actual production of a product, and activities that are not. These and other results related to futility have important practical implications, because many serious practical issues are tied to situations in which the action of individuals is in some sense futile – for example, whether it is permissible to consume animal products given the terrible way that many animals are treated, whether it is permissible to consume products that are produced in a way that harms or exploits other people, whether individuals are required to contribute to charities that help innocent people who are suffering from easily curable afflictions, and, of course, whether individuals are required to reduce their greenhouse gas emissions.

In the final chapters, I show that morality and other forms of normativity are sometimes *dramatically directly collectively self-defeating*, where a normative system has that property when it requires everyone to act in a way that everyone can see is certain to be dramatically worse along the relevant normative dimension than if everyone did not follow the requirements of that normative system instead. This shows that a wide range of normative theories are either false, or at least don't have the consequences that their adherents take them to have, including consequentialist theories, contractualist theories, Kantian theories, universalization theories, enlightened self-interest theories, and many other normative theories. It also means that morality and other forms of normativity cannot be relied upon to solve collective action problems even in a world of normatively flawless agents. One practical upshot is that even when a disaster will ensue if everyone

acts in a particular way or on a particular principle, that doesn't settle the question of whether individuals are permitted to act in that way or on that principle.

In sum, I provide answers to the most pressing questions about what we should do about difficult collective action problems, especially when non-cooperation would lead to catastrophe.

Part One: What Should We Do?

Solving Seemingly Intractable Collective Action Problems

There is now some agreement among scientists, economists, and other experts on how to answer the following questions:

> From a scientific perspective, how and to what extent are human activities causing climate change, and what will be the effects of those changes in the future under various possible alternative courses of action?
>
> From an economic perspective, what are the costs and benefits of the various possible strategies for combating climate change, and what are the most efficient mechanisms for distributing the costs of such strategies, on the assumption that the nations of the world would be perfectly willing to implement such strategies and mechanisms?
>
> From an ethical perspective, what is the best response to climate change and what is the best way of distributing the costs of such a response, on the assumption that the nations of the world would be perfectly willing to adopt such a response and such a distribution of costs?

Given answers to the questions above, it might seem that the obvious next step is to advocate national policy that reflects those answers, and to modify our lives as individuals in a way that would complement such national policy – and it might seem that anyone who resists taking this next step must be either ignorant, unethical, or irrational.

Unfortunately, even if we agree about the answers to the questions above,[1] those answers do not really tell us what we ought to do about climate change either collectively or individually. The problem is that the answers to the questions above tell us only what course of action would be best *if nations of the world were willing to cooperate and adopt an ideal course of action*. However, in the real world, the problem of how to secure cooperation on a chosen course of action is the most important and the most intractable problem that stands in the way of an effective response to climate change. For example, even if we agree that in an ideal world nations would ratify an international treaty creating a cap and trade scheme, where tradable emissions permits would be allocated on an equal per-capita basis, it is nonetheless clear that in the real world such a treaty has no chance of being ratified by the United States and many other major emitters, at least not in any form that has been proposed. As a result, insistence on such a treaty would ensure that the problem of climate change is never successfully addressed, which makes such an approach the wrong approach in the real world, even if it is the approach that would be best in an ideal world.

[1] It is worth noting that there is more disagreement among economists about how to think about the costs and benefits of a global response to climate change than is communicated to the educated public. For a sample of the spectrum of views, see *The Economics of Climate Change: The Stern Review*, William Nordhaus, *A Question of Balance,* Bjorn Lomborg, *The Skeptical Environmentalist*, and Bjorn Lomborg ed., *Global Crises, Global Solutions* Second Edition.

Such real-world difficulties also threaten the idea that individuals have an obligation to fight climate change by reducing their carbon footprints, because if it is clear that nothing effective is being done about climate change, then it is hard to see how individuals could be required to reduce their carbon footprints in a way that seems both costly and futile. For example, if the rest of the world is not doing anything and has no likelihood of doing anything effective about climate change, then it is hard to see how you and your family could have a genuine ethical obligation to reduce your carbon footprint by making costly sacrifices of consumption activities that make significant contributions to your well-being, such as travel, safe transportation, and affordable goods – especially in light of the fact that reducing your carbon footprint does not really do any good for anyone else, and would therefore be futile as well as costly.

As a result, questions about what we as a nation and questions about what we as individuals should do about climate change are not settled by, and have no obvious connection to, questions about what it would be best to do in the sort of idealized world that is the focus of most academic research, because the idealized world of such research abstracts away from the collective action problems that are the most important and most intractable part of the real-world problem. So, in order to reach confident conclusions about what we should actually do about climate change – both collectively and individually – we must first grapple with the question of what we should do when we face such seemingly intractable collective action problems, especially when non-cooperation would lead to catastrophe. That is the main focus of this dissertation.

To appreciate the challenge involved here, it is important to see that previous work on collective action problems is of little or no help in answering these questions. For example, game theory tends to abstract away from the facts that are often crucial to solving real-world collective action problems,[2] and in any event provides no insight into how to bring about a desirable solution in the case of climate change, serving only to dramatize the apparent intractability of the situation. At the same time, more practical discussions of collective action problems have focused only on what might be called *easy collective action problems* – i.e., collective action problems in which it is clear after sufficient empirical investigation how to bring about collectively optimal solutions, where the only important challenges are gathering the relevant empirical evidence, constructing a theory that systematizes our knowledge of the best solutions in various circumstances,[3] and constructing a philosophical theory that explains why such solutions are *justified* from an ethical and political perspective. As a trivial example, if a serious tragedy of the commons arises within a nation's borders, and if it is clear that the problem can only be solved by a particular kind of coercive governmental response, and if it is clear that every citizen should want the government to adopt such a response, then it is clear how best to solve the problem. This is a trivially easy collective action problem, because it does not require any interesting insight or empirical investigation to see how best to solve the problem. As a highly non-trivial example, Elinor Ostrom has argued on empirical grounds that if a small group must cooperate to manage resources that are

[2] For example, see Thomas Schelling, "Bargaining, Communication, and Limited War".

[3] For example, see Elinor Ostrom, "A diagnostic approach for going beyond panaceas", and Elinor Ostrom, "A General Framework for Analyzing Sustainability of Social-Ecological Systems".

under threat of overuse beyond their carrying capacity, where such collective overuse would be contrary to the interest of each individual, and if the actions of each individual can be monitored by the others, and if other realistic conditions are met, then the best solution is to allow natural 'market-based' coordination and bargaining among the parties involved, with minimal governmental interference except to provide appropriate background institutions to facilitate that coordination and bargaining.[4] This is an important result because it shows that the kind of 'command and control' solutions via governmental regulation and coercion that might seem a priori to provide the best solutions to this class of problems are in fact inadvisable. Nonetheless, such collective action problems are still 'easy' in the sense defined above, because it is still true that the best solution to such problems becomes clear after careful empirical investigation.

In general, the best solution to easy collective action problems can generally be 'read off' from general features of the case given sufficient background knowledge of the relevant empirical facts. In contrast, there is nothing even resembling an empirical precedent for solving the collective action problems that give rise to anthropogenic climate change, and there is no way we could discover the best solution through further empirical investigation. Instead, the collective action problems associated with climate change are impervious to our standard tools for achieving social cooperation and avoiding tragic outcomes. In particular, we are unable to rely on benevolent human dispositions, bargaining, natural 'market-based' coordination, or governmental coercion to avoid a tragic outcome. That is because a climactic tragedy of the commons arises not only at the

[4] Elinor Ostrom, *Governing the Commons*.

level of individual people, but also at each higher level of collective abstraction, and there is no effective authority to provide the typical failsafe of governmental coercion. Furthermore, those who will suffer the most as a result of climate change are members of future generations who are incapable of influencing the current decisions that will determine their fate. Finally, ideology seems to preclude the possibility of any effective international agreement, because many negotiators refuse to consider any agreement that would make their nations worse off than the status quo, while others refuse to consider any agreement that isn't both fair and effective; this seems to make a stalemate inevitable, because it seems impossible for a single agreement to satisfy all of these constraints even in broad approximation.[5]

If this pessimistic view of our situation is correct, are we doomed?

We are not. Even when our standard tools for avoiding tragic outcomes are useless, we can often escape a social dilemma by turning the force of such a dilemma against itself.

[5] Perhaps an analogy to the kind of novel strategic challenge posed by our situation with respect to climate change is the situation faced by theorists who had to develop a nuclear weapons strategy during the Cold War. Derek Parfit provides a taxonomy of solutions to collective action problems and a classic discussion of solutions involving benevolent human dispositions in *Reasons and Persons*, Chapters 2-6; Ronald Coase provides a classic discussion of 'free-market' solutions in "The Problem of Social Cost"; Thomas Schelling provides a classic discussion of explicit bargaining in *The Strategy of Conflict*, and suggests such an approach to climate change in "What Makes Greenhouse Sense?"; Garrett Hardin provides a classic discussion of governmental coercion in "The Tragedy of the Commons"; Elinor Ostrom provides a taxonomy of solutions and a classic discussion of solutions involving natural coordination and bargaining, facilitated by appropriate background institutions in *Governing the Commons*. If the pessimistic assumptions outlined in this paper are correct, then Schelling's bargaining response to climate change would be unavailable, because according to the assumptions outlined below, an effective bargaining response would require the United States and other nations to act contrary to the interests of their current citizens, which would be impossible given those assumptions. In the final chapter of this dissertation, I show that Parfit's assumptions about morality and benevolence are mistaken by showing that morality and benevolence cannot be relied upon to solve collective action problems even in a world of normatively flawless agents.

The trick is to intentionally create a new social dilemma, the predictable outcome of which is a solution to the initial, otherwise insoluble dilemma. In more detail and applied to the problem of climate change, the idea is that we can intentionally create social dilemmas, which are situations in which we can predict that everyone will 'defect', thereby making everyone worse off than they would have been if everyone had chosen to 'cooperate' instead; furthermore, we can engineer and introduce international treaties that create such social dilemmas, the predictable outcome of which is that all nations will agree to the treaty, thereby making themselves worse off than they would have been if they had all chosen not to agree to the treaty instead; finally, we can engineer a climate treaty that has this 'tragic' flavor, but where the treaty makes the current generation of only a few high-emitter nations worse off as a mere side-effect of a mechanism within the treaty that effectively combats climate change. Such a treaty might be our only effective option if nations are disposed to reject or pull out of treaties whenever they make their current citizens worse off, and if it is impossible to effectively combat climate change without making the current citizens of a few influential nations worse off than the no-treaty status quo.

The motivation for such a treaty emerges most clearly from reflection on the following two claims, which constitute a worst-case description of our situation with respect to a climate treaty:

> No nation will agree to a treaty if agreeing would make their current citizens worse off than not agreeing. (Call this the *Realist Assumption*.)

> Any effective climate treaty requires the agreement of the United States, China, Brazil, and other high-emitter nations, and any effective treaty would make the current citizens of those nations worse off than the no-treaty status quo – i.e., worse off than they would be without the existence of a climate treaty. (Call this the *Current Generation Assumption*.)

From these two assumptions, it might seem to follow that the United States, China, Brazil, and other high-emitter nations will never agree to an effective treaty, and thus that there is no hope for an effective treaty. However, those conclusions do not really follow from these assumptions. That is because the introduction of a treaty can change what it is in the interest of nations to do, thereby making it the case that nations are better off by agreeing to the treaty than not agreeing, even if agreeing to the treaty makes nations worse off than the no-treaty status quo – i.e., worse off than they would have been if the treaty had never been introduced. Although this might seem paradoxical at first glance, denying it would imply that social dilemmas can never arise – but such dilemmas can arise, even at the level of nations, and the threat of such dilemmas provides part of the intellectual justification for global authorities like the World Trade Organization (WTO). For example, imagine that global trade rules did not exist except for the rule that duties could only be introduced by treaties open to all nation; if so, then it would be possible to create a social dilemma at the level of nations by introducing a treaty that imposed duties on non-signatory nations: given plausible assumptions, all nations would ultimately sign on to such a treaty in order to avoid having non-reciprocated duties imposed on their

exports, even though universal ratification of such a treaty would make all nations worse off than they would have been if the treaty never existed, because imposing symmetrical duties on all nations would have no real positive effect for any nation, but would have the effect of imposing additional economic inefficiency on all nations. This illustrates the way in which the introduction of a treaty can change what it is in the interest of nations to do, thereby making it the case that nations are better off by agreeing to the treaty than not agreeing, even though agreeing to the treaty makes nations worse off than the no-treaty status quo.

At first glance, a treaty that creates a social dilemma at the level of nations might seem necessarily irrational and undesirable. However, a treaty with this tragic flavor is exactly what we need if we face a situation like an intergenerational tragedy of the commons that is impervious to our standard tools for solving collective action problems, and that can only be solved by finding a novel way of changing our situation to make it the case that some influential nations would best serve the interests of their current citizens by acting in a way that makes those citizens worse off than the status quo. The only practical way of solving such a problem seems to be by creating something like a social dilemma at the level of nations, the predictable outcome of which is collective action that makes some nations worse off than the status quo.

To see how such a strategy might work in practice, imagine a treaty that has two components: a duty imposed on imports from nations that are not signatories to the treaty, where this duty is determined by the normalized emissions differential between signatory

nations and non-signatory nations,[6] and a cap and trade scheme among signatories to the treaty, where the cap decreases each year along a path that is insensitive to the number of signatories to the treaty.[7] Many nations would initially join such a treaty, in part because it would be to the advantage of many nations to impose an unreciprocated duty on imports from nations that are not initially signatories to the treaty, especially in light of the fact that the treaty would not require deep emissions reductions early on, but would only require significant reductions later as the cap decreased and the largest emitters joined the treaty, at which point the cost to early signatories would tend to be offset by payments from later-joining nations in exchange for emissions permits; furthermore, many other nations would initially join such a treaty out of a long-run concern for their own citizens and perhaps even the world, even if such a treaty was not to their short-run economic advantage. As nations joined the treaty, incentives would increase for other

[6] For ease of exposition, I use 'signatory' to refer to nations that have *ratified* the treaty. There are many ways of normalizing the emissions differential into units that allow 'apples-to-apples' comparison of the emissions differential between signatory and non-signatory nations: the simplest are either *by population*, or *by economic output* – i.e., tying duties to the difference in emissions *per person* in signatory nations versus non-signatory nations, or to the difference in emissions *per unit of GDP* in signatory nations versus non-signatory nations. Perhaps the most attractive approach would be a complex metric that takes both population and economic output into account, as well as other factors such as economic development and growth. Many ethicists would object to taking economic output into account on the grounds that fairness demands normalization by population only. However, these ethicists overlook the fact that normalizing by population only unfairly punishes nations for having a history of rational political and economic policies and thus lower birth rates, and unfairly rewards nations for having a history of irrational policies and thus higher birth rates. This objection is independent of objections that claim that equal per-capita metrics and allocations of emissions permits create perverse incentives, and therefore is not disarmed by effective replies to perverse incentives objections. Instead, this objection focuses on backward-looking considerations of fairness, and not on forward-looking consideration of incentives. For other good reasons not to normalize by population only (and not to distribute emissions permits on an equal per capita basis), see Eric Posner and David Weisbach, *Climate Change Justice*, Chapter Six, especially pp. 131-135.

[7] Such a treaty would also include an allocation scheme for emissions permits, any number of which would be consistent with the suggestions in this paper. Such a treaty might also include a mechanism to redistribute to developing nations some of the duties collected, in order to secure the universal compliance of developing nations and to offset the more ethically significant costs of the treaty to developing nations.

nations to join as well, because additional signatories would mean additional nations imposing duties on imports from non-signatory nations; in addition, as more nations joined the treaty and the cap decreased over time, the cap and trade scheme would become increasingly effective, thereby increasing the emissions differential between signatories and non-signatories, thereby increasing the magnitude of each individual duty imposed. As a result, even if the initial set of signatories was comparable to the limited set of nations that initially ratified the Kyoto Protocol, the treaty would set in motion a chain reaction that would make it in the interest of an increasing number of nations to join over time, eventually driving the costs of not joining so high that even the United States and China would ultimately find it in their interest to join rather than continue to hold out.

Of even greater importance, once nations joined such a treaty, it would never be in their interest to pull out, because the costs of pulling out always outweigh the costs of staying in after the point at which it is initially in a nation's interest to join, if the treaty has the novel 'self-enforcing' structure described above.[8] This solves the most serious problem for climate treaties, which is the problem of securing not merely initial agreement to the treaty, but more importantly *long-run participation and compliance*.[9] This problem is

[8] A treaty is *self-enforcing* in the relevant sense if and only if it is individually rational for each nation to maintain agreement to the treaty, and 'collectively rational' from the point of view of all nations. (For this definition, see Scott Barrett, *Environment and Statecraft*, pg. xiv, who adds the condition that the treaty must also be fair.)

[9] This self-enforcing structure is also consistent with other complementary incentive schemes to encourage compliance. For example, one promising addition would be for duties from all nations to be held by a single global administrator until the end of each year, at which point each nation's proceeds would be disbursed only if that nation complied with the treaty's provisions in the previous year; duties could then be subtracted from the accounts of non-compliant signatory nations based on their degree of non-compliance, with the proceeds distributed to compliant signatories.

insurmountable for more familiar treaty proposals. For example, even on the unrealistic assumption that the United States and China might initially to agree to a more familiar treaty that is costly enough to be effective, once the negative effects of such a treaty became salient to the populations and leaders of those nations, those nations would certainly pull out, resulting in yet another failed climate treaty and a decade or more of additional wasted time.[10] More importantly, the failure of a treaty that is costly enough to be effective under such circumstances would create deep hostility and resistance to future treaties, especially among nations that incurred significant costs under the failed treaty. For such reasons, it is likely that the world has only one shot at a climate treaty that is costly enough to be effective – and it is almost certain that more a familiar treaty proposal would waste that single opportunity, whereas a treaty with the self-enforcing structure described above would provide a realistic means to lock in universal participation and compliance over the long run.

The main hurdle for the particular treaty proposal described above is compliance with WTO regulations. Fortunately, there is clear precedent for its permissibility under WTO rules.[11] Perhaps even more importantly, it is clear that the WTO *should* permit such a

[10] In particular, a Republican president would almost certainly pull out of such a treaty if pulling out was much less costly to the US in the short run than staying in. (Although the US Constitution requires a two-thirds majority in the Senate and a Presidential signature for ratification, it allows a President to pull out of any treaty unilaterally without consulting either the Senate or the House of Representatives.)

[11] For example, see the discussion of 'border tax adjustments' and other instruments in Joost H. B. Pauwelyn, "U.S. Federal Climate Policy and Competitiveness Concerns: The Limits and Options of International Trade Law", available at http://www.nicholas.duke.edu/institute/internationaltradelaw.pdf, Frank Biermann and Rainer Brohm, "Implementing the Kyoto Protocol without the USA: the strategic role of energy tax adjustments at the border", and Roland Ismer and Karsten Neuhoff, "Border tax adjustment: a feasible way to support stringent emission trading". Precedent also establishes that initial non-signatories like the United States and China would not be permitted to levy reciprocal duties on signatory countries to cancel the effect of the signatories' duties, because the duties imposed by signatories would have protected

treaty upon its inevitable review, because, in addition to existing precedent, if the Realist Assumption and the Current Generation Assumption above are true, then allowing such a treaty is the only way to correct the most catastrophic case of market failure in history[12] – which means that allowing such a treaty is then the only way that the WTO can act in accord with its mission and treaty obligations of promoting ethical and economically efficient free trade.

In any event, the main point is not to defend the details of the particular treaty proposal described above, but rather to illustrate the general way in which a treaty can secure universal agreement and long-run compliance by incorporating the self-enforcing structure described above, which creates something like a social dilemma at the level of nations, the predictable outcome of which is universal ratification and adherence to the treaty over the long run, despite the fact that agreeing to the treaty makes some nations worse off than the no-treaty status quo. If the Realist Assumption and the Current Generation Assumption are roughly true, then a treaty with this structure is the only hope for an effective climate treaty, regardless of which of the existing architectures for a global response are chosen to fill in the details of the treaty. In this way, the current proposal solves the seemingly intractable problem of identifying a 'meta-architecture' that ensures cooperation and compliance to a chosen global response, while remaining

---

status as a result of their function to internalize externalities in a way that is permitted by existing WTO rules.

[12] "Climate change is the greatest market failure the world has ever seen…", *The Economics of Climate Change: The Stern Review*, pg. xviii.

relatively agnostic about the details and architecture of the global response that ought to be chosen.[13]

The novel self-enforcing structure described above is also the key to constructing the most *ethical* climate treaty possible, because it is the key to avoiding the ethical and practical failings of existing approaches to climate treaties. For example, the *idealist approach* identifies the climate treaty that it would be best to propose with the treaty that would be best from an ethical perspective if adopted.[14] The problem with this approach is that treaties that would be best from an ethical perspective if adopted seem to have no chance of being ratified and followed over the long run, given the actual dispositions of powerful nations. As a result, simple insistence on such treaties would ensure that the problem of climate change is never successfully addressed, which cannot be the best course of action even from an ethical perspective, because of the disastrous consequences

[13] For the distinction between 'architecture' and 'meta-architecture', see Jonathan Wiener "Incentives and meta-architecture", in Joseph Aldy and Robert Stavins (eds.) *Architectures for Agreement*, pg. 67. For discussion of alternative architectures for a global response, see *The Economics of Climate Change: The Stern Review*, Part IV and Part VI, the Intergovernmental Panel on Climate Change (IPCC) report *Climate Change 2007: Mitigation of Climate Change*, Chapter 13, Aldy and Stavins (eds.) *Post-Kyoto International Climate Policy*, and Aldy and Stavins (eds.) *Architectures for Agreement*.

[14] For example, see the various "allocation methods" for national emission targets and emissions trading surveyed in the IPCC report *Climate Change 2007: Mitigation of Climate Change*, pg. 770; in addition, almost all work by ethicists related to global responses to climate change falls under the rubric of the idealist approach. A possible exception is Peter Singer, who explicitly notes the problem with the idealist approach in "One Atmosphere", in Stephen Gardiner et. al. eds. *Climate Ethics*. However, because Singer's proposed cap and trade scheme, which allocates permits on an equal per capita basis, would likely cost the United States more than one hundred billion dollars per year and would be prohibitively costly to other large-emitter nations as well (see Eric Posner and David Weisbach, *Climate Change Justice*, pp. 123 and 212-213), a simple-minded insistence on Singer's proposal would be misguided for the same reason that idealist approaches are misguided in general: such an approach is not advisable even from an ethical perspective, because it has no chance of actual success, given the actual dispositions of powerful nations. The only promising way around such a problem would be to make an equal per capita allocation of permits part of a treaty that has something like the novel self-enforcing structure described above.

that would ensue. As a result, the idealist approach fails because it mistakenly equates the question of what we should do in our actual, imperfect circumstances with the question of what it would be best to do if we found ourselves in more ideal circumstances.

Another approach, the *hard-headed realist approach*, identifies the climate treaty that it would be best to propose with the treaty that would leave all nations at least as well off as the no-treaty status quo, and, subject to that constraint, would be best from an ethical perspective if adopted.[15] The problem with this approach is that it has unacceptable consequences of its own, because it implies that the treaty that it would be best to propose is a treaty on which poor nations that are vulnerable to climate change must make large transfer payments to the rich nations that are causing climate change in order to ensure that the rich nations are made no worse off by emissions reductions. Defenders of the hard-headed realist approach insist that this regrettable feature is necessary for an effective treaty, because they claim that there is no way that nations would agree to treaties that make themselves worse off than the no-treaty status quo. However, as the discussion above reveals, this idea is confused even from a rational choice perspective that takes the Realist Assumption for granted, because the introduction of a treaty can change what it is in the interest of nations to do, thereby making it the case that nations are better off by agreeing to the treaty than not agreeing, even if agreeing to the treaty

[15] For example, see Eric Posner and Cass Sunstein, "Climate Change Justice", especially pp. 1569-1570, Eric Posner and David Weisbach, *Climate Change Justice*, pp. 6, 86, and 143, Richard Stewart and Jonathan Wiener, *Reconstructing Climate Policy: Beyond Kyoto*, pp. 102-103, Jonathan Wiener, "Incentives and meta-architecture", in Aldy and Stavins (eds.) *Architectures for Agreement*, pp. 75-76.

makes them worse off than the no-treaty status quo.[16] As a result, the hard-headed realist approach fails because it implies that we should favor a treaty that is dramatically worse from an ethical perspective than other treaties that would be equally effective, even given the Realist Assumption – namely, well-crafted treaties with the self-enforcing structure described above.

The other main approach, the *incrementalist approach*, does not identify the treaty that it would be best to propose with a treaty that would actually be effective, but rather with a treaty that would serve merely as a starting point for future agreements to reduce emissions, with the hope that nations would, over time, accept incremental additions to the initial treaty, thereby gradually deepening the emissions-reduction effect of the initial agreement.[17] One problem with this approach is that it has already been tested over several decades, with little or no success. More importantly, the world now has only a few decades left to prevent the accumulation of potentially catastrophic levels of greenhouse gasses in the atmosphere, and it is implausible that the incrementalist approach could lead to sufficient emissions reductions within anything like that timeframe – in particular, it is implausible that such an approach could generate a reduction in global emissions of more than 50% from current levels, as is necessary

[16] More generally, the reasoning of hard-headed realists overlooks the fact that acts of coercion and influence can change what is in the interest of nations to do relative to the status quo. For further discussion, see Thomas Schelling, "An Essay on Bargaining" in *The Strategy of Conflict*.

[17] For example, see Axel Michaelowa, "Graduation and Deepening", and Jeffrey Frankel, "Formulas for quantitative emissions targets", both in Aldy and Stavins (eds.) *Architectures for Agreement*, and Robert Keohane and David Victor, "The Regime Complex for Climate Change", *Perspectives on Politics*.

merely to halt the accumulation of greenhouse gasses in the atmosphere.[18] Perhaps most importantly, the incrementalist approach is by its nature piecemeal, and would therefore fail to create a single global price for greenhouse gas emissions, which economists tend to agree is a necessary condition for any effective strategy, because a single global price for emissions is necessary not only to prevent 'emissions leakage' and to ensure that reductions are efficient, but also to create sufficient incentives for the kind of large-scale global investments in low-carbon energy and green technology that are necessary to reduce global emissions by more than 50% from current levels over the coming decades.[19] Because the incremental approach has no chance of creating such an unprecedented Archimedean lever of incentives, it has no real chance of being effective.

---

[18] From the authoritative Intergovernmental Panel on Climate Change (IPCC) report *Climate Change 2007: The Physical Science Basis*: "The concentration of greenhouse gas in the atmosphere depends on the competition between the rates of emission of the gas into the atmosphere and the rates of processes that remove it from the atmosphere. For example, carbon dioxide (CO2) is exchanged between the atmosphere, the ocean, and the land through processes such as atmosphere-ocean has transfer and chemical (e.g. weathering) and biological (e.g. photosynthesis) processes. While more than half of the CO2 emitted is currently removed from the atmosphere within a century, some fraction (about 20%) of emitted CO2 remains in the atmosphere for many millennia. Because of slow removal processes, atmospheric CO2 will continue to increase in the long term even if its emission is substantially reduced from present levels. … More specifically, the rate of emission of CO2 currently greatly exceeds its rate of removal, and the slow and incomplete removal implies that small to moderate reductions in its emissions would not result in stabilization of CO2 concentrations, but rather would only reduce the rate of its growth in coming decades. A 10% reduction in CO2 emissions would be expected to reduce the growth rate by 10%, while a 30% reduction in emissions would similarly reduce the growth rate of atmospheric CO2 concentrations by 30%. A 50% reduction would stabilize atmospheric CO2, but only for less than a decade. After that, atmospheric CO2 would be expected to rise again as the land and ocean sinks decline owing to well-known chemical and biological adjustments" (pp. 824-825). For further ramifications for greenhouse gas stabilization levels, see *The Economics of Climate Change: The Stern Review*, Chapter 8; for a summary, see Hal Harvey and Sonia Aggarwal, "The Costs of Delay", ClimateWorks Foundation.

[19] *The Economics of Climate Change: The Stern Review*, Part IV and Chapter 22; William Nordhaus, *A Question of Balance*, pg. 29: "…placing a near-universal and harmonized price or tax on carbon is a necessary and perhaps even a sufficient condition for reducing the future threat of global warming".

In light of the problems with these existing approaches to climate treaties, the possibility of a treaty with the self-enforcing structure described above holds the key to the most ethical climate treaty possible in our imperfect circumstances, and transforms the project of crafting such a treaty into a well-defined engineering problem that can be readily solved by experts on the relevant economic mechanisms and dispositions of nations. The primary constraint on such a treaty is that it must predictably lead to universal ratification and compliance in the long run via the cascade of economic incentives described above. Subject to that constraint, the treaty should be designed to distribute the costs of emissions reductions as fairly as possible.

At first glance, it might seem that the devil is in the details regarding fairness. However, on further reflection, the details are relatively uncontroversial insofar as they are relevant to the engineering problem of creating the basic structural features of the treaty, because the basic structural features can only ensure fairness within a range, letting the ultimate details fall where they may in the inevitable round of political adjustment before the treaty is formally introduced for ratification – and it is easy to make accurate comparative judgments of the fairness of treaty proposals whenever one proposal is within the general range of fairness while the other falls outside that range. As a result, the most important consideration is to ensure that the basic structure of the treaty is crafted by experts with good ethical and practical judgment, and that such experts design the treaty to ensure fairness subject only to the constraint of feasibility. Fortunately, the self-enforcing structure described above is a significant advantage even here, because it ensures that an effective treaty can be developed by a team of independent experts motivated only by a concern for fairness and efficiency, with no need to accommodate myriad demands from

national representatives and others in order to ensure universal participation and compliance in the long run once the treaty is introduced.

In sum, when the novel approach to solving collective action problems outlined in this chapter is available, it allows us to bring about solutions to collective action problems that cannot be achieved by any other means. This is especially important when we face an intergenerational tragedy of the commons, or some other situation in which we have decisive reason to do something that we are unable to do using any of the familiar approaches to collective action. More generally, the result of this chapter is a straightforward, realistic, and non-coercive framework for constructing economically and ethically optimal solutions to many previously intractable collective action problems. In such cases, the approach outlined above is often the best approach possible in our imperfect world.[20]

[20] For example, on a relatively small scale, the collective action problems that stand in the way of a coordinated global aviation fuel tax and a coordinated global financial instruments trading tax are examples of otherwise intractable problems that are amenable to a similar treatment. On a much larger scale, runaway deficit spending by nations that benefits current citizens at an unacceptable cost to future generations and at an unacceptable risk to the current generation of other nations is an example of another serious intergenerational tragedy of the commons that may be amenable to a similar treatment.

Strategy for Introducing a Climate Treaty

Any climate treaty should be introduced under only the most favorable conditions that can be realistically expected, especially if the world has only one shot at a climate treaty that is costly enough to be effective. As a result, any climate treaty should be introduced only if the following two conditions are satisfied:

> (1) The treaty should be introduced early in the term of a US President who supports the treaty, enabling him or her to sign the treaty and publicly endorse its permissibility under all international laws and treaties, and to use his or her power and influence to ensure that international court decisions establish as *precedent* that the treaty is permissible. This will *institutionalize* the permissibility of the treaty in a way that cannot be reversed by anything short of the most dramatic (and hence unlikely) power politics. The US President can do all of this unilaterally even if the US Senate is initially disposed to reject the treaty unanimously, and even if powerful nations such as China oppose the treaty.

(2) Upon introduction, the treaty should be ratified by a substantial proportion of developed nations, as well as a substantial proportion of developing nations.

The possibility of a treaty with the self-enforcing structure described above is the key to satisfying both conditions (1) and (2), because the structure of such a treaty ensures that it would actually be effective, which means that nations will recognize that the costs of the treaty would be non-futile, including the political costs to leaders who support the treaty. This increases the likelihood that (1) will be satisfied relative to any particular US President, and the design of such a treaty ensures satisfaction of condition (2), partly for economic reasons, and partly because many nations are already disposed to ratify such a treaty out of a long-run concern for their own citizens and the world, as demonstrated by the nations that ratified the Kyoto Protocol.[21]

Furthermore, if both (1) and (2) are satisfied, then other strategic challenges do not arise. For example, with respect to the detailed treaty proposal discussed above, on the realistic assumption that (1) is satisfied by a supportive US President and that (2) is satisfied by an initial coalition that includes most European nations, Japan, Russia, and other nations that ratified the Kyoto Protocol, it is unrealistic to think that China would have the power and influence to convince international courts to rule against both the US administration and

[21] It is worth noting that even the flawed and ineffective Kyoto Protocol was ratified by an impressive set of nations, and was signed by US President Bill Clinton. As a result, it is realistic to think that a superior treaty with the self-enforcing structure described above would ensure satisfaction of (2), and that a US President might well be willing to expend the political capital necessary to satisfy condition (1) relative to such a treaty. (If that is not immediately clear, it might help to imagine a US President such as Al Gore.)

that coalition regarding the permissibility of the treaty under WTO rules, and it is thus unrealistic to think that China could retaliate against signatory nations by imposing reciprocal duties on imports, because a favorable WTO ruling would mean the impermissibility of such reciprocal duties.

As a result, it is realistic to think that a treaty with the self-enforcing structure described above would succeed if introduced under favorable conditions, even given maximally pessimistic assumptions about opposition to the treaty from China and the US Senate.[22]

[22] Thanks to Robert Keohane for encouraging me to discuss these issues.

Part Two: What Should You Do?

Suppose that [a] person proposes regulating the fuel efficiency of cars, or requiring high-efficiency lightbulbs, or subsidizing ethanol, or providing research support for solar power – but nowhere does the proposal raise the price of carbon. You should conclude that the proposal is not really serious and does not recognize the central economic message about how to slow climate change. To a first approximation, raising the price of carbon is a necessary and sufficient step for tackling global warming. The rest is at best rhetoric and may actually be harmful in inducing economic inefficiencies.

William Nordhaus

Climate Change and the Ethical Significance of Futility

1. The Intuitive Argument

Climate change may be the most serious problem facing humanity today. If we do nothing about it, billions of people will almost certainly suffer in one way or another, and a worldwide catastrophe may ensue. In short, climate change is a serious problem.

However, it doesn't follow that you are required to do anything about it. To see why, suppose for a moment that you knew that nothing could be done about it; then you certainly wouldn't be required to do anything about it. For example, if there were a giant asteroid headed toward the earth that was sure to kill us all, that would count as a serious problem, but if you knew that nothing could be done about it, you wouldn't be required to do anything about it. So, if there is a serious problem, it doesn't follow that you are required to do anything about it.[23]

Of course, there are important differences between the asteroid story just described and our situation with respect to climate change. One difference is that we *can* do something about climate change, whereas in the asteroid case just described there is nothing we can

[23] Here and in the rest of this paper by 'required' I mean 'ethically/morally required'.

do about the threat. Does this difference alone make for a difference in what individuals are required to do about climate change?

To investigate this, consider a different version of the asteroid story. This time, we discover that a giant asteroid is headed toward earth and will kill us all if we do nothing, but this time we can do something to stop it: we can stop it if and only if we all cooperate and do something together. In particular, suppose that we will avoid destruction from the asteroid if and only if we all congregate in Kansas and fire our ray guns into the sky in unison; if some of us are absent, the collected beam from our guns will be inadequate to destroy the asteroid and it will kill us all; but if we are all there, the collected beam will be just powerful enough to vaporize the asteroid.

In this case, it seems like you would be required to go to Kansas and fire your ray gun *if* you had good reason to think that everyone else would do the same thing. However, if it was clear that not nearly enough people would show up, then you would not be required to sacrifice your remaining days on earth by taking a futile trip to Kansas.[24] This shows that even if we can collectively do something about a serious problem, it does not follow that you as an individual are required to do anything about it, because you might know that although we can collectively do something about it, in fact we won't.

[24] If it isn't immediately clear that you wouldn't be required to make the trip to Kansas, focus on the fact that if you went to Kansas, most of your friends, neighbors, extended family, and other people would remain in their hometowns enjoying their remaining time together with their friends and immediate families, while you'd spend your remaining days on an unpleasant journey filled with traffic congestion and uncomfortable anguish, which would ultimately be shared with strangers in a strange place – and all for nothing, since you know that your going to Kansas has no chance of making any positive difference for anyone.

There are further important differences between the last asteroid story and our situation with respect to climate change. For example, in the last version of the asteroid story we assumed that the threat from the asteroid was a natural threat, whereas in the real world the threat from climate change is a threat that *we have caused*; in addition, in the last story we imagined that an individual going to Kansas would have no positive effect at all, whereas in the real world you know that reducing your greenhouse gas emissions *will have some very small effect* on the level of such gasses in the atmosphere. Do these differences make for a difference in what you are required to do?

Suppose again that a giant asteroid will kill us all if we do nothing. As before, we can save our planet if and only if we all cooperate, but it is clear that we won't all cooperate, and so our planet is doomed. However, this time the destruction of the Earth will be *delayed* based on the number of people who show up in Kansas and fire their ray guns in unison at the asteroid. In particular, for every person who shows up and fires their gun from Kansas, the destruction of the Earth will be delayed by 1/1,000 of a second.[25] In addition, suppose that the only reason we are threatened by the asteroid is that one of our mineral-extracting expeditions to the asteroid knocked it into its threatening orbit, and we do not have time to deflect it using any other means.

Do these changes make any difference to what you as an individual are required to do? They do not. It remains true that if you know that only a small percentage of people will show up in Kansas, thereby sealing our fate, then you aren't required to sacrifice your

[25] Why can we delay the end of the world in this way? Imagine that this is our only opportunity to influence the orbit of the asteroid, and that each shot will result in the asteroid taking a slightly wider orbit, thereby very slightly delaying the moment when it will return, unstoppable, to destroy the Earth.

remaining days on earth by taking a trip to join them. If this isn't immediately clear, focus on the fact that if you went to Kansas, most of your friends, neighbors, extended family, and others would remain at home, enjoying much of their remaining time on Earth with their friends and families, while you'd spend the next week on an unpaid leave from work, on a costly and unpleasant journey filled with traffic congestion and uncomfortable anguish, which would culminate in a bizarre scene in a strange place, surrounded by strange people – and all of that merely to delay the inevitable end of everything by 1/1,000 of a second. Then, after the depressing ride home, many would ask you why you wasted so much of your remaining time and money in such a foolish way. Could you really be required to incur such significant costs to benefit the rest of the world in such a trivial way, given that most of the rest of the world is not willing to do the same? No. In such a situation it would be permissible for you, like the others, to skip the trip to Kansas and spend your time doing something more valuable to you and your family instead.

This is the basis for my intuitive argument that you, as an individual, are not required to reduce your greenhouse gas emissions by a significant amount: just as you wouldn't be required to make the trip to Kansas in the story above, so too you aren't required to reduce your emissions by a significant amount. In more detail: You know that billions of people in China, India, and the rest of the world don't care about climate change and won't care until it is too late; so, you know that the non-cooperation of others will lead to a bad, non-cooperative outcome regardless of your own cooperation, and you know that if you cooperate it will not have any tangible effect on the outcome, except to make it

worse for you; therefore, since there are no other ethically significant factors that differ from those in the story above, you are not required to cooperate.

This is my intuitive argument. In the next sections I will reply to objections and make the argument more rigorous. However, before doing that, I will outline some of the other conclusions I will ultimately reach, so that no unnecessary confusion arises about what I take my arguments to show.

In particular, although I am arguing that *as things stand now* individuals are not required to reduce their emissions by a substantial amount, this is perfectly consistent with the idea, which I also endorse, that in the future individuals would be required to reduce emissions if the rest of the world somehow managed to cooperate and reduce emissions in a way that had some hope of making an important difference to the outcome. These two claims are perfectly consistent, because there is an important distinction between, on the one hand, what individuals are required to do in a collective action problem in which almost everyone else is *defecting*, and, on the other hand, what individuals are required to do in an otherwise similar collective action problem in which others are *cooperating*. Refusing to cooperate when almost everyone else is defecting is refusing to be a *sucker*, whereas refusing to cooperate when almost everyone else is cooperating is *free-riding*. The view that you are not required to be a sucker is perfectly consistent with the view that free-riding would be impermissible in an otherwise similar case in which almost everyone is cooperating.

In addition, I am not arguing that *we* shouldn't do anything about climate change. Instead, I am arguing that *individuals* are typically not required to reduce their emissions

by any significant amount, which is perfectly consistent with the idea that *we* are *collectively* required to reduce emissions. As a result, my arguments do not have the consequence that we shouldn't do anything about climate change. Instead, my claim is merely that as things stand now, individuals are typically not required to reduce emissions by a significant amount.

This view might seem incoherent, because it might be seem that if we are required to reduce our emissions collectively, then it follows that each of us is required to reduce our emissions individually. However, that does not follow, because the ethically relevant properties of a group are often different from those of its members. In particular, a group of people Φ-ing typically has more significant Φ-effects than an individual Φ-ing; as a result, in a situation where Φ-effects are ethically important, it might be impermissible for the group to Φ, but permissible for each member of the group to Φ.

For example, imagine that we find ourselves in the middle of a stampede. If the stampede continues, it is clear that an increasing number of innocent people will be killed. Luckily, each of us has a button in our hands: once more than 50% of us press our buttons, a painless 'sleep' device in all of our spines will be activated, causing the stampede to stop peacefully. Other than pressing the button and then continuing to stampede, the only other options we have are to continue stampeding without pressing the button, or else to stop stampeding and certainly be run over and killed. In this case, the thing for each of us to do is to press our button and then continue stampeding, hoping that enough others press their buttons to end the stampede.

So, suppose you press your button, and the stampede continues; this means that less than a sufficient number of the others have pressed their buttons. You know that the stampede will stop once a sufficient number do press their buttons, but now, in the meantime, the only relevant options that you have are to continue stampeding and live, or else stop stampeding and die, given that you have already pressed your button.

In this case, you are permitted to continue stampeding, and so is every other individual; at the same time, we are collectively required to stop stampeding. This shows that the objection to my view under consideration fails, because individual requirements do not follow from collective requirements.

Perhaps more importantly, in the case just described, individuals are not required to *stop stampeding*, although individuals are required to *press their button*. Analogous remarks apply to our situation with respect to climate change. On my view, individuals are not required to *reduce emissions by a substantial amount*, although individuals are required to *favor effective intergovernmental solutions to climate change*.[26] It is crucial to distinguish the second political act from the more direct act of reducing one's own emissions by a significant amount, just as in the stampede story it is crucial to distinguish the more political act of pressing your button from the more direct act of stopping stampeding. With this distinction in mind, I agree that even individuals are required do

[26] When I say 'individuals are required to favor effective intergovernmental solutions to climate change', this should be understood as elliptical for 'individuals are required to favor effective intergovernmental solutions to climate change *that are not dramatically inferior to other realistic and effective solutions that we could adopt instead*'. I argue that such solutions are realistic despite the collective action problems that arise at the level of nations in the first part of this dissertation. Here and in the rest of the paper, I use 'S is required to favor p' as shorthand for something like 'If S were idealized in ethically relevant ways, S would be required to be in favor of p occurring in S's actual situation, even if S is not actually required to recognize or vocalize this fact, etc.'.

something about climate change; however, that something is different than it might initially appear – the thing that individuals are required to do about climate change is *political*, and not directly tied to personal emissions.

Is this an important difference? Does it make a difference in practice?

It does. To begin to see why, consider a controversial remark by Barack Obama that might initially seem confused. During a 2008 debate, Brian Williams asked Obama what he was *personally* doing about climate change in his day-to-day life. Obama later described Williams's question as "stupid", telling his staff: "What I'm thinking in my head is, 'Well, the truth is, Brian, we can't solve global warming because I fucking changed light bulbs in my house. It's because of something collective'."[27]

The explicit content of Obama's statement is true: climate change is caused by "something collective", and so it cannot be solved by one man changing light bulbs in his house. However, Obama's point is undoubtedly something further. At first glance, it might seem that Obama is arguing that he isn't required to change light bulbs in his house because a person isn't required to make a contribution toward solving a problem if that contribution won't solve the problem all by itself. However, this is not a charitable way to interpret Obama's reasoning, because the stampede case above shows that such a view is clearly mistaken: in that case, you pressing your button cannot solve the problem all by itself; nonetheless, you are required to press your button.

[27] http://www.huffingtonpost.com/2008/11/05/obama-we-cant-solve-globa_n_141407.html

A better interpretation is that Obama's view is essentially the same as the view I am defending here: there is something that we as individuals are required to do about climate change, but it isn't to change light bulbs in our house, or to do other costly emissions-reducing actions at a personal level. Instead, each of us is required to favor and not impede promising political solutions to the problem of climate change because such solutions are the only ones that have any real prospect for success; furthermore, insofar as we are inclined to go above and beyond what we are merely required to do, we as individuals should focus our energy on promoting such solutions and encouraging others to do the same, rather than focusing our efforts on changing light bulbs in our houses and other individual-level actions that are not an important part of a genuine solution to the problem. Williams's question mistakenly assumes that climate change is a problem to be solved by the direct action of individuals motivated by the recognition of individual obligations to reduce emissions, when in fact individuals have no such obligations, and would do better to focus on promoting collective solutions to climate change instead.

If this is Obama's view, then why doesn't he explicitly say it? Part of the explanation might be that the mistake made by Williams is subtle and could only be demonstrated by lengthy arguments such as those in this chapter. But a more obvious and more tragic part of the explanation is that most left-leaning Americans have already concluded that Williams's way of thinking about the ethics of climate change is correct, and they reject without examination anything inconsistent with that view. As a result, Democratic politicians are afraid to question that view publicly, and instead ensure that their policies are consistent with and reward those who hold such views. This is tragic because if it continues, it will ensure that the problem of climate change is never successfully

addressed. That is because if left-leaning Americans continue to believe that climate change is a problem to be solved at an individual level by "doing the right thing", and continue to believe that they are "doing their part" by changing light bulbs, etc., then the collective upshot will be that left-leaning Americans fail to understand and fail to support a real solution to the problem (namely, a dramatic intergovernmental solution that creates a single global price for emissions of carbon dioxide and equivalent greenhouse gasses), and instead continue to direct their limited political capital toward bogus solutions that conform to their mistaken view about the nature of the problem and the nature of effective solutions to the problem[28] – and if left-leaning Americans never demand a dramatic intergovernmental solution to the problem, then there will never be sufficient support for such a solution in America; and if there is never sufficient support for such a solution in America, then much of the rest of the world will recognize that dramatic action on their part would be futile, thereby ensuring that the problem of climate change is never successfully addressed.[29]

---

[28] As William Nordhaus observes, "Suppose that [a] person proposes regulating the fuel efficiency of cars, or requiring high-efficiency lightbulbs, or subsidizing ethanol, or providing research support for solar power – but nowhere does the proposal raise the price of carbon. You should conclude that the proposal is not really serious and does not recognize the central economic message about how to slow climate change. To a first approximation, raising the price of carbon is a necessary and sufficient step for tackling global warming. The rest is at best rhetoric and may actually be harmful in inducing economic inefficiencies. … Although other measures might usefully buttress this policy, placing a near-universal and harmonized price or tax on carbon is a necessary and perhaps even a sufficient condition for reducing the future threat of global warming" (*A Question of Balance*, pg. 22 and pg. 29). See also *The Economics of Climate Change: The Stern Review*, Section IV; for example, "…a common price signal is needed across countries and sectors to ensure that emissions reductions are delivered in the most cost-effective way" (pg. 352).

[29] This reasoning is not undermined by opinion polls, some of which claim to show that left-leaning Americans support some form of collective action on climate change. For one thing, supporting some form of collective action is not the same thing as supporting *dramatic* collective action: in particular, the relevant polls provide some reason to think that left-leaning Americans support a token incentive scheme (which would have no important effect), but do not provide any reason to think that left-leaning Americans support

If anything like this is true, then the arguments in this paper have some practical importance, because they can help us avoid inaccurate and counterproductive views about what individuals are required to about climate change.[30]

---

the more dramatic action that would be necessary to change the ultimate outcome in any meaningful way. Furthermore, nothing dramatic will happen unless left-leaning Americans *demand* such action, and there is no chance of that happening as long as they continue to conceptualize the problem as one to be solved largely by individuals "doing the right thing" and "doing their part" by changing light bulbs in their houses, etc.

In any event, the results of climate surveys vary widely because of massive framing effects that depend on whether the questions are framed in a way that emphasizes *costs*, or instead emphasizes *environmental benefits*; as a result, individual polls provide no real evidence for conclusions about individuals' 'real opinions' on climate policy. In fact, reflection on the dramatic nature of the framing effects suggests there is probably no such thing as American's 'real opinions' about climate policy. (For some related discussion, see Cass Sunstein, *Laws of Fear*, pg. 48.) An additional consideration is that left-leaning Americans tend to demand action to counteract any substantial increases in gas prices, which provides more reliable evidence about their true preferences regarding policy changes that essentially involve such consequences.

Although climate change will not be successfully addressed until left-leaning Americans demand dramatic action, that does not mean that individuals are required to demand such action as things stand now. Instead, individuals are required to favor demanding such action, and perhaps should also stand ready to demand such action conditional on a sufficient number of other people demanding such action – but if few others are demanding such action, an individual is not required to take to the streets and demand it all by him or herself. Participating in the early stages of an important movement is generally a supererogatory act – and participating in the later stages is often a supererogatory act as well, depending on the features of the case.

[30] Compare the following from Peter Railton: "By altering social and political arrangements we can lessen the disruptiveness of moral demands on our lives, and in the long run achieve better results than freelance good-doing. A consequentialist theory is therefore likely to recommend that accepting negative responsibility is more a matter of supporting certain social and political arrangements (or rearrangements) than of setting out individually to save the world. Moreover, it is clear that such social and political changes cannot be made unless the lives of individuals are psychologically supportable in the meanwhile, and this provides substantial reason for rejecting the notion that we should abandon all that matters to us as individuals and devote ourselves solely to net social welfare", "Alienation, Consequentialism, and the Demands of Morality" in *Facts, Values, and Norms*, pg. 172.

2. Objections and Replies

In the preceding section I offered an intuitive argument, based on an analogy, for the conclusion that you are not required to reduce your emissions by a substantial amount. My argument might initially seem like it has to be mistaken for one reason or another.

For example, it might seem that my argument has to be mistaken because it overgeneralizes and implies that it is always permissible to pollute, no matter how wasteful or egregious the pollution. However, my argument does not have that implication. Instead, my argument depends essentially on the fact that when reasonable people go about their daily lives, they cause emissions that *do not tangibly harm anyone*, even when repeated over many years, and that *could be avoided only by making significant sacrifices of personal and familial well-being*; furthermore, these emissions occur in a context in which it is clear that *almost everyone else would be unwilling to sacrifice such emissions if they were in similar circumstances*. The claim that emissions that meet all of these conditions are permissible does not in any way imply that wasteful or egregious pollution is permissible. For example, wasteful emissions are emissions that could be avoided *without significant sacrifices of well-being*, and so my argument does not apply to such emissions; similarly, egregious pollution is pollution that *tangibly harms others*, and so my argument does not apply to such pollution. For these reasons, my argument does not overgeneralize and lead to absurd conclusions. Instead, my argument points the way toward a carefully reasoned basis for condemning wasteful and

egregious pollution, while allowing that reasonable emissions are often permissible – something that previous accounts have failed to do.[31]

A different objection might be raised by theorists who believe that individual requirements are determined by a principle of *universalizability*. These theorists are likely to argue that my conclusions must be mistaken because if everyone fails to reduce emissions by a substantial amount, then a disaster will ensue, which (these theorists claim) makes it the case that everyone is required to reduce emissions by a substantial amount.

The problem with this reasoning is that it leads to the wrong verdict on the asteroid case and stampede case in the previous section and many other cases involving collective action problems in which most people are 'defecting'. For example, in the asteroid case it is also true that a disaster will ensue if everyone fails to go to Kansas; however, it does not follow that you are required to go to Kansas.

[31] Existing accounts do not plausibly distinguish between permissible and impermissible emissions. For example, John Broome's deontological account (discussed below) seems to overgeneralize and yield the verdict that emissions are never permissible. That is because his premises that *emissions cause harm* and *it is wrong to cause harm for one's own benefit* seem to imply that emissions are never permissible. In reply to this objection, the second premise could be restricted in some way, but it is unclear how that could be done in a way that resulted in both a plausible premise and a sound basis for distinguishing permissible and impermissible emissions. For example, Broome could claim that *it is wrong to cause harm for one's own benefit unless it is necessary for survival* – but then his argument would imply that you are required to reduce your emissions to a bare subsistence level, which is not plausible, especially in light of my discussion in this paper. These considerations show that distinguishing permissible and impermissible emissions is a difficult task that requires careful argument and reflection. Somewhat relatedly, Henry Shue has noted that emissions that are necessary for survival are undoubtedly permissible, and must in effect be guaranteed to each person on the planet as part of any ethically permissible intergovernmental emissions permit allocation scheme. Although true, this does not provide a plausible general distinction between permissible and impermissible emissions, because even if everyone is given a right to emissions that are necessary for survival, there is still the further crucial question of what *additional* emissions would be permissible – and it is obvious that some further emissions would be permissible ("Subsistence Emissions and Luxury Emissions", in Stephen Gardiner et. al. (ed.) *Climate Ethics*.)

In general, universalizability, Kantian, and contractualist theories are either false because they deliver the wrong verdict on cases like the asteroid case, or else, if they are understood in such a way that they do not deliver the wrong verdict on such cases, they provide no reason to think that the conclusions of my arguments are mistaken. As a result, there is no good objection to my arguments that arises from such theories. Instead, the asteroid case and stampede case show that there is an interesting range of collective action situations in which ethical actors are 'thrown into' situations in which a bad outcome looms in the background through no fault of their own, and in a way that makes non-universalizable action permissible. Our actual situation with respect to climate change is a particularly dramatic and regrettable collective action situation of this type. (I discuss these issues related to universalizability further in the final chapter of this dissertation.)

It might still seem that there is a crucial disanalogy between the asteroid case and our actual case with respect to climate change, because in the asteroid case by not going to Kansas you would merely *allow* harm to befall others, whereas by continuing to emit greenhouse gasses you would *actively* harm others. John Broome offers a clear statement of such an argument:

> …almost everyone recognizes (with some exceptions) the elementary moral principle that you should not do something for your own benefit if it harms another person. ... Climate change will cause harm. … In going about our daily lives, each of us causes greenhouse gasses to be emitted.

> … In that way, what we each do for our own benefit harms others. Perhaps at the moment we cannot help it, and in the past we did not realize we were doing it. But the elementary moral principle I mentioned tells us we should try to stop doing it and compensate the people we harm.[32]

As Broome notes, there are "some exceptions" to the principle with which he begins; however, he does not notice that these exceptions undermine his argument. To see why, imagine that there is a single power plant in your region, which emits water vapor as its only by-product, and which as a result causes a snowstorm about once per year in the lesser-developed nation that borders your nation. In such a case, are you required to make costly reductions in your consumption of power, simply because your consumption of power makes an intangible causal contribution to harm via the snowstorms that are caused by the power plant? In answering this question, it is important to keep in mind that snowstorms cause significant harm to people – for example, they cause transportation accidents in which people are harmed and killed, they cause power disruptions that can have substantial harmful effects, and they cause other undesirable things to happen. However, at the same time, even if such significant harms make it the case that the power plant's customers must pay *compensation* to those who are negatively impacted by the snowstorms, that does not settle the question of whether you are *required to reduce your consumption of power*. With that in mind, could you be required, at

[32] John Broome, "The Ethics of Climate Change", pg. 69. The context makes it clear that Broome is correctly interpreted as arguing that *individuals* are required to reduce emissions. I raise additional objections to this argument in the preceding footnote and in the following footnote.

significant cost to yourself, to reduce your power consumption simply because the power plant in your region occasionally causes a snowstorm?

To make the case more concrete, suppose that if you continue to consume electricity from the power plant, you will cause a tiny amount of water vapor to be released which would not otherwise be released, where this additional vapor will have no effect on the once-per-year snowstorm that is caused by the power plant, except to cause a few extra flurries to fly. Given that assumption, it seems clear that you are not required to reduce your power consumption by a costly amount, even though you know that your power consumption plays a direct causal role in significantly harming some people in the adjacent country. This shows that Broome's principle fails to apply in just the sort of cases that are relevant to the ethics of climate change, because it shows that when the causal contribution of your action is negligible, and the cost to you of avoiding that action is substantial, it does not follow from any "elementary moral principle" that you must avoid that action, even if your action plays a direct causal role in harming others.[33]

[33] Unlike previous responses to arguments like Broome's, my response does not depend on denying that emissions play a *direct causal role* in significantly harming people. (For a response that depends on such a denial, see Walter Sinnott-Armstrong, "It's Not My Fault: Global Warming and Individual Moral Obligations", in Stephen Gardiner et. al. (ed.) *Climate Ethics*. Sinnott-Armstrong's ultimate conclusions are very similar to mine, although his arguments for those conclusions are different.) Another problem for harm-based arguments like Broome's is that they seem to imply that seizing property via eminent domain is impermissible in a wide range of cases in which it is in fact permissible. This is relevant to the current discussion because it might seem, for example, that Bangladesh will have to be inundated with water for the greater good of others *in a way that is relevantly similar to other justified uses of eminent domain*, which means that it may be permissible for us *collectively* to act in a way that we know will inundate Bangladesh with water as long as we take sufficiently dramatic steps to resettle and compensate Bangladeshis. If that is right, then even some issues about what we are *collectively* required to do might come out differently than ethicists typically assume. I do not mean to endorse this line of reasoning – I merely intend to point out its significance, and the fact that it does not involve any *obvious* mistake. (Thanks to Gideon Rosen for noting the connection to eminent domain– I don't know that he'd favor any aspect of its consideration here.)

This leads to a reply to another objection – namely, that my arguments must be mistaken because they ignore the fact that you will *violate other people's rights* unless you reduce your emissions. The snowstorm case shows that this objection is mistaken, because in that case the people who are harmed as a result of the relevant snowstorms have rights, but you do not violate anyone's rights by continuing to consume electricity from the power plant – and you certainly do not act impermissibly.

The upshot is that it is sometimes permissible to play a causal role in harming others even when there are no dramatic ethical reasons for doing so. This is true for the general reason that your own interests are ethically significant, just like everyone else's interests, and so in some cases where your own interests are *fully at stake* they outweigh the interests of others that are only *intangibly at stake* – and this can be true even when your own interests do not constitute dramatic ethical reasons.[34]

Once we've identified this phenomenon, it is easy to find more cases to illustrate it. For example, imagine that you must drive to work, and that the quickest route is a well-traveled road that passes by another person's house. If you take this route, you will often splash some water onto his property, which will cause some intangible amount of erosion of the foundation of his house, which is near the road. It is clear that over twenty years the collective action of everyone who drives on the road will certainly undermine the

[34] Analogous reasoning suggests that rights infringement can sometimes be permissible even when there are no dramatic ethical reasons for that infringement. For example, imagine that you buy a snow-making machine to make snow for your children in your backyard. You know that when you operate it, a few flurries (but no more than a few flurries) will fly over onto your neighbor's land and will even occasionally fall on his head without his consent. Do you infringe his rights by operating the snow machine? Perhaps – but the infringement is so minimal that it cannot possibly make it impermissible for you to operate the snow-making machine.

foundations of his house; however, it is also clear that your driving on the road will not make any tangible difference, even over twenty years. In such a case, if it would be costly to you to take another route, then it is permissible for you to drive on the road every day, even though you will thereby play a causal role in harming this person by undermining the foundation of his house.[35]

Another objection might arise from the idea that you *owe it to others* to reduce your emissions by a substantial amount, even if you do not violate anyone's rights by failing to reduce your emissions. To see the problem with this objection, note that in the snowstorm case discussed above, although it is plausible that you owe it to others to *compensate them* to an appropriate extent if they are harmed by a snowstorm caused by your power plant, it is not plausible that you owe it to them *not to use power from the power plant*. Similarly, even if it is plausible that you owe it to others to *compensate them* to an appropriate extent if they are harmed by climate change caused by global emissions, it is not plausible that you owe it to them to *reduce emissions that have an important effect on your personal and familial well-being*. Although such a combination of claims can seem unintuitive in the abstract, consideration of the snowstorm case and other cases above indicates that it is the correct thing to say upon reflection.[36]

---

[35] If necessary, suppose that the road is located on land that is unowned and uncontrolled by anyone or any agency, whereas the man's house is located on property that he owns.

[36] It might seem that you are required to pay *compensation* to those who are harmed in such cases, even if you are not required to avoid playing a role in causing the harm. However, even this is not necessarily true. For example, imagine that the power plant from which you get your power is owned by your government, and that your government refuses to compensate those who are harmed by the snowstorms that are caused by the power plant. Are you then personally required to compensate those who are harmed? What if the costs to you of providing the compensation are much larger than the amount of compensation that you owe? What if it is not your fault that your government decided to build such a power plant rather than a

Another objection arises from arguments like the following:

> If we do nothing about climate change, then at least millions of people will die; so, unless you reduce your emissions by a substantial amount, your proportional share of the harm that is done will be very great. Therefore, you are required to reduce your emissions by a substantial amount.

This argument commits the *Average Effects Fallacy*, which is the fallacy of equating the ethically relevant effects of an act with the average effects of all of the actual acts of that type.[37] To see why this is a fallacy, imagine that you are leaving a stadium after an event. As you are about to cross the street and walk toward the parking garage, you hear an ambulance. You initially stop and are inclined to wait for the ambulance if other people do, but you see hundreds of other people cross the street in front of the ambulance, and

---

different power plant that would not harm anyone? Reflection on such questions reveals that if the costs to you are significant enough and the benefit to others is small enough, it can even be true that you as an individual are not required to compensate others for the harm you knowingly play a role in causing in pursuit of your self-interested goals.

[37] One version of this fallacy is simply to equate the actual effects of an act with the average effects of actual acts of that type. Many forms of carbon footprint reasoning (lifecycle analysis) commit this version of the fallacy. For example, in *The Food Revolution*, John Robbins claims that a pound of beef raised in California has a larger water footprint than taking a shower each day for six months; from this, he concludes that you can save more water by not eating a pound of beef than by not showering for six months (Tenth Anniversary edition, pp. 236-237). Robbins's conclusion is false and his argument is invalid because foregoing a pound of beef will not actually affect the amount of water that is consumed, whereas foregoing a shower for six months would affect the amount of water that is consumed. Robbins's mistake is to equate the actual effects of particular acts with the 'footprint' associated with those acts, where footprints are simply a measure of the *average effects* of acts of that type. In other words, his mistake is to commit the Average Effects Fallacy. I discuss other problems with carbon footprint reasoning several footnotes below in connection with an analogous argument due to Michael Pollan.

based on your knowledge of the situation you know that your crossing the street will not make any difference, given that there is no way you can convince the crowd to wait for the ambulance. In light of this, you cross the street to get to your car, because waiting for the entire crowd to disperse would be a significant inconvenience for you, and would not do any good for anyone else. Because so many people cross the street in front of the ambulance, the ambulance is delayed, and the patient inside dies, who would otherwise have lived. In this case, we can suppose that the average effect of an act of crossing the street is negative on balance, because the good that accrues to all of the people as a result of getting to their cars faster is greatly outweighed by the harm that is caused by their collective action. Nonetheless, it is permissible for you to cross the street, because the outcome only would have been worse if you had not crossed the street, because then no good would have accrued to you and every other relevant feature of the outcome would have been the same. The explanation is that the ethically relevant effects of a particular act are the marginal effects of that particular act, and not the average effects of acts of that type. (Similar remarks apply in more dramatic fashion to the asteroid and stampede cases described in the first section of this paper.)

As another example, imagine that you live in a different world in which everyone is required to make all of their charitable contributions on the last day of each year. There are only two charities to which contributions can be made, and the only options are to give $1,000 to one of those charities, or to give no money to any charity at all. Furthermore, there is a website where you can track everyone's decisions throughout the last day of the year – the website is highly accurate, updated every 5 seconds, etc. This system has been in place for several decades, and the trend is for almost everyone who is

going to contribute to a charity to make their decision early in the day, and in fact very few people (less than one million) ever decide to give to charity after 10pm. Usually, about 100 million people decide not to give to any charity at all, which is the 'default' option. Knowing all of this, at 11:45pm you are deciding how to make your personal charitable contribution, and by looking at the website you see that 111 million people have given to charity A, 11 million have given to charity B, and that 100 million people have not made any decision. It is well known that the good done by charity A is not a linear function of the number of contributors, because there are dramatic thresholds involved; in particular, for each increment of exactly 100 million contributors, charity A saves exactly 200 million people from death and ensures that they live happily ever after, but charity A does not otherwise help anyone in any way – so, because of this threshold effect, exactly 200 million people will be saved by charity A regardless of whether 111 million people, 121 million people, or 199 people give to that charity. In contrast, for each and every contributor to charity B, one person is saved from death and lives happily ever after. In this case, the thing to do is to give to charity B, despite the fact that the average effect of giving to charity B is significantly worse than the average effect of giving to charity A.[38]

This shows that the argument against my conclusion displayed above is unsound. The problem is that the conclusion of that argument does not follow from its premise about your proportional share of harm, given that 'your proportional share of harm' in the

[38] And despite the fact that the 'categorical imperative' says to give to charity A because everyone [in the same relevant circumstances] [always acting on a maxim that implies] giving to charity A would be significantly better than everyone [in the same relevant circumstances] [always acting on a maxim that implies] performing any of the relevant alternative acts.

argument must be interpreted as meaning, roughly, 'the average effects of acts of failing to reduce emissions' if that premise is to be plausible. The argument is initially seductive because we are often disposed to commit the Average Effects Fallacy in our reasoning by equating the ethically relevant effects of an act with the average effects of actual acts of that type, rather than with the marginal effects of the act itself.[39]

The street-crossing case also illustrates the mistake made by causal consequentialist theories that proportion credit for consequences based on agents' causal responsibility for those consequences. The street-crossing case is a counterexample to such views because, given suitable assumptions about that case, you are as causally responsible for the

[39] To reinforce the idea that we should focus on marginal effects rather than average effects, it is worth noting that even from a collective perspective we should focus on marginal effects rather than average effects. For example, consider a policymaker who is deciding whether to advise people to grow vegetables in their backyard, and suppose it is clear that individuals would lower their carbon footprint by growing vegetables in their backyard, because the average emissions associated with store-bought vegetables are higher than the average emissions associated with vegetables grown in a backyard. At first glance, it can seem obvious that the policymaker should advise people to grow vegetables in their backyard in light of these facts, if the only relevant goal is to reduce emissions. However, that does not follow, and to think it does follow is to commit a version of the Average Effects Fallacy. To see the problem, consider the realistic assumptions that only a small percentage of the population will actually follow the advice, and that the followers will be fairly evenly distributed with respect to the supermarkets where they currently buy vegetables; if so, the actual effect of such advice would be to increase emissions, because each follower will generate some new emissions in the course of creating and tending their garden, and any collective positive effects will be negligible compared to the collective negative effects, because the followers are dispersed and few in number, thereby having no important effect on the production of vegetables and hence emissions outside their backyards, because their dispersed consumption acts will not make a difference further up the supply chain. For such reasons, advising a population to reduce their carbon footprint in some way will often predictably lead to greater overall emissions, even if everyone who follows the policy succeeds in lowering their carbon footprint, and even if none of the non-followers increase emissions or their carbon footprints. This illustrates why it is crucial to focus on marginal effects rather than average effects at every level of individual and collective abstraction – and why decision makers at all levels should recognize that carbon footprints and lifecycle analysis provide unreliable evidence about what policies would have the best effects on emissions levels. (Michael Pollan commits the Average Effects Fallacy in his article "Why Bother?" when he concludes that growing your own vegetables would have good environmental effects because you would thereby reduce your carbon footprint. See also the discussion of an analogous argument due to John Robbins several footnotes above.)

outcome as every other person (imagine that everyone else engages in similar reasoning before crossing the street), and therefore your proportional share of the harm that is done is unacceptably negative according to a causal consequentialist view. Nonetheless, there is nothing wrong with your decision to cross the street, which shows that such views are mistaken.[40] (Similar remarks apply regarding the asteroid and stampede cases discussed in the first section of this paper.)

In response to all of this, proponents of the argument above might insist that 'your proportional share of harm' should be understood in terms of the *marginal effects* of particular acts rather than the *average effects* of acts of a general type. However, if 'your proportional share of harm' is understood in terms of marginal effects, then the premise of the argument above is no longer true that *your proportional share of the harm that is done will be very great unless you reduce your emissions*, because the harm that would actually be avoided by you reducing your emissions is negligible, whereas the harm that would be imposed on yourself by a substantial emissions reduction is considerable. Thus, the (expected) marginal effect of reducing your emissions by a substantial amount is worse than the alternative.

At first glance, it might seem that this reasoning must be mistaken for one reason or another. However, lingering doubts about this reasoning are probably driven either by the

[40] For discussion of causal consequentialist views, see Alvin Goldman "Why Citizens Should Vote: A Causal Responsibility Approach", and Dan Moller "The Morality of Overdetermination: Wrongdoing When it Makes No Difference", in *Abortion, Killing, and Overdetermination – Three Essays*, PhD dissertation, Princeton University.

Average Effects Fallacy, or else by a mistaken conception of the empirical facts regarding climate change.

For example, one of the most depressing facts about climate change is that greenhouse gasses are currently *accumulating* in the atmosphere, and will continue to accumulate into the foreseeable future even if emissions are reduced by as much as 50% from current levels, because the rate of emissions is much greater than the rate at which greenhouse gasses are removed from the atmosphere by natural processes.[41] In light of this, it would be a mistake to equate the ethics of climate change with the ethics of 'voting cases' in which your actions have a very small chance of making a dramatic difference, as in the following argument:

[41] From the authoritative Intergovernmental Panel on Climate Change (IPCC) report *Climate Change 2007: The Physical Science Basis*: "The concentration of greenhouse gas in the atmosphere depends on the competition between the rates of emission of the gas into the atmosphere and the rates of processes that remove it from the atmosphere. For example, carbon dioxide (CO_2) is exchanged between the atmosphere, the ocean, and the land through processes such as atmosphere-ocean has transfer and chemical (e.g. weathering) and biological (e.g. photosynthesis) processes. While more than half of the CO_2 emitted is currently removed from the atmosphere within a century, some fraction (about 20%) of emitted CO_2 remains in the atmosphere for many millennia. Because of slow removal processes, atmospheric CO_2 will continue to increase in the long term even if its emission is substantially reduced from present levels. … More specifically, the rate of emission of CO_2 currently greatly exceeds its rate of removal, and the slow and incomplete removal implies that small to moderate reductions in its emissions would not result in stabilization of CO_2 concentrations, but rather would only reduce the rate of its growth in coming decades. A 10% reduction in CO_2 emissions would be expected to reduce the growth rate by 10%, while a 30% reduction in emissions would similarly reduce the growth rate of atmospheric CO_2 concentrations by 30%. A 50% reduction would stabilize atmospheric CO_2, but only for less than a decade. After that, atmospheric CO_2 would be expected to rise again as the land and ocean sinks decline owing to well-known chemical and biological adjustments" (pp. 824-825). For further ramifications for greenhouse gas stabilization levels, see *The Economics of Climate Change: The Stern Review*, Chapter 8; for a clear summary, see Hal Harvey and Sonia Aggarwal, "The Costs of Delay", ClimateWorks Foundation.

> There is some danger that we will cross an emissions tipping point that will lead to a catastrophe; so, unless you reduce your emissions by a significant amount, there is some chance that you will cause this catastrophe. Therefore, you are required to reduce your emissions by a significant amount.

The problem with this argument is that it ignores the fact that greenhouse gas levels are currently accumulating, and will continue accumulating into the foreseeable future. To see why this is a problem for the argument, imagine that this week, as you engage in some emissions-generating activity, your emissions cause a catastrophic tipping point to be crossed. Nonetheless, even if you had avoided those emissions and thus hadn't tipped the scales yourself, it is certain that someone else's emissions would have tipped the scales at the same time, because an entire planet of other people would still have been emitting at the same time even if you had not been emitting. This shows that, given the empirical facts, there is no chance that you could delay a catastrophic tipping point from being crossed today or in the foreseeable future by reducing your emissions, and thus there is no good reason for reducing emissions that arises from the possibility of tipping points being crossed now or in the foreseeable future.[42]

[42] More precisely: there is no chance that you could delay a tipping point from being crossed today, *at least not by any tangible amount of time*. Here it is worth noting that catastrophic tipping points are *global* in the sense that they are a function of emissions that tend to be dispersed quickly around the entire earth, which makes it unrealistic to believe that some particular, localized facts about your emissions could conspire to cause a catastrophe that would otherwise have been avoided. In response to my reasoning in the main text, someone might suggest that we should focus on your emissions over your entire lifetime, rather than on your emissions on a particular day. This wouldn't affect my point, since a reasonable individual's

This shows that it is a mistake to think that the ethical issues regarding climate change are essentially the same as the ethical issues regarding voting in an election. That is because in an election a crucial consideration is that your vote has some chance of tipping the scales and making a dramatic difference to the outcome, whereas a depressing fact about climate change is that, as things stand now and as they will stand into the foreseeable future, your personal emissions have no real chance of making a tangible difference to the outcome, except by making things substantially worse for yourself.

There are also further important differences between voting and climate change. For example, it is not costly to vote in an election, whereas it is costly to reduce your emissions by a substantial amount; in addition, it is plausible that you have important deontological reasons to vote, whereas you do not have important deontological reasons to reduce your emissions by a substantial amount (as I argued above). This shows that the considerations relevant to the ethics of climate change are importantly different than the considerations relevant to the ethics of voting, and that the challenge of explaining how individuals could be required to vote in an election is an easy problem compared to the challenge of explaining how individuals could be required to reduce their greenhouse gas emissions by a substantial amount.[43]

---

emissions over an entire lifetime would still be much less than the emissions happening elsewhere on the planet in any single second.

[43] To say it is an easy problem compared to another problem is not to suggest that anyone has ever offered a plausible argument that citizens are required to vote in elections, especially from a utilitarian perspective. (Thanks to Michael Cox and Richard Chappell for encouraging me to discuss the connection between the ethics of climate change and the ethics of voting.)

One important qualification to all of this is that there is some chance that in the future humans will change their behavior dramatically (i.e. there is some chance that in the future humans will reduce emissions by more than 50% from current levels), thereby creating a situation in which greenhouse gasses are being removed from the atmosphere faster than they are added. If such a scenario ever obtains, then the argument above would not fail for the main reason just given. However, to argue that individuals are currently required to reduce emissions by a substantial amount because of the possibility of such an unlikely change in the future is like arguing that you would be required to go to the polling place and attempt to cast a ballot even if the polls were already closed and the costs to you were high, because there would still be *some* chance that precedent would be reversed and somehow your vote would count, and thus there would be *some* chance that your vote would end up making a dramatic difference. This shows that insofar as your situation with respect to climate change can be accurately described as analogous to a voting case, it is analogous to a voting case in which you are clearly not required to vote.[44]

The most decisive way of showing that tipping point reasoning is misguided is to note that it overgeneralizes in absurd ways. For example, the probability of causing a climactic catastrophe by your personal emissions is roughly the same as the probability of causing a climactic catastrophe by taking a swim in the ocean or having a cup of tea every day, given that the effects of your emissions on the potential causes of climactic catastrophe such as sea level rise and air temperature increase are about the same as the effects on

[44] Thanks to Joe Rachiele for encouraging me to discuss the issues in this paragraph.

those things of taking a swim in the ocean and boiling water for tea every day, respectively. As a result, if it is impermissibly risky not to reduce your emissions by a substantial amount because of the chance that you will cause a tipping point to be crossed, then it must also be impermissibly risky to swim in the ocean or have a cup of tea every day, since the very same risk is involved, and the potential benefit to yourself is of no greater significance. But it is absurd to think that swimming in the ocean and having a cup of tea are impermissible because they risk climactic catastrophe. This shows that there has to be something irreparably wrong with the tipping point reasoning under consideration.[45]

A different kind of objection arises from the idea that large numbers of intangible harms can be just as bad as smaller numbers of substantial harms. According to this sort of objection, you are required to reduce your emissions because the aggregate harm caused by your personal emissions, via many intangible harms to others, outweighs the significant good your emissions do for you, your friends, and your family.

To evaluate this objection, it is important to distinguish two different issues. One issue is whether intangible harms *count for something*, and another issue is whether intangible harms *count for enough to outweigh your own personal and familial interests*. To make progress here, it might help to consider a possible case that illustrates these issues. Suppose that 10,000 men are currently set to be tortured with intensity i for the next hour, and that you must choose between either increasing the intensity of their pain by an imperceptible amount (whatever that means), or else causing your innocent niece, who is

[45] Thanks to Jane Willenbring for helping me appreciate the importance of this point.

currently set not to be tortured, to also be tortured with intensity i for the next hour along with the 10,000 men. As you make your choice, one question is whether there is a reason that counts against burdening the men with an additional amount of intangible pain; however, even if there is such a reason, a further question is whether that reason is outweighed by the reason you have not to burden an additional innocent person with an hour of intense torture that she would not otherwise experience.

In this case, it is clear that you must choose to spare your innocent niece from torture, even though that implies causing each of the men an imperceptible amount of additional pain. On reflection, this undermines the idea that intangible harms caused by emissions could aggregate in a way that provides strong reasons against those emissions. To see why, suppose that the torture will be delivered by a device that has adjustable dials that determine how much pain is inflicted on each subject. Each subject's dial can be set anywhere between 0 and over 10,000, and the settings have been carefully designed so that each increment delivers an imperceptible amount of additional pain, with 0 delivering no pain, and 9,999 delivering the fairly intense torture that the men are antecedently set to endure.[46] So, as you make your decision, each man's dial is set to 9,999. If you choose to spare your niece, you cause 10,000 additional units of pain to be inflicted (distributed equally and imperceptibly among the men, one unit each); however, if you choose not to spare your niece and have her tortured along with the men, you cause only 9,999 additional units of pain to be inflicted (all inflicted on your niece). Even in

[46] We can also suppose that the increments have been carefully designed so that each additional increment results in linearly decreasing utility in the subjects, insofar as it is possible to determine a utility curve for pain in such a case.

light of all these details, it is obvious that you should choose to 'inflict' the pain on the men rather than on your niece. Furthermore, this remains true even if we increase the number of men who would be affected to 100,000 or more, given that your choice is still between causing someone to be tortured who will not otherwise be tortured, and merely doing something that will have no tangible effect on anyone. This shows that it is false to claim that intangible harms aggregate in anything close to a simple additive way. Instead, the case shows that on any plausible view, substantial harm inflicted on particular people tends to be much worse than an equal 'sum of harm' that is distributed intangibly among many people. In light of this, it is hard to see how there could be a plausible argument that the intangible harms caused by your emissions could outweigh the important goods that those emissions bring to you and your friends and family.[47]

A different fallacy is committed by the following reasoning:

---

[47] All of this is consistent with thinking that intangible effects *count for something*. For important arguments that they do count for something, see Jonathan Glover, "It Makes No Difference Whether or Not I Do It", and Derek Parfit, *Reasons and Persons*, pp. 75-82. Glover and Parfit slide from reasons for thinking that intangible harms count for something to the unwarranted conclusion that intangible harms aggregate in a simple additive way. For example, Glover advocates a simple additive aggregation principle that he calls 'The Principle of Divisibility' (see pp. 174-175). However, if such a principle were correct, then it would be a matter of moral indifference for a person who has to steal in order to survive whether he steals all of what he needs from one person, thereby making that person seriously worse off, or instead steals what he needs from a very large number of people in a way that does not tangibly harm anyone but imposes the same 'aggregate sum of harm'; as far as I can tell, this consequence is just as perverse as the unintuitive consequences of traditional consequentialism that Glover and Parfit are attempting to explain away by means of simple additive aggregation principles – and by the same token it shows that simple aggregation principles cannot really explain the entire range of cases that Glover and Parfit intend to explain by means of such principles. This undermines any reason for favoring such principles even from a traditional consequentialist perspective. Furthermore, from the perspective of non-consequentialist theories, there is nothing in need of explanation in the first place, because the so-called problem cases (Glover's cases involving bandits, Parfit's cases involving torturers) can be readily explained on non-welfare-based grounds. For further discussion of issues about intangible harms, aggregation, and analogous distributional issues, see Larry Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*, and Shelly Kagan, "Do I Make a Difference?". I discuss other related issues at greater length in below.

> If we reduced our emissions collectively by a substantial amount by means of carefully designed regulations, we would be better off on balance. Therefore, even as things stand now, the marginal effects of reducing your emissions by a substantial amount would be positive, and so you are required to reduce your emissions by a substantial amount.

The problem with this argument is that even if the premise were true, it would provide no reason to believe that the conclusion is true, given what we know about the relevant empirical facts. For example, insofar as the premise is plausible, that is because many emissions in our economy are inefficient, and inefficient emissions are the emissions that would be reduced by carefully designed regulations. However, that does not show anything about your personal emissions, because insofar as my arguments apply to your personal emissions, your personal emissions are among the most efficient emissions in our economy. The main reason why this is true is that your personal actions have surprisingly little effect on the total level of emissions, and so each unit by which you actually reduce emissions comes at a surprisingly high cost of well-being. For example, it is easy to think that you are reducing emissions by turning out lights in your house, or by turning down your thermostat. But in reality such actions have no effect on emissions at all, because they have no effect on how much energy is produced by the power plants that

supply the power, and thus have no effect on the power plants' emissions.[48] Similarly, it might seem that growing a vegetable garden in your backyard is an effective way of reducing emissions, but in fact growing your own vegetables won't really have any effect on the number of vegetables produced outside of your backyard, which means that the net effect of your garden will be to increase the total amount of emissions, since you will generate some additional emissions in the course of developing and tending your garden. For numerous reasons like these, there is no good reason for thinking that, as things stand, the marginal effect of attempting to reduce personal emissions by a significant amount would better than the alternative, unless you happen to have a strong personal preference for 'green living'. This is perfectly consistent with the idea that everyone would be better off if we collectively reduced emissions by a substantial amount by means of carefully designed regulations. Understanding that these claims are consistent is crucial to understanding the consequentialist and economic logic of our situation with respect to climate change. As a result, other than forgoing use of an automobile, it is largely inaccurate to think that your actions are even capable of reducing emissions by a significant amount. I will simply ignore this complication for ease of exposition, because including this complication would only strengthen my arguments.

A different kind of objection is that my arguments mistakenly imply that charitable acts are never required. However, my arguments do not have that implication, because it is perfectly consistent to think that you are required to be charitable, and even that you can satisfy that duty by reducing your emissions, but that nonetheless you are not required to

[48] Assuming that your power is produced in a typical way.

reduce your emissions. For example, it is plausible to claim that individuals are periodically required to perform some charitable acts or other, such as giving money to charity, helping at soup kitchens, helping at animal shelters, or other such things, even though, for each of those acts, no individual is required to perform that particular act – including reducing emissions, on the assumption that reducing emissions is a charitable act. In other words, although you might be required to be charitable, it doesn't follow that for any particular (type of) charitable act, you are required to do that particular (type of) charitable act. For this reason, the idea that individuals are required to be charitable would not undermine the claim that individuals are not required to reduce emissions by a significant amount, even on the assumption that reducing emissions is a charitable act.[49]

Other objections to my arguments come from versions of virtue ethics that claim, roughly, that an act is required if and only if a virtuous person would perform that act. From the perspective of virtue ethics, the objection is that you should reduce emissions, because reducing emission is what a virtuous person would do, regardless of the consequences, and regardless of any other deontological considerations, etc.

However, it is implausible that a virtuous person would reduce her emissions by a significant amount, because a virtuous person would not sacrifice her own well-being and the well-being of her family in a way that does not benefit anyone else, and a virtuous person would not act on the basis of the kind of mistaken reasoning that is required for thinking that such individual action is an effective way of combating climate change. For example, if you reduce your emissions by canceling family vacations and family trips,

[49] Thanks to Heather Berginc for encouraging me to discuss the issues in this paragraph.

that imposes significant costs on your family, and does not really help anyone else, and is not an effective way of contributing to a solution to climate change. And if you walk or ride your bike to work every day, you impose a non-trivial risk of catastrophe on your family and friends, because there is a much greater chance that you will be killed in an accident if you walk or bike than if you drive – and, again, imposing this risk on your family does not really benefit anyone else, and is not an effective way of contributing to a solution to climate change. In light of all this, there is no reason for thinking that a virtuous person would impose such costs on her friends and family by reducing her personal emissions by a significant amount. Instead, a virtuous person would favor and perhaps support promising intergovernmental solutions to climate change, because such solutions are the only ones that have any real prospect of success.[50]

A related objection is that my conclusions must be mistaken because they imply that hypocrisy is permissible. However, my arguments do not have that implication. To see why, consider a serious tragedy of the commons in which your family must choose between either using the commons to ensure their own flourishing, or else foregoing use of the commons in a way that is certain to be both very costly and futile, because the commons will be destroyed by the overuse of others regardless of your family's actions. What is the correct and virtuous thing to do in such a situation? On the one hand, it is true that your family and every other family should favor an effective solution to the social dilemma. On the other hand, if you know that such a solution is not forthcoming, and that

[50] It is worth noting that climate change advocates such as Al Gore have enormous carbon footprints as a result of their activities, which focus largely on convincing others to favor large-scale political solutions to climate change. Nonetheless, such people are quite virtuous, abstracting away from their non-climate-related activities.

the commons will be destroyed regardless of what your family does, then it would be wrong for you to sacrifice your family in a way that is certain to be futile by foregoing use of the commons. In such a situation, you are not a hypocrite if you say "As things stand, I am not required to sacrifice the well-being of my family, even though I favor a course of action in which everyone makes sacrifices for everyone's benefit, perhaps by means of mutual coercion, mutually agreed upon". Such a statement expresses correct judgment, not hypocrisy – and is a straightforward analogue to my view about the individual ethics of climate change.[51]

Another related objection is that you are required to reduce emissions by a significant amount because you should want to be the kind of person who reduces emissions by a significant amount in order to express something like a *symbolic protest* of the unfolding climate catastrophe, even if such reductions are costly to you and your family and cannot be expected to have any other important effect. There are two main problems with this objection. The first problem is that even if it were true that you should want to be a person who performs the futile but symbolic act of reducing emissions in a way that is very costly to you and your family, that would not show that such an act is *required*, because it would be much more plausible to think of such an act as *supererogatory*. But more importantly, you should not want to be such a person, because you should want to be the kind of person who thinks clearly and correctly about climate change, and who recognizes that imposing significant costs on yourself and your family does not really

[51] The phrase "mutual coercion, mutually agreed upon" is from Garrett Hardin, "The Tragedy of the Commons". Like many commentators, I don't agree with Hardin that coercion is always the best way to solve collective action problems, but I agree that it is sometimes the best way. (Thanks to Gideon Rosen for encouraging me to discuss hypocrisy.)

benefit anyone else and would not even begin to be an effective solution to climate change even if very large numbers of others acted in a similar way; instead, you should want to be the kind of person who recognizes that promising intergovernmental solutions to climate change are the only solutions that have any real prospect of success, and you should want to be the kind of person who favors and clearly explains the reasoning behind such solutions to others, instead of counterproductively acting on the self-righteous and tragic conceit that reducing emissions unilaterally as an individual in a way that is very costly to you and your family has some connection to a real solution to the problem.[52]

At this point, I have replied to a number of important objections to my view. I will now offer a minimalistic analysis of the ethical significance of futility, and then offer a more rigorous argument that you are not required to reduce your emissions by a significant amount.

[52] Thanks to Cheshire Calhoun and Jeff Downard for encouraging me to consider this kind of objection. For related discussion, see Thomas Hill, "Symbolic Protest and Calculated Silence". My discussion here is perfectly consistent with thinking, as I think we should, that *some* acts that do not have any important effects are genuinely required. I provide examples of such acts and a principled way of distinguishing them from non-required acts in the next section. (In connection with Hill's paper, it is worth noting that his examples tend to rely on cases in which individuals are intuitively required to "disassociate from irreparably corrupt groups" (92). However, there is no clear analogy between such cases and our situation with respect to climate change.)

3. The Ethical Significance of Futility

The intuitive idea behind the ethical significance of futility is that you are not required to do something if it would be both costly and futile. However, this intuitive idea could be interpreted in a number of different ways. For example, it could be understood to imply that it is permissible for you to steal other people's property if you are in a situation of looting and rioting and you know that someone will steal or destroy the property regardless of what you do; it could also be understood to imply that it is permissible for you to ignore the suffering of distant strangers if nothing you could do would make a difference to the underlying problems that cause such suffering. To avoid interpreting the intuitive idea as having such mistaken implications, while keeping the idea as uncontroversial as possible, I suggest the following:

Minimal Analysis of the Ethical Significance of Futility

> If you know that a course of action would be costly to you, and that there are no significant welfare-based reasons that support that course of action, and that there are no significant rights- or respect-based reasons that support that course of action, then you are not required to take that course of action.

Once the intuitive idea is developed in this way, it does not imply that it is permissible for you to steal another person's property even if that property would be stolen by someone else anyway, because there are significant rights-based reasons not to steal property; it also does not imply that it is permissible for you to ignore those who are suffering if you can help them in some way, because there are significant welfare-based reasons to aid whenever you can make a difference to even one person, even if you cannot do anything about the underlying situation that is causing such suffering.[53]

The Minimal Analysis also helps to clarify the correct way of thinking about many other cases in which actions are required even though they would be costly and in some sense futile. As a particularly horrible but instructive example, imagine that a would-be rapist considers raping my unconscious body after I have been raped by many men. In such a case, there are no welfare-based reasons why the would-be rapist should not also rape me, at least if we understand the facts of the case in a suitable way. Nonetheless, it would be

[53] Of course, there is still the residual problem of what you should do when you know that a particular aid organization, although quite effective, is so large that your personal contribution to the organization would not make any difference to the outcome. To get some traction on this issue, it might help to consider a case in which you must choose between making a donation to such an organization, where you know that the marginal effect of your donation is zero, and giving the same amount to smaller organization, where you know that your donation really would make a significant positive difference to the outcome. In such a case, if you know that your donation would have these differing effects, it would be *better* from an ethical perspective to give to the second organization, even if you know that the average positive effect of a donation to the first organization is greater than the average positive effect of such a donation to the second organization. This suggests that the gift that would be *best* from an ethical perspective really does depend on the (expected) marginal effects of the gifting options open to you. Of course, such gifts are (arguably) generally supererogatory, so a further complication is that it would (arguably) generally be permissible to give to suboptimal charities instead. More vivid support for all of this is provided by reflection on the charity case discussed several footnotes above in connection with the Average Effects Fallacy.

dramatically wrong for him to rape me, because there are dramatic non-welfare-based reasons not to do so.[54]

As a more challenging example, imagine that I am deciding whether to eat a hamburger at a restaurant, and I know that the hamburger comes from factory farms. While I enjoy eating hamburgers, I also believe that suffering lies behind their production, and that the pleasure derived from eating hamburgers is insignificant compared to that suffering. Nonetheless, although I am unwilling to cause such suffering myself, I recognize that this does not give me a clear reason not to eat the hamburger, because my eating a hamburger at the restaurant will not have any effect on the number of cows that suffer and the extent of that suffering. In fact, eating a hamburger would not even support the factory farms

[54] Unfortunately, this case is quite realistic, as thousands of people have been raped in such circumstances in the past several decades, especially in war zones. (For some not widely-recognized examples, see http://www.guardian.co.uk/society/2011/jul/17/the-rape-of-men). Classical utilitarianism implies that, from both a subjective and objective perspective, such a rapist is required to rape me, given suitable assumptions about the case. Versions of utilitarianism that appeal to preference satisfaction could attempt to explain the wrongness of such an act in terms of my preference not to be raped, but such explanations are ultimately unsuccessful – for example, when I reflect honestly on a related case in which I am dead rather than unconscious, I find that I would be willing to accept having my dead body raped by one additional man after it has been raped by many in exchange for a small financial benefit for my surviving family. It is unbelievable that this fact about my preferences would make it permissible (or mandatory!) to rape me in such a situation if the would-be rapist merely had a more weighty preference for raping me, as presumably he would. (Other people might have similar preferences regarding even a case in which they are still alive but unconscious, given a suitable description of the case.) Further examples of analogous problems for traditional consequentialism are provided by many clearly impermissible consumption activities that individually have no effect on the outcome – for example, an individual's decision to watch freely downloadable child pornography in the privacy of his own home, etc. (If it is really necessary, we could imagine that such people take a pill that guarantees that they will subsequently forget having engaged in such activities, thereby cutting off an uninteresting line of consequentialist reply to such cases.) Some theories that identify themselves as versions of 'consequentialism' avoid these problems, but such theories are best classified as deontological theories (or some other type of non-consequentialist theories) that are merely expressed using the conceptual structure of consequentialism. Here I have in mind 'consequentialist' theories that specify the good in a way never encountered in consequentialist theories until the late twentieth century, such as agent-relative consequentialist theories that imply that there is something especially bad about performing particular types of acts oneself, etc. The upshot is that no traditional consequentialist theory is plausible after reflection upon such cases.

that cause such suffering in any tangible way, because although buying this hamburger will benefit the restaurant, it will not tangibly benefit factory farms.[55]

Here it is important to resist the temptation to commit the Average Effects Fallacy. While it is true that the *average effect* of eating a hamburger at a restaurant includes some harm to cows and some benefit to beef producers, that does not show that the *marginal effect* of buying a hamburger on this occasion would include harm to cows or benefit to beef producers. In fact, it is plausible that the marginal effect of an entire lifetime of consuming hamburgers would not include any significant harm to cows or any significant benefit to beef producers. That is because cows, like most other products we consume, are delivered by a massive and complex supply chain in which there is waste and other forms of inefficiency at each link, where that inefficiency serves as a buffer to absorb any would-be marginal effects from the links before. As a result, it is implausible that even a lifetime of decisions to consume animal products by an individual would make a significant difference to the number of animals produced and harmed over that time.

For a particularly clear illustration of this, consider the supply chain for American beef. When ranchers who own their own land decide how many cattle to raise, their decisions are sensitive to their own financial situation, the number of cattle their land can support, the expected price of feed, and the expected price that the cattle will fetch when they are

[55] Here I disagree with Tristram McPherson and Tom Regan, who argue that it is wrong to consume animal products because doing so supports morally objectionable institutions ("Why I am a vegan", pg. 11, and *The Case For Animal Rights*, updated edition, pg. 351, respectively.) For similar reasons, and for the reasons discussed in the previous footnote, I disagree with traditional consequentialist explanations of such cases – for example, I disagree with Peter Singer's explanation in *Animal Liberation*. (I follow McPherson and many others in noting such 'inefficacy' objections to traditional consequentialist explanations, which I discuss in detail in the following chapter.

ultimately sold to feedlots. Of these, small changes in the last are of the least importance, because if ranchers think that ranching will continue to yield a greater rate of return than other investments, they will tend to raise as many cattle as they can afford to raise and feed or that their land can support, letting the ultimate extent of their profits fall where it may at the feedlot. Many ranchers also use the nutritional well-being of their herd as a buffer to absorb adverse changes in market conditions, feeding the cattle less and less to whatever point maximizes the new expectation of profits as adverse conditions develop, or even sending the entire herd to premature slaughter if, say, feed prices rise to levels that are unacceptably high. This serves to shift the ranchers' emphasis in decision making even further away from herd size. As a result, even if an individual's consumption of beef somehow (implausibly) managed to have a tiny effect on the price of cattle at feedlots, the effect on the number of cattle produced would be much smaller than it would have to be in order for the possibility of such a threshold effect to justify equating the expected marginal effect of an individual's consumption of beef with the average effect of such consumption decisions. Similarly, when ranchers lease grazing land from the government, if cattle ranching can be expected to be profitable, ranchers will collectively purchase all of the scarce and independently determined number of grazing permits and raise the maximum number of cattle that are allowed by those permits, because it does not make economic sense to hold such permits unless one grazes the maximum number of cattle allowed on the relevant parcels of land. As a result, the number of animals that are raised on land leased from the government is almost completely insensitive to tiny changes in the price of cattle at feedlots.

More importantly, because animal production is so many links in the supply chain away from grocery stores and restaurants, and because each of the intervening links involves waste and inefficiency that serve as a buffer to absorb any effect that your personal consumption might otherwise have, it is unrealistic to think that your personal consumption could really have any effect on decisions made at the beginning of the supply chain, even when your consumption is considered over the course of an entire lifetime.[56] That is because the actual mechanisms by which information is conveyed and decisions are made throughout the supply chain do not give rise to the sort of threshold effects near the margin that philosophers tend to imagine as driving the expected marginal effect of an act of consumption toward the average effect of such consumption; instead, waste and inefficiency ensure that the real expected marginal effect of an individual's consumption is essentially zero, because the change in the signal received at the production end of the supply based on a change in a single individual's consumption decisions is almost certainly zero, and does not have a significant enough chance of amounting more than noise to give rise to any important expected effects.

Similar reasoning applies in the case of American poultry and pork, although the relevant market mechanisms are less transparent because of the vertical integration of those industries. Despite that complication, it remains true that the actual mechanisms by which information is conveyed and decisions are made throughout the supply chain are

[56] Here it is important to keep in mind that the amount of beef that you have personally consumed in the non-proximate past cannot have any real effect on the current or future price of beef at feedlots. So, it is simply a confusion to imagine, as many philosophers do, that the marketplace somehow keeps track of your personal cumulative lifetime consumption, and that once your cumulative lifetime consumption reaches some threshold level it causes something to happen in the world that would not happen otherwise. Upon reflection, it should be clear that such reasoning is philosophical fantasy.

insufficiently fine-grained to be sensitive to the consumption decisions of individual consumers in the way that would be necessary for there to be important welfare-based reasons for individuals not to consume those products, and it remains true that every link in the supply chain contains inefficiency that serves as a buffer to absorb any effect that an individual's consumption decisions could otherwise have.

This is not to deny that there are genuine welfare-based reasons not to eat meat. On the contrary, you would be healthier if you didn't eat meat, and your vegetarian lifestyle would influence others in a similarly positive way. However, those are not strong enough reasons to require you to adopt a vegetarian lifestyle, even from a utilitarian perspective. To see why, note that you would be healthier if you didn't consume alcoholic beverages, and that your abstention from alcohol would have positive health effects on others, but that does not mean that you are required to give up alcoholic beverages if you really enjoy those beverages and are able to enjoy them without harming others, even from a purely utilitarian point of view.

Another important point is that even if you would convince many others to be a vegetarian by becoming one yourself, that does not show that you have strong welfare-based reasons to become a vegetarian, because even if your vegetarian lifestyle caused, say, one hundred others to become vegetarians who would not otherwise have done so, their collective consumption decisions would still not have any appreciable effect on the number of animals that are raised and mistreated, because the actual mechanisms in the marketplace are not sensitive to the effects of even one hundred consumers. Of course, this reasoning does not hold true when applied to an influential person like Peter Singer

who really does influence enough people to make a difference, but it does hold true when applied to almost everyone else, which means that utilitarianism does not require most individuals to become vegetarians, even if it requires a few influential people like Peter Singer to be vegetarians.[57]

A particularly dramatic way of illustrating the problem with welfare-based arguments is to imagine a future course of events in which factory farmers attempt to neutralize the ethical objections to their practices by simply producing more animals to increase the amount of spoilage in the supply chains, thereby undermining any possible effect that a single individual's consumption decision could have. If factory farmers adopted such a strategy, then utilitarianism implies that even if eating meat from factory farms were wrong now because of welfare considerations, factory farmers could succeed in making it permissible to consume animal products from factory farms merely by behaving in a way that was more evil and that intentionally inflicted more suffering on animals.[58]

In light of all this, utilitarianism and other welfare-focused theories cannot explain how it could be wrong to eat meat from factory farms. (I discuss these issues in greater detail in the following chapter.)

---

[57] Analogy: utilitarianism does not require us to act as if we had the talents, influence, and resources that Warren Buffett has; similarly, utilitarianism does not require us to act as if we had the talents, influence, and resources that Peter Singer has.

[58] Even setting aside the other arguments in this paper, it is bizarre to think that practical questions about what individual consumers should do could be hostage to empirical fortune in such a way – even utilitarians such as Peter Singer should be uncomfortable with the consequence that such an intentional increase in the suffering associated with a product that is itself completely unnecessary for well-being could make it the case that individuals are required to consume that product.

A different strategy is to argue that eating meat from factory farms is impermissible for non-welfare-based reasons. As the preceding discussion reveals, such an argument cannot be as straightforward as it initially appears, because an individual's decision to consume meat does not actually result in any additional animals being harmed or killed, and so does not obviously violate deontological constraints or any other ethical principles. However, it could be claimed that there are decisive reasons for you not to eat meat that arise *indirectly* from the fact that eating meat *depends on* or *supports* industries that directly violate significant constraints themselves.

Unfortunately, this reasoning fails for several reasons. One problem is similar to the problem with welfare-based arguments: initial appearances aside, your purchases at supermarkets and restaurants do not really have any tangible effect on the revenues of factory farms, and therefore do not support factory farms in any tangible sense. (Although your individual purchases have a tangible effect on the revenues of *supermarkets and restaurants*, the expected marginal effect on the revenues of *factory farms* is essentially zero.) More importantly, this reasoning overgeneralizes and implies that you are almost never permitted to consume anything at all, because petroleum companies routinely violate significant constraints,[59] and almost every consumption activity depends on and supports such companies to a much greater extent than buying animals at supermarkets and restaurants depends on and supports factory farms, because we often purchase gasoline directly from the petroleum companies themselves, or are at least only one step in the supply chain removed from such companies, and in the

[59] For examples of routine abuses, see Peter Maass, *Crude World*, especially chapters 2, 3, and 4.

petroleum supply chain there is much less of a buffer caused by waste and inefficiency than in the supply chain for animal products.[60]

In light of all this, the most plausible account of what we are required to do when products are produced in a morally objectionable way depends on a distinction between activities that are *essential* to a product or to the actual production of that product, and activities that are not – although this distinction is sometimes a matter of degree or dimension.[61] To illustrate the basic idea, consider a can of vegetables that is produced in a normal way. Although the production of those vegetables might depend on petroleum products, there is nothing about the vegetables that depends *essentially* on causing suffering or violating significant constraints, because there is nothing in the nature or the actual production of the vegetables that necessitates suffering or the violation of significant constraints. As a result, there are no strong reasons not to consume a can of vegetables even if you know that its production depends in some way on petroleum products that happen to be produced by corporations that often violate significant constraints.

Evidence for the importance of this distinction is provided by consideration of other cases, such as the difference between consuming ordinary goods like watches that were

[60] Here I seem to disagree with Tom Regan, who argues that "Since [factory farming] routinely violates the rights of these animals…it is wrong to purchase its products" (*The Case For Animal Rights*, updated edition, pg. 351). I suggest a more plausible way of developing Regan's view further below. More clearly, here I disagree with Tristram McPherson's view in "Why I am a vegan".

[61] Thanks to Ryan Jenkins, Ryan Robinson, and Ángel Pinillos for helpful discussions about such a distinction; I don't know whether they'd be sympathetic to what I'm saying here. It is important to note that the following discussion should not be interpreted as suggesting that there are no other reasons not to consume products when they are produced in a way that does not essentially involve harm or the violation of constraints – the point is merely that other reasons are less decisive and more easily outweighed.

merely produced by the Nazi regime, and consuming soap and other products that were made directly from the bodies of those killed in Nazi concentration camps. In both cases, consuming the products now would have no important effect on others, and could not, of course, *support* the Nazi regime. However, there are still strong reasons not to consume products that were made in a ruthless way from the bodies of innocent people at the direction of an evil regime, whereas there are not strong reasons not to consume products like watches that were manufactured in an ordinary way merely at the direction of such a regime. The distinction above provides a straightforward and compelling explanation of the difference, because it is fairly essential to the actual production of the soap that humans are used in unacceptable ways that violate significant constraints, whereas it is fairly inessential to the actual production of the watches that anyone is used in unacceptable ways.[62]

[62] Even if the soap is made only from dead bodies that would otherwise waste away, it still violates important constraints to use human bodies in such a way. If the watches were actually produced by slave labor, then consuming them is more objectionable than if they were not actually produced by slave labor, but less objectionable than consuming the soap, because it is still less essential to the product and its actual production that significant constraints are violated, because the watches could easily have been produced in much the same way without slave labor; insofar as we imagine the watches being produced in a way that essentially involves the severe mistreatment of workers (at the limit, imagine that the hands of the watch are made from human bones that were removed from still-living humans), it becomes more and more wrong to trade in such watches. The distinction above provides the most natural explanation of all of this. A more difficult case is that of sweatshop clothing. On the one hand, it could be claimed that the mistreatment of workers is not an essential part of the production of sweatshop clothing. However, on reflection, the distasteful use of human labor seems to be, at least along one important dimension, an essential part of the method by which clothing from sweatshops is produced. Nonetheless, that does not settle the question of whether it is permissible to consume clothing from sweatshops, because there is much variation in sweatshop behavior, and it is sometimes unclear how significant the constraints are that are being violated. For example, some sweatshops violate significant constraints by chaining people to their sewing machines and forcing individuals to work in horrible conditions that could never be justified, whereas other sweatshops merely pay workers a very low wage and insist that they work very long hours, while still arguably offering such workers better prospects than they would have if the sweatshops did not exist. In light of this, the ethical issues regarding the consumption of products from sweatshops are subtle even by the standards of this paper, and should be addressed on a case-by-case basis depending on the

The distinction above also offers better explanations than the theoretical alternatives, such as an explanation in terms of *complicity in evil*. In the example just given, both the soap and the watches were (let us suppose, for the sake of a clear example) produced by the Nazi regime itself, which means that the producer of both was equally complicit in violating constraints – because it was, after all, the very same producer in both cases. Nonetheless, there is an important difference in the permissibility of consuming those goods. As another example, imagine a beauty cream that is produced by extracting a substance from the bodies of innocent human beings who are painfully killed in the process, where this substance cannot be procured in any other way. Although the process is as horrible as it sounds, suppose the amount of the substance obtained from one innocent person is enough for about 100,000 bottles of beauty fluid; at the same time, other beauty creams are made in a normal way that does not harm anyone, and they are only slightly less effective in promoting beautiful skin. However, the factories that produce these other beauty creams require more inputs than their horrible competitor, and they turn a blind eye to the misconduct of their suppliers and the petroleum companies that provide the fuel to transport their goods and run those factories. As a result, both kinds of beauty fluid have the same 'human rights violation footprint' once the violations of suppliers and petroleum companies are taken into account, although the footprint of the operations under their direct control differs dramatically. In light of all these facts and the assumption that it will not make any difference to human suffering whether you personally consume either of these products, it would still be worse to consume the first

---

specifics of the case – and require further reflection on the significance of the constraints that are violated by various kinds of exploitation, etc.

kind of beauty fluid than the second. If this is not immediately clear, imagine that, because of a medical condition, you had to use one of these two kinds of skin treatments. Would it be a matter of moral indifference which you chose? If your affliction was serious, it would be better to choose the second than the first.

Furthermore, if complicity in evil really did give rise to strong reasons not to consume products, then it would be wrong to consume petroleum products because of the oil industry's complicity in serious harm, and it would be wrong to consume almost everything else as well, because transportation companies turn a blind eye to oil companies' abuses, which are known to benefit transportation companies in the form of lower fuel costs. Despite all of this complicity in evil, it is nonetheless permissible to consume such products, and the explanation is that it is not *essential* to such things that suffering is caused and significant constraints are violated. This is not to say that complicity in evil provides no reason to forego petroleum products – the point is merely that such reasons can much more easily be outweighed by our own interests, because they are much weaker than the reasons that arise when products essentially involve causing suffering and the violation of significant constraints.[63]

In order to develop the distinction into a more complete theory, it is important to note that in many cases there are important further distinctions between different activities that fall under the general rubric of *consuming a product* – for example, there are distinctions between merely purchasing a product, merely displaying a product, physically consuming a product, using a product in some other way that does not involve physical consumption,

[63] As a result, I disagree with the view in Christopher Kutz, *Complicity*.

and other ways of being a consumer. In many cases the permissibility of these more specific activities go together, but in some cases they come apart.[64]

For example, consider a case involving soap made from the bodies of innocent political prisoners during a past war. Suppose that some individuals want to sell the soap at a profit to dealers who want to sell it at yet another profit to others who want to purchase the soap for various reasons – some want to acquire the soap in order to destroy it respectfully and remove it from the face of the earth forever, some want to acquire the soap to display in museums to illustrate the brutality of previous regimes, some want to display the soap in their own homes to demonstrate their own personal wealth and consumptive power, and some want to use the soap in their bathrooms, perhaps for novelty effect, or perhaps as a token of their dominance over other human beings. Imagine that as the market for this soap develops, a billionaire takes notice and is emotionally struck by the situation, and decides to personally intervene in order to ensure that all of the soap is collected and respectfully destroyed once and for all. The billionaire assembles a powerful worldwide syndicate in order to find all of the soap and acquire it – either by simply purchasing it, or by using other kinds of financial leverage in order to ensure cooperation. Many members of the syndicate agree to participate in the billionaire's scheme despite their previous refusal, on ethical grounds, to participate in the marketplace for the soap.

In this story, there are important ethical differences between the actions of the participants in the marketplace for the soap. Most obviously, even on the assumption that

[64] Thanks to Ryan Jenkins for encouraging me to discuss different types of consumption activities.

no individual's actions will make any difference to human suffering, there is an important ethical difference between those who purchase the soap merely to ensure that it is respectfully destroyed, and those who purchase the soap and use it for purposes of distasteful consumption. Furthermore, there is an important ethical difference between the dealers who are willing to sell to the highest bidder regardless of the bidder's plans for the soap, and the other dealers who perform a similar market function, but only participate in the marketplace in order to acquire the soap for the billionaire. Perhaps most strikingly, the billionaire's actions are permissible, despite the fact that they significantly increase demand for the soap, which significant increases the market price for the soap, which significantly enriches those who sell the soap, including those who initially acquired the soap in abhorrent ways. Despite all of these undesirable effects, the billionaire's actions are permissible, even if the billionaire's actions cannot be expected to have any substantial effect on anyone's well-being. It is an important constraint on a theory of the ethics of the marketplace – and on a normative ethical theory in general – that it respect and explain these differences.

As a further clue to the underlying ethical facts, it is useful to note that there is a huge difference between the case just described and an otherwise analogous case in which the aggregate demand for the abhorrent product really does affect the amount of harm and violation of constraints that occurs in the background. For example, if the story above were redescribed so that the billionaire's action caused much more of the soap to be produced, perhaps from the bodies of current political prisoners, then the billionaire's actions would be wrong – and it would be also be wrong for the others in the story to traffic in the soap as well if their collective action would lead to a similar result.

In light of these observations, a more complete theory should distinguish between different consumption acts, and should also include additional parameters that track the important distinctions illustrated above, as in the following:

> Very strong ethical reasons not to engage in a consumption act A arise if, and then insofar as, it is essential to the relevant product P or the actual production of P that significant constraints are violated or significant harm is caused, and, if so, insofar as:
>
> > A involves (in decreasing order of significance) disrespectful, personal, or physical consumption of P, and
> >
> > in the context of A, there are no ethically weighty reasons to consume P in way A, and
> >
> > the relevant harm and violation of constraints associated with P-type products is caused by or (less significantly) counterfactually depends on the collective action of A-type consumption acts, and
> >
> > the footprint of harm and violation of constraints associated with a unit of A-type consumption of P-type products is large.[65]

[65] This analysis is intended to generalize in an obvious way to other collective action situations in which the actions of a single individual cannot be expected to make any difference to the outcome. A complete analysis would explain in detail how the ultimate degree of wrongness is a function of varying degrees of satisfaction of these (and perhaps other) conditions. I will not attempt to do that here, except to note that the conditions appear in roughly descending order of importance, insofar as they are of different levels of importance, and that dramatic reasons not to consume a product do not seem to arise from satisfaction of

This analysis captures and clarifies the intuitive thought that, unless there are strong countervailing reasons, it is wrong to consume a product when it is essential to that product that harm and violation of constraints lie behind it, especially insofar as the magnitude of the harm and violation of constraints is significant, and especially insofar as there is something disrespectful about the act of consumption itself.

My own view about the more controversial underlying explanation of all of this is that even when our individual decisions make no difference, others have a *claim* against us that we not engage in the activities outlined in the analysis above, where such claims are an unnoticed part of what it is to have more easily glossed *basic rights* such as the right not to be harmed without good reason, etc., and where such claims are possessed by, but are less stringent in, non-human animals than humans, where such stringency is a function of whatever ultimately explains the general differences in moral status between humans and non-human animals. As far as I can tell, it is no less plausible to think that animals have such claims against us than it is to think that very young, uncared-about children have such claims – and, as reflection on cases involving, say, child pornography reveals, it is hard to deny that such children have such claims. For example, unless we believe that such children have such claims, it is hard to see how we could *correctly* explain why it is wrong for an individual to watch freely-downloadable child pornography in the privacy of his home, given that such an individual's actions make no

---

the latter conditions unless the initial condition *that it is essential to the relevant product P or the actual production of P that significant constraints are violated or significant harm is caused* is satisfied.

difference to anything outside his home, including to child abuse. Such cases are, of course, simply further embarrassment for traditional consequentialism.[66]

This more complete analysis is able to explain in greater detail the spectrum of cases in which the consumption of ethically contentious products runs from acceptable to distasteful to clearly wrong to unconscionable. For example, in addition to explaining the wide range of cases in the examples involving the billionaire above, it provides a compelling explanation of why it is generally wrong to purchase and eat animals from factory farms. The explanation is that the actual production and distribution of animals from factory farms essentially involves the infliction of extreme suffering and the violation of significant constraints; as a result, it is usually wrong to purchase and eat animals from factory farms given that we usually lack strong overriding reasons for doing so, given that such consumption acts are personal, physical, and arguably even disrespectful, given that the relevant suffering and violation of constraints counterfactually depends on such consumption behavior, and given that the footprint of harm and violation of constraints associated with such consumption is large.[67] At the same time, the analysis also explains why eating meat acquired via dumpster diving is less objectionable than eating meat bought in a supermarket: the explanation is that

[66] For related discussion, see Joel Feinberg, "The Rights of Animals and Future Generations". Once such natural 'deontological' explanations are adopted for these cases, it is then easy to offer similarly natural explanations of other cases in which individual action intuitively doesn't make a difference – for example, the cases discussed in Jonathan Glover, "It Makes No Difference Whether or Not I Do It", and Derek Parfit, *Reasons and Persons*, Chapter 3. I discuss these issues in greater detail above, where I argue that Glover and Parfit's consequentialist explanation of such cases does not work.

[67] To say that suffering *counterfactually depends* on aggregate consumption behavior is not to say that the aggregate consumption behavior *causes* suffering – as noted above, aggregate consumption does not cause animal suffering, because animal suffering is caused on the 'supply side' by decisions of producers and also by governmental policies.

animal suffering counterfactually depends on aggregate consumption of the supermarket kind, but not on aggregate consumption of the dumpster diving kind; at the same time, the analysis also explains why there is still arguably something ethically objectionable about eating meat acquired via a dumpster, because such consumption still involves personal, physical, and arguably even disrespectful consumption of a product that was produced in a way that essentially involves significant suffering and the violation of important constraints. More generally, the analysis explains why, other things being equal, eating a vegan meal is less objectionable than eating a meal involving trace amounts of animal products as in bread and cereal, and why the latter is less objectionable than eating a salad containing small pieces of bacon, and why the latter is less objectionable than eating a steak, and why the latter is less objectionable than eating veal, etc.[68] As these cases show, the spectrum of cases involving consumption of animal products provides a useful test of the analysis above, and confirms its superior plausibility and explanatory power.

At an even more subtle level, the analysis also provides a satisfying account of complications that arise from the fact that agents typically perform multiple acts by means of a single course of action. For example, consider a case in which, after several days of backcountry hiking, we emerge from the wilderness to find a single store that sells only meat. Because we are famished, we have sufficient reason to eat meat on this

[68] The analysis can also explain why it would be much worse to eat factory farmed *humans* than factory farmed non-human animals, if we imagine for a terrible moment a scenario in which choice cuts of human flesh are available in supermarkets and produced and distributed in a way similar to factory farmed animal products. The explanation is that the production of factory farmed humans would involve the violation of much more stringent and dramatic constraints even if the facts about suffering were held constant, thereby ensuring more dramatic reasons not to consume the human flesh.

occasion, assuming that meat is the only nourishing food available in the vicinity. Nonetheless, if the store offers certified organic meat as well as factory farmed meat, we do not have good reasons to eat factory farmed meat rather than organic meat, assuming a typical difference in price between those types of meat. As a result, if we decide to purchase and eat the factory farmed meat, then although our act of *purchasing and eating meat* is unobjectionable, our act of *purchasing and eating factory farmed meat* is objectionable. Once again, the analysis above gets such subtleties right for the right reasons.[69]

In case the relevant facts regarding factory farming are unclear, it is worth noting explicitly that the actual production and distribution of animals from factory farms essentially involves the infliction of extreme suffering and the violation of significant constraints because of the production, confinement, and slaughter practices that are essential to factory farming. In particular, factory farming essentially involves causing intense suffering to almost all pigs and birds because of how those animals must be confined, handled, transported, and slaughtered in such operations, and it essentially involves causing suffering to an unacceptable number of cows because of how cows must be transported, maintained at feedlots, and then slaughtered in such operations. The main difference is that most factory farmed pigs and birds are confined for most of their lives in a way that causes intense suffering (because they are kept in enclosures too small to allow basic free movement), whereas cows tend to be raised in good conditions on

[69] As another example, if, as a token of his purported dominance over other human beings, a man chooses to wash his body with soap made from the bodies of executed political prisoners, then although his act of *washing his body with soap* is unobjectionable, his act of *washing his body with this soap* is seriously wrong.

ranches, and only encounter factory farming operations when transported to feedlots and then to slaughterhouses – and even then, significant suffering tends to be visited probabilistically on only some of the cows.[70]

The actual slaughtering methods for all of these animals cause suffering, because they inevitably result in an unacceptable number of the animals being dismembered while fully conscious – or, in the case of birds, drowned in boiling water while fully conscious. That is because slaughterhouses depend essentially on mass production processing lines in which it is obvious and unavoidable that a significant percentage of the animals are not be properly stunned before being dismembered, and 'factory farm' processing plants tend to insist on maximum line speeds for maximum economic efficiency, even though it clear to everyone involved that this dramatically increases the percentage of animals that are dismembered while fully conscious, and dramatically increases the occurrence of disabling injuries among the workers on the line, who are treated as expendable cogs in the processing plant machine. Clearly, the treatment of these humans and non-human

[70] An additional consideration is that it is quite essential to the production of meat that healthy animals are slaughtered in the prime of their lives for no compelling reason. Many believe that such unnecessary killing is itself deeply objectionable. However, the issue is somewhat opaque, and so I ignore it here, except to note that the arguments here only gain further force if one believes that such killing is itself wrong. For further discussion, see Tristram McPherson, "Why I am a vegan". The issues are difficult because, for one thing, if an animal has a life even minimally worth living, it could be argued that it is better off than it would have been if it had never been brought into existence. But more importantly, if animals are cared for under good conditions until they are slaughtered, it is plausible to argue that if they were capable of having reflective preferences, they themselves would prefer such a life to non-existence, and would even prefer such a life to a natural life that would predictably involve greater suffering and stress, and would predictably end in starvation, predadation, or some other painful death. (I have heard this line of argument suggested by many students and philosophers.) It is also worth noting that factory-farmed milk cows often share many of the life-long horrors of confinement and mistreatment that are visited upon pigs and birds. One interesting consequence is that it is not obvious that continuing to consume dairy products while giving up eating beef is ethically superior to giving up dairy products while continuing to eat beef.

animals involves the violation of significant constraints and the infliction of serious suffering.

In the future, it is possible that American agribusiness will move away from the factory farm paradigm just described, and so it is also important that the analysis above clarifies the way in which changes in production conditions would affect the permissibility of consuming animal products. As a realistic example, consider the case of American beef. According to cattle industry sources, there have been significant improvements in cattle slaughterhouse practices over the past decade, with the alleged result that very few cows are now slaughtered in a way that involves serious suffering. Over the same time period, there have also allegedly been significant improvements in the conditions under which cattle are transported and maintained at feedlots. If these alleged improvements are verifiable, then, by the analysis above, this removes much of the force of ethical objections to eating beef in America, although it does not, of course, undermine the objections to factory farmed chicken and pork.[71]

[71] Thanks to Travis Hoffman and Nathan Meyer for helpful discussion of these issues. Another relevant factor, the size of the animal consumed, is discussed by Joel MacClellan in "Size Matters: Animal Size, Contributory Causation, and Ethical Vegetarianism". To illustrate the relevance of animal size, note that the amount of suffering necessary to feed 10,000 people with a blue whale is much less than the amount of suffering necessary to feed that number with a suitably large number of pigs. The analysis above incorporates MacClellan's point via the final condition, because the 'welfare footprint' associated with a (serving of a) particular animal product will depend in part on the size of the relevant animal. When these ideas are conjoined with the alleged improvements in cattle slaughter and transportation conditions discussed above, and with the relatively humane conditions under which cattle are generally raised on ranches, it is easy to imagine an argument that it is permissible to eat beef on the grounds that the welfare footprint associated with a serving of beef is sufficiently small, even if the welfare footprint associated with pork and chicken is impermissibly large. Of course, whether such an argument would be sound is a further question – at the very least, such an argument seems to overlook the environmental footprint associated with a serving of beef, which is arguably impermissibly large. Such environmental worries also explain why consumption of mollusks and other similar animals is ethically problematic even on the assumption that such animals are physically incapable of suffering. Along related lines, at the level of advocacy it

In sum, the discussion and analysis above clarifies what consumers are required to do when products are produced in morally objectionable ways, and reveals that the correct explanation is more subtle than it initially appears. At the same time, the analysis is consistent with intuitively appealing ideas such as that the most decisive fact relevant to the ethics of eating meat is that a person's gustatory pleasure is of little ethical significance compared to the suffering that animals must experience in the service of that pleasure. The important philosophical point is that such facts do not lead to conclusions about the ethics of the marketplace in the simplistic way that theorists have imagined, but must instead be marshaled into more subtle arguments.

The preceding discussion helps to clarify the conditions under which consumption activities are permissible when the collective result of those activities is in some way ethically undesirable, and illustrates the way in which even the most subtle and difficult ethical issues can be illuminated by an exacting application of the Minimal Analysis. The main virtue of that analysis is to focus our attention on the subtle welfare-based and non-welfare-based reasons that are ultimately the key to thinking correctly about such issues. After carefully investigating the contours of such reasons, no reason has emerged for doubting the Minimal Analysis, or for thinking that it is merely an ad-hoc device that has been crafted only to justify idiosyncratic conclusions. As a result, careful reflection on the Minimal Analysis should increase our confidence that individuals are not required to reduce their emissions by a substantial amount, because it is not *essential* to the kind of

---

seems worth striving to make the welfare footprints of various products readily available in the same way that environmental activists strive to make information about environmental footprints readily available – perhaps even making such information a mandatory part of the labeling for particular products.

activities that generate emissions that anyone is harmed or that any important constraints are violated, and because significantly reducing emissions involves a much more substantial sacrifice of personal well-being than, say, foregoing factory-farmed meat or sweatshop clothing.

An even stronger argument arises from a principle that could be true even if the Minimal Analysis were false:

Non-Cooperation Principle

If you face a collective action problem in which you know that most people will defect, thereby ensuring a collectively bad outcome, you are not required to cooperate if you know that:

(i) cooperating would require significant sacrifices of personal well-being, and

(ii) there are no significant welfare-based reasons to cooperate, and

(iii) there are no significant rights- or respect-based reasons to cooperate, and

(iv) you do not owe it to the others to cooperate, and

(v) it is not *essential* to non-cooperative action (in the sense discussed above) that harm or rights-violations lie in the background, and

(vi) most of the others would defect even if they were in your exact situation, and

(vii) cooperating would not allow you to be part of a successful solution to a different collective action problem.

It is difficult to see how this principle could be false, because it captures what is most compelling about the idea that individuals are never required to be *suckers* – and this principle together with the facts of our situation with respect to climate change implies that individuals are typically not required to reduce their emissions by a substantial amount.

It is important to see that the Non-Cooperation Principle does not imply that free riding is permissible, because in cases where free riding is intuitively impermissible, you owe it to the others to cooperate, which means that condition (iv) fails to hold. As a result, the principle is consistent with the view that you would be required to reduce emissions by a significant amount if sacrifices by a significant number of others ensured a sufficiently good emissions outcome; however, because the actions of others are not currently or foreseeably securing a sufficiently good outcome, you are not required to make such sacrifices as things stand.[72]

[72] As Robert Nozick has shown, free-riding is sometimes permissible. For example, suppose that your neighbors spend a very large amount of time cleaning the sidewalks around your houses – in particular,

A possible objection to the argument from the Non-Cooperation Principle is that condition (vii) fails to hold in connection with our situation with respect to climate change, because by reducing emissions you would succeed in solving a different collective action problem: namely, the problem of reducing emissions by an amount greater than r, where r is the amount that emissions are reduced by the small proportion of other people who are reducing their emissions. As a result, it could be claimed that this undermines the argument from the Non-Cooperation Principle that you are not required to reduce your emissions.[73]

To see the problem with this objection, consider how it would apply to the asteroid case discussed in the first section of this paper. In that case, it is also true that if you cooperated you would all succeed in helping people to a degree greater than d, where d is the degree that the other cooperators will succeed in prolonging life on Earth. Nonetheless, this does not undermine the idea that you are not required to cooperate in that case, even though the good you can do with the other cooperators in that case is even more significant than the good you can succeed in doing with those who reduce

---

imagine they get down on their knees and use toothbrushes, etc. to ensure that the sidewalks are as clean as humanly possible. In such a case, you are not required to help pay for this excessive maintenance even if you enjoy having the sidewalks so clean, and thus benefit from their cooperative actions; similarly, your neighbors would not be justified in forcibly taking money from you to pay 'your share' of such a project. Nozick uses cases like this to show the falsity of the principle (due to H.L.A. Hart and endorsed by John Rawls) that if you benefit from the cooperative actions of others, you are required to pay your proportional share of the costs, and/or the state is justified in extracting those dues from you by force. (See *Anarchy, State, and Utopia*, pp. 90-95.) A more plausible principle regarding free-riding cases is suggested by my discussion above: free-riding is impermissible if and only if you owe it to the others to cooperate. In the Nozick-style case just discussed, you do not owe it to the others to cooperate in their sidewalk-cleaning scheme, and so you are not required to cooperate. (Of course, this leaves unanswered the question of when exactly you owe it to others to cooperate.)

[73] Thanks to Heather Berginc, Tim Clark, and Chris Griffin for encouraging me to discuss this objection.

emissions by a significant amount. This shows that the current objection does not undermine the idea that you are required to reduce your emissions by a significant amount, just as it does not undermine the idea that you are not required to travel to Kansas in the asteroid story discussed in the first section. As far as I can tell, the general lesson of this objection is that for purposes of the Non-Cooperation Principle, the problem of reducing emissions by an amount greater than r is not a *different* collective action problem in the sense relevant to condition (vii) than the more general problem of reducing emissions to combat climate change.

At this point, I am unaware of further objections to my arguments, and so I conclude that you are not required to reduce your emissions by a substantial amount. Perhaps the most rigorous argument I've given for this conclusion is the one that relies on the Non-Cooperation Principle. However, even if there is a problem with that argument, there are other arguments for the same conclusion that are difficult to resist, such as the intuitive argument by analogy discussed in the first section of this chapter.

It is also worth noting that none of the arguments in this chapter depend on subtleties of our situation with respect to climate change such as our prior ignorance of the facts about climate change, our uncertainty about cause and effect relationships, our uncertainty about the long-run effects of climate change, or complications regarding the identity of persons in future generations. Instead, my arguments show that even if we ignore those issues there is still decisive reason to think that individuals are not required to reduce emissions by a significant amount. If those subtleties are taken into account, then there is even more reason to think that the conclusions defended in this chapter must be true.

4. The Political Significance of Futility

In the previous sections I argued that individuals are not required to reduce their emissions by a significant amount. In this section I explain why, at the same time and partly as a result, individuals are required to favor effective intergovernmental solutions to climate change (on the assumption that climate change is a serious problem, and on the assumption that there are effective intergovernmental solutions available, despite the collective action problem that arises at the level of nations), and I explain why this makes it the case that nations are justified in adopting such solutions, even if those solutions are quite costly, unfair to their citizens, and could be reasonably rejected by their citizens.[74]

More generally, in this section I show that it can be justifiable for a nation to act in a way that makes its citizens worse off, that is unfair to its citizens, that its citizens would not agree to, and that its citizens could reasonably reject. Such a situation arises most clearly when the citizens of a nation are required to favor a national course of action that is unfair to them, contrary to their own interest, and contrary to the interest of future

[74] When I say 'individuals are required to favor effective intergovernmental solutions to climate change', this should be understood as elliptical for 'individuals are required to favor effective intergovernmental solutions to climate change *that are not dramatically inferior to other realistic and effective intergovernmental solutions that might well be adopted instead*'. An overview of the arguments that climate change is a serious problem is provided by David Archer, *The Long Thaw*, which summarizes the more authoritative and complete arguments in the Intergovernmental Panel on Climate Change, *Climate Change 2007: The Physical Science Basis*. In the first chapter of this dissertation I argue that there are effective intergovernmental solutions available even in our imperfect world; see also Eric Posner and David Weisbach, *Climate Change Justice*.

generations of their descendants and fellow citizens. The fact that governmental action can be justifiable under such conditions shows that many political theories are false, because many political theories entail that it is impossible for governmental action to be justifiable under such conditions. The practical upshot is that nations can be justified in adopting effective intergovernmental solutions to climate change even if such solutions are costly, unfair, not to their advantage, and unacceptable to their citizens. In what follows, I show that this is true on all plausible political theories, including all plausible contractualist theories (e.g. Rawlsian and Scanlonian contractualism) and rights-based theories (e.g. Nozickian libertarianism).

To get some initial intuitive bearing on these issues, and to illustrate what our default view ought to be regarding the political ethics of climate change, imagine that a large ship from a distant nation has wrecked in the waters outside your community, and that the survivors will drown unless they are quickly saved. Furthermore, the survivors cannot be saved by military and law enforcement alone, because there are far too few government boats and helicopters available to save them. However, all of the survivors can be saved if most of your community's privately-owned boats are quickly deployed to the scene of the shipwreck. As it happens, because of an obscure but respected maritime law, you alone have the legal authority to 'federalize' and deploy the community's privately-owned boats, because you are the Commodore of the local yacht club; furthermore, you have the ability to deploy the boats quickly, because they almost all docked in a marina that you control, and you can easily draft skippers to pilot the boats as the result of a large conference that is currently underway at the marina. However, the government will maintain control over the general rescue operation, and as you review their plan of action,

it is clear that their plan is unfair to the owners of private boats, because the government is insisting that private boats bear the burden of all of the 'close-in' rescue operations, which will foreseeably lead to damage to the boats involved, which you know will not be covered by the insurance policies maintained by the owners of the boats. Unfortunately, it is clear that the government officials are unwilling to negotiate on this or any other aspect of their planned response to the crisis, and so there is no way of changing the plan to make it more fair to the owners of the boats. Instead, it is clear that if you do not accept the terms of the government's rescue plan, the result will be that many of the survivors drown who would otherwise be saved.

In this story, you are justified in deploying the privately owned boats, even if the boat owners bear no responsibility for the shipwreck, even if such a response is unfair to them, even if such a response is contrary to their interests, and even if you know that most of the boat owners would never agree to aid foreigners in such a way. By analogy, the default view regarding the political ethics of climate change ought to be that just as you are justified in such actions in the story above, given that you must cooperate with others who are ethically-deficient in order to avoid a catastrophe, so too a nation can be justified in adopting an analogous course of action, if it must cooperate with other ethically-deficient nations in order to avoid a catastrophe.

This is even more clear when nations also bear *responsibility* for the existence of the problem in the first place – as in our actual situation with respect to climate change. To make the shipwreck case analogous in this way to our actual situation with respect to climate change, we would have to imagine, for example, that a yacht race organized by

your yacht club crisscrossed the shipping lanes outside your city in a way that largely caused the shipwreck. If that detail is added to the case, then it is obvious that you would be required to deploy the yacht club's boats to rescue the survivors, even if the expected cost to yacht club members was significant, and even if members balked at such a rescue operation.

The most fundamental explanation of these cases is that, other things being nearly equal, when there is a particular course of action that every member of a collective is *required to favor*,[75] then the collective is justified in adopting that course of action, even if that course of action is unfair to each of its members and contrary to their interests. This general ethical fact implies that something like the following principle is true in the political realm:

Ethical Principle

> If citizens are required to favor a particular governmental act, then it is required that the government perform that act, and such an act is justified.[76]

[75] Here and in the rest of the paper, I use 'S is required to favor p' as shorthand for something like 'If S were idealized in ethically relevant ways, S would be required to be in favor of p occurring in S's actual situation, even if S is not actually required to recognize or vocalize this fact, etc.'.

[76] Because nations are legally permitted to enter into the kind of intergovernmental agreements that offer the best hope of effectively combating climate change, it would not affect my argument if one thought (as I do not) that a government must also be *legally entitled to perform an act* in order for that act to be justified. (Thanks to Gideon Rosen for this point.)

Many theories of political justification do not endorse anything like this principle, and are implausible as a result. For example, there are many contractualist theories that do not endorse such a principle, and that instead attempt to explain the entire realm of political justification with a principle like the following:

Contract Principle

> If citizens would agree to (principles that require) a particular governmental act under suitable idealizing conditions, then it is permissible that the government perform that act, and such an act is justified.

The problem with the Contract Principle is that when it is taken to explain the entire realm of political justification it has unacceptable consequences for non-citizens. For example, consider a government that enslaves and oppresses non-citizens whenever it is advantageous to do so. Clearly, such action would be monstrous and unjustified. However, theories that rely on the Contract Principle alone tend to imply that such action is justified. For example, consider John Rawls's theory in *A Theory of Justice*. According to that view, a government's current citizens are the only parties to the relevant agreement, and they have no particular interest in the good of other people. As a result, if their deliberations are not constrained in some way that goes beyond the Contract Principle, then it is clear that they would adopt principles that imply that non-citizens

should be enslaved and oppressed whenever it is to the advantage of citizens to do so. In fact, it is much clearer that they would accept principles that have that implication than any of the other things that Rawls claims they would accept.[77]

Some contractualists are inclined to respond to these problems by claiming, roughly, that all persons on the planet, including all future persons, should be understood as parties to the relevant agreement. This leads to a version of cosmopolitanism that mandates massive redistribution of wealth and other radical sacrifices of national self-interest.[78] Many people find such a theory implausible, and so I will not consider it further here, except to note that if such a theory were true, my conclusions about the permissibility of governmental action regarding climate change would also be true for obvious reasons.

Rawls's own response to this problem is to introduce a 'two-level' contractualist theory in *The Law of Peoples* in which the relations between nations are governed by a second social contract at the level of nations. Unfortunately, Rawls never really explains how the two-level theory is supposed to solve the relevant problems, in part because he never provides a sufficiently detailed explanation of how to adjudicate tensions between the two levels, especially in the sort of 'non-ideal' cases that are commonplace in the real world in which some nations and individuals are known to act in unjustified ways.[79] This

[77] Citizens in the decision situation Rawls describes have no particular interest in the good of other people, and so would not recognize any reason to extend basic rights to non-citizens, and so would not agree to extend such rights to non-citizens.

[78] For cosmopolitan views of this type, see Charles Beitz, *Political Theory and International Relations*, and Thomas Pogge, *Realizing Rawls*. For important related discussion, see Martha Nussbaum, *Frontiers of Justice*.

[79] For discussion, see Martha Nussbaum, *Frontiers of Justice*, chapters 4 and 6, esp. pp. 262-264.

problem for Rawls's account seems particularly insurmountable with respect to challenges such as climate change, in which a collective action problem arises at the level of nations, and different nations stand to gain or lose very different amounts, and different nations are disposed to cooperate in solving the problem to very different degrees, and different nations bear very different and often unclear levels of responsibility for the problem, and different nations are disposed to conform to very different degrees to the demands of justice. In such situations, Rawls's theory seems incapable of providing any guidance whatsoever as to what kind of governmental actions would be justified or unjustified.

This illustrates a characteristic weakness of Rawls's theory and many other major political theories: the theories provide no guidance on important real-world problems, while offering only dubious explanations of things that are already obvious to all reasonable people. As an illustration, consider the obvious fact that governments are justified in enacting laws that prohibit cruelty to animals. Intuitively, this is because animal cruelty is wrong and animals are so defenseless that there could be no ethically defensible reason for opposing such laws. This intuitive explanation is in perfect alignment with the Ethical Principle above. However, many contractualist theories do not endorse the Ethical Principle and can only attempt to explain why such laws are justified via theoretical epicycles that are an embarrassment and ultimately implausible, given that animals are not plausibly thought of as parties to the relevant contractualist agreements.[80]

[80] To illustrate just one aspect of the problem with thinking of animals as parties to the relevant agreements: if animals were parties to the agreements, then we would have to devote the same resources to the care and

A more promising way for contractualists to respond to these problems is to attempt to idealize the parties to the relevant agreement in such a way that they would not agree to offensive principles like enslaving non-citizens whenever it is to their advantage. This strategy appears to be successful if and only if it is equivalent in practice to endorsing the Ethical Principle above. Because this seems like the only plausible way of responding to these problems while leaving much of the rest of standard contractualist views unchanged, this shows that standard contractualist theories of political justification are implausible unless they endorse the Ethical Principle or some other theoretical device that is equivalent in practice.

To illustrate the way in which contractualist theories might endorse a theoretical device that is equivalent in practice to the Ethical Principle, consider a Scanlonian contractualist theory that claims that governmental action is justified if and only if it is permitted by principles that citizens cannot reasonably reject.[81] If such a theory is to avoid the problems regarding non-citizens discussed above, then the occurrence of 'reasonably' in the theory must be understood in a way that is not restricted to merely prudential reasons, but includes ethical reasons, and takes the balance of ethical reasons to be all-things-considered decisive. That is because if 'reasonably' is understood only in terms of prudential reasons, then citizens can reasonably reject the principle that *non-citizens cannot be enslaved even if it is to the advantage of citizens*, because rejecting that

---

well-being of many animals as we devote to human citizens that are mentally handicapped. Most people find this idea absurd.

[81] Such a political theory has some similarity to Thomas Scanlon's ethical theory in *What We Owe to Each Other*.

principle is what they have most prudential reason to do. Instead, the only way that the balance of reasons will always tell in favor of that principle is if ethical reasons are included in the calculation and taken to be decisive. Once this is noted, it becomes clear that the mechanism by which the problems involving non-citizens are intended to be solved by Scanlonian contractualism is equivalent in practice to endorsing the Ethical Principle above.[82]

The important upshot is that contractualists should endorse the Ethical Principle, and should therefore embrace the idea that there are cases in which governmental action is justified even when it is unfair and makes citizens worse off, even when citizens would not agree to such action, and even when citizens can reasonably reject such action both from a prudential point of view and from an all-things-considered point of view that does not take ethical reasons as decisive. Depending on the costs of an effective solution to climate change and the dispositions of other nations, our situation with respect to climate change might be an example of such a situation. In any event, regardless of whether our actual situation is an example of such a case, the practical upshot remains that on any plausible contractualist view, rich nations are justified in adopting effective intergovernmental solutions to climate change because their citizens are required to favor such solutions, given that catastrophic harm will otherwise befall others largely because

[82] More explicit reliance on the Ethical Principle would also mitigate (but not entirely remove) the need for contractualists to provide a 'non-ideal theory' in order for their view to have any relevance to the real world. Here it is worth noting that contractualists like Rawls and Scanlon already explicitly appeal to a range of ethical facts as explanatorily prior inputs to their contractual reasoning – namely, facts about what the original position must be like in order for choices in that situation to be fair and ethically illuminating, and facts about what we have ethical reasons to do, respectively.

of their earlier actions, and because the cooperation of rich nations is clearly essential to any successful solution to the problem.

Hard-headed realists are apt to respond to all of this by insisting that such reasoning is naïve, and that the only thing that really matters for determining what nations are justified in doing is what it is in their interest to do. One way of responding to this brand of hard-headed realism is to note that it implies that we would be justified in enslaving non-citizens whenever it is to our advantage, which is clearly an absurd and unbelievable consequence of any view. More importantly, even if some hard-headed realists have a more sophisticated theory that, at least in their own minds, does not have such embarrassing consequences, it is sufficient to note that any version of hard-headed realism that is inconsistent with the intuitive reasoning about climate change above implies that in the shipwreck case above you should refuse to save the drowning sailors if it is not to the advantage of your yacht club to do so, even if the shipwreck was caused by your yacht club. Because such a consequence is itself absurd, it follows that any brand of hard-headed realism has absurd consequences if it is inconsistent with the reasoning regarding climate change above.[83]

At first glance, it might seem that Robert Nozick's theory in *Anarchy, State, and Utopia* has the same problem as hard-headed realism, because it might seem that Nozick's theory implies that deploying the fleet in the story above would be unjustified, even if the shipwreck was caused by the yacht club, because it amounts to infringing the property

[83] Perhaps the ultimate response to hard-headed realism in this context is to show, as in the first part of this dissertation, that there is a realistic way of making it in the interest of every nation to agree to an effective intergovernmental climate treaty.

rights of others without their consent. In response, defenders of Nozick's view might insist that the shipwreck situation constitutes an emergency in which property rights can justifiably be treated as less stringent than they normally are. If that is right, then it follows by similar reasoning that even on Nozick's view governmental action can be justified in response to climate change even if citizens wouldn't agree to such action, and even if such action would be redistributive and unfair, given that climate change is an emergency, and that there are effective intergovernmental solutions that can be adopted by governments.

From a theoretical perspective, a better response is for defenders of Nozick's view to explicitly endorse the Ethical Principle above, because that principle offers a more compelling and less ad hoc explanation of cases of emergency and other related cases in which property rights can justifiably be treated as less stringent than they normally are, and, more importantly, because Nozick's main arguments in *Anarchy, State, and Utopia* are already committed to the truth of the Ethical Principle. In particular, Nozick tacitly relies on the Ethical Principle when he argues that a government could arise naturally by a series of individually justifiable steps from a state of nature. The crucial part of the argument is the explanation of how a protective agency could enforce a monopoly on the use of force in a region in a way that was justifiable. In order for the enforcement of such a monopoly to be justifiable, the agency must be willing to extend protection to everyone in its region, including those who are unable to pay for its protection; however, that means that there must then be some redistribution of wealth, since those who can afford to pay for protection will have to pick up the tab for those who cannot. However, it is hard to see how such redistribution could be justifiable on Nozick's view, given that there

is no emergency in such a case. Nozick's solution to this problem is to claim that citizens would be *required* to favor the free provision of protection to those who cannot afford to pay once the agency declares a monopoly on the use of force in the region, and that this explains why the protective agency is justified in forcing the wealthier to pay for protection for those who cannot afford it.[84] This reasoning tacitly invokes the Ethical Principle – and it also seems to be exactly the right thing to say about such a case, setting aside other aspects of Nozick's view, and setting aside what Nozick is inclined to say about other cases. However, this undermines Nozick's idea that infringement of rights is always impermissible except in the face of an emergency, because there is no emergency in such a case, but yet the infringement of rights is required nonetheless. This shows that Nozick's own arguments entail that the infringement of rights can be permissible even in non-emergency situations whenever there are sufficient ethical reasons for such infringement – just as the Ethical Principle entails, and just as is intuitively correct.[85]

As a result, the most plausible way to develop a rights-based political theory like Nozick's is to explicitly endorse the Ethical Principle above, together with a modified

[84] "[I] argue that the transition from an ultraminimal state to a minimal state morally must occur. It would be morally impermissible for persons to maintain the monopoly in the ultraminimal state without providing protective services for all, even if this requires specific 'redistribution'" (*Anarchy, State, and Utopia*, pg. 52; see also chapter 5). More explicit reliance on the Ethical Principle would also mitigate (but not entirely remove) the need for neo-Lockeans like Nozick to provide a 'non-ideal theory' in order for their view to have any relevance to the real world, given that in the real world facts about the historical acquisition and transfer of property are usually opaque and almost always ethically problematic.

[85] In "Justice and Climate Change", Dan Shahar notes that rights-based political theories can also deliver plausible verdicts on many climate change issues if they can rely on the premise that victims of climate change have their rights violated by those who generate emissions. If my arguments above are correct, then this explanation is unavailable, because if the arguments above are correct then it is false that victims of climate change have their rights violated in any important way by those who generate emissions. (Shahar's discussion is agnostic about whether rights are violated in such a way, and about whether the possible explanation he notes is ultimately a plausible explanation.)

and more plausible view about the nature of rights, on which rights are *defeasible* side-constraints that play a crucial but *defeasible* role in determining what government and individuals can justifiably do.[86] The result of such a modification is a more appealing political theory that softens the edges of Nozick's avowed brand of libertarianism by allowing that redistributive aid is permissible even when there is no pressing emergency, provided that the aid will succeed in doing dramatic good at relatively little cost to the wealthy, and perhaps provided that other important ethical conditions are met. (It is of course an empirical question how often such conditions are satisfied in the real world.)[87]

The important practical upshot of all of this is that on any plausible political theory – including consequentialist, realist, contractualist, cosmopolitan, and rights-based theories – nations can be justified in adopting effective intergovernmental solutions to climate change even if such solutions are costly and unfair to their citizens, and even if such solutions would make their citizens worse off and would be rejected by their citizens, assuming that climate change is as serious of a problem as scientists claim, and assuming that there are effective intergovernmental solutions available, despite the collective action problem that arises at the level of nations.[88]

---

[86] For an elegant and plausible general theory of the nature of rights that vindicates the idea that rights are *defeasible* side-constraints as described here, see Judith Jarvis Thomson, *The Realm of Rights*.

[87] My arguments here largely follow those of Judith Jarvis Thomson in "Some Ruminations on Rights", and, at a more theoretical level, in *The Realm of Rights*. It would be hard to overstate the influence Thomson's arguments have had on my thinking about the issues discussed in this paper, especially in this final section. In those works, Thomson provides more extensive and more decisive arguments for many of the claims made in this section in connection with rights-based views.

[88] In the first part of this dissertation I argue that there are effective intergovernmental solutions available even in our imperfect world; see also Eric Posner and David Weisbach, *Climate Change Justice*. (Thanks to Chris Griffin for encouraging me to clarify the way in which the conclusions in this section are dependent

Consequentialism, the Inefficacy Objection, and the Ethics of the Marketplace

Worries about inefficacy arise in almost every area of practical ethics, and threaten to undermine many normative theories. A careful examination of inefficacy in this chapter reveals that existent theories are unable to offer a plausible account of what individuals are required to do in the kind of collective action problems that are common in a market-based society, and are therefore unable to explain many of the most important facts about modern moral life. In particular, results from economics and political science that have been unappreciated by philosophers help to clarify that for straightforward mathematical and empirical reasons an appeal to expected consequences cannot possibly deliver the verdicts on such cases that consequentialists assume. In addition, alternative theories that appeal to 'universalizability', 'direct harm', and other notions also cannot deliver plausible verdicts on such cases. For these and other reasons, a plausible account of what individuals are required to do in a large marketplace must invoke the distinction developed in the previous chapter between activities that are essential to a product or to the actual production of a product, and activities that are not.

---

on the assumption that there is are effective intergovernmental solutions available, despite the collective action problem that arises at the level of nations.)

To begin the discussion of inefficacy and individual-level requirements, note that we would be healthier if we dramatically reduced our consumption of animal products, and any desirable level of animal consumption could be easily sustained without factory farming; furthermore, eliminating factory farms would save billions of animals from intense suffering each year, improve public health, reduce food prices in lesser developed countries, reduce energy costs by reducing the price of biofuels, reduce air and water pollution, reduce the irrational consumption of scarce groundwater resources, reduce greenhouse gas emissions, reduce health-care costs and thereby reduce national deficits, and allow family farmers to compete more effectively against agribusiness conglomerates.

In light of the preceding, and the fact that governments could easily adopt efficient regulations that would largely eliminate factory farming without causing any ethically weighty negative side effects, governments should adopt such regulations.[89] However, in

[89] In my paper "The Case for a Revenue-Neutral Meat Tax", I outline the virtues of one possible set of regulations: first, levy a flat tax on animals slaughtered at USDA-regulated facilities based on a fixed percentage of the animals' wholesale price; second, waive the tax when animals are certified organic and raised in a way that is humane by common-sense standards; finally, refund the proceeds from the tax on an equal per-taxpayer basis. This would create incentives for Americans to reduce their unhealthy overconsumption of meat without forcing anyone to change their diet, and it would do this without introducing any new bureaucracy, because slaughterhouse production numbers are already closely monitored by the USDA, and so the prospect of significant and likely fines would ensure near-universal compliance without additional oversight. Most importantly, the tax is ethically and economically desirable because it serves primarily to correct externalities in the current marketplace – in other words, the tax does not funnel money into the pockets of government or special interests, but merely corrects a flaw in the marketplace whereby some people are not paying the real costs of their consumption, while others are forced to pick up the tab in the form of higher health insurance premiums, higher taxes, and other undesirable effects, such as higher levels of pollution – not to mention the costs imposed on animals raised in factory farms. For a different proposal for a meat tax, see Peter Singer, "Make meat-eaters pay", *New York Daily News*, 25 October 2009.

the meantime factory farming continues. Does this mean that each of us is required to stop consuming animal products from factory farms, at least when we can do so without dramatic inconvenience?

At first glance, it might seem that the answer to this individual-level question is 'yes', just as the answer to the collective-level question of whether we should eliminate factory farms is 'yes'. However, as in the previous chapter, on further reflection the question of what individuals are required to do about factory farms is more complicated than it initially appears – and, for similar reasons, most questions about what individuals are required to do when products are produced in a morally objectionable way are more complicated than they initially appear. These complications reveal that existent normative ethical theories are unable to offer a plausible account of what individuals are required to do in the kind of collective action problems that are ubiquitous in a market-based society, and are therefore unable to explain many of the most important facts about modern moral life.

To see how things get complicated, it is useful to begin by examining utilitarian arguments that it is wrong to eat meat from factory farms, such as those offered by Peter Singer. According to Singer, eating meat from factory farms is wrong because it has unacceptable consequences on balance for welfare. For example, if I eat a steak from a factory farm, Singer would claim that my gustatory pleasure is greatly outweighed by the suffering that the cow experiences in order to bring me that pleasure; as a result, Singer would claim that the effects of my eating a steak from a factory farm are unacceptably

negative on balance, even if I really enjoy eating steak – and this means that it is impermissible for me to eat a steak from a factory farm.

At first glance, Singer's reasoning can seem decisive even if one endorses an ethical view other than utilitarianism. That is because on any view, the welfare effects of one's actions are ethically important. In particular, even on deontological theories and virtue theories, it is generally wrong to bring about significant harm or suffering for no significant reason, even if doing so would not violate any rights or agreements. For example, on any plausible view, if I know that bad weather has destroyed the food supply for animals in my area except for some berries that grow deep in the woods on my property, it would be wrong for me to burn all of those berries in my outdoor fireplace, where their combustion would provide a very fleeting pleasant aroma, if I know that this would cause all of the animals in my area to starve to death painfully. As a result, if Singer's arguments about eating meat are sound on the assumption that utilitarianism is true, then it seems like they are also sound on the assumption that any other plausible ethical theory is true, because if Singer is right about the welfare effects of eating meat, those considerations would trump any other considerations that a plausible ethical theory could say are relevant to determining what we are required to do.[90]

[90] It could be objected that there are good reasons for consuming meat from factory farms because we need meat for nutritional purposes, and because factory-farmed meat is cheaper than organic meat. However, both of these claims are false. In fact, the evidence suggests that for people who live in developed countries with easy access to a wide variety of fruits, vegetables, legumes, etc., consuming animal products is analogous to consuming alcoholic beverages that are not necessary for our nutritional well-being and are in fact toxic to our bodies, at least in non-trivial quantities. (For discussion of these issues, see T. Colin Campbell, *The China Study*.) To see why this undermines the idea that there are good reasons for consuming meat from factory farms, imagine that there are two types of intoxicating substance: one is produced in a normal way, and the other is produced using the slave labor of children. If the slave labor variety is merely cheaper, that does not mean that there is then a good reason for consuming it – because

Unfortunately, Singer's account of the welfare effects is mistaken, because even if one agrees (as one should) with Singer's premises about the magnitude of animal suffering and the comparative unimportance of gustatory and other human pleasures, his conclusion about the welfare effects of consumption by individuals does not follow, and, upon careful reflection, turns out to be false. That is because an individual's decision to consume animal products does not really have any effect on the number of animals that suffer or the extent of that suffering, given the actual nature of the supply chain that stands in between individual consumption decisions and production decisions; at the same time, an individual's decision to consume animal products does have a positive effect on that individual's own welfare.[91] As a result, Singer's premises about animal

---

there is no significant reason for consuming intoxicating substances in the first place, and so the fact that a particular type of that substance is cheaper than the others does not amount to a good reason for consuming it, especially if it is produced in a way that is more ethically objectionable than the slightly more expensive alternatives. It could also be objected that there are good reasons for eating meat on aesthetic grounds, because meat is an essential part of 'sophisticated' culinary dishes, etc. The problem with this objection is that it mistakenly assumes that aesthetic experiences that are fleeting, easily replicable, and intellectually insignificant can provide good reasons for torturing animals – which is surely false. For example, suppose that a distinctive odor is released when a particular species of pig is slowly burned alive in an outdoor fire pit, and that some people find this odor to be 'sophisticated' and a good aromatic match for a variety of fine wines. Nonetheless, the prospect of such an insignificant aesthetic experience could not provide good reason to slowly torture a live pig to death in such a way. To put the point another way: while it is arguable (but not obvious) that a pig should be tortured to death if that is the only way to produce a great work of art of everlasting importance, it cannot sensibly be maintained that the shallow and fleeting aesthetics of a fine meal are of sufficient importance to justify such cruelty.

[91] At least insofar as individuals' are made better off by having their preference to eat meat satisfied. This qualification shows that there is conceptual space for an interesting *paternalistic* welfare-based argument against eating animal products: it could be claimed that consumers are so radically and systematically mistaken about the bad health effects of eating animal products that eating such products is wrong because of its negative effect on consumers' *own* welfare. This is actually a more plausible *utilitarian* argument than standard utilitarian arguments that rely on considerations of animal welfare. Unfortunately for Singer, his brand of utilitarianism not amenable to this sort of argument (and is less amenable than more 'classical' forms of utilitarianism), because he takes the satisfaction of individuals' actual preferences as much more important than other, more 'physical' and 'classical' aspects of well-being. As a result, if Singer's utilitarian theory were modified to make it amenable to this sort of paternalistic argument, then a change in view would also be required on many related issues in which considerations of paternalism arise, such as

suffering and human pleasures, together with the actual empirical facts about the workings of the marketplace, entail that the welfare effect of an individual's decision to consume animal products is *positive* on balance, in contrast to what Singer assumes. This undermines the idea that considerations of welfare tell against eating meat from factory farms, and it shows that Singer's utilitarian view actually implies that most individuals are *required* to consume animal products from factory farms, which is obviously the opposite of what Singer and most other utilitarians want to believe.[92]

Because this argument flies in the face of conventional wisdom, especially among moral philosophers, it is worth taking some time to verify that it is sound. As a first step, note that everyone can agree that there is a dramatic ethical difference between the following two ways of consuming a T-bone steak: in the first case, a dumpster diver snags a T-bone steak from the garbage and eats it; in the second case, a man enjoys a T-bone steak at Jimmy's You-Hack-It-Yourself Steakhouse, where customers brutally cut their steaks from the bodies of live cows, which are kept alive during the excruciating butchering process. (Once a cow bleeds to death, customers shift their efforts to a new live cow.)[93]

---

euthanasia – for example, the resulting view would then presumably imply that it is permissible to euthanize fully conscious adults whenever their lives are not worth living, even when they explicitly insist that they want to continue living.

[92] This sort of 'inefficacy' problem for consequentialism is familiar: for previous attempts to develop it into an objection to utilitarian arguments that it is wrong to eat meat, see R.G. Frey, "Utilitarianism and Moral Vegetarianism Again: Protest or Effectiveness", Russ Schafer-Landau, "Vegetarianism, Causation and Ethical Theory", Tristram McPherson, "Why I am a Vegan", and many others; previous developments of the inefficacy objection have been indecisive because they do not explain what is wrong with the various common replies on behalf of utilitarianism, which I discuss below.

[93] Those who find this example offensive probably would also find it offensive that some cows are dismembered while fully conscious because of mistakes made in the stunning process at slaughterhouses. Although some such mistakes are inevitable, the actual number of such mistakes is inexcusable, because

Everyone can agree that enjoying a steak at Jimmy's You-Hack-It-Yourself is objectionable, whereas enjoying a steak acquired through dumpster diving is relatively unobjectionable, especially from a traditional consequentialist perspective.[94] On any ethical view, at least part of the difference is that the welfare effect of eating a steak at Jimmy's is substantially negative, whereas a dumpster diver's consumption has no negative effect on welfare; according to welfarists like Singer, that is the only important difference between the two cases.

But now consider the question: when individuals purchase animal products from factory farms, are the welfare effects of their consumption decisions more like buying a steak at Jimmy's You-Hack-It-Yourself, or more like acquiring a steak through dumpster diving? Conventional philosophical wisdom says that they are more like eating at Jimmy's; however, the empirical facts say that they are more like dumpster diving, because it is virtually impossible for an individual's consumption of animal products at supermarkets and restaurants to have any effect on the number of animals that suffer and the extent of that suffering, just as it is virtually impossible for an individual's consumption of products acquired through dumpster diving to have any effect on animal welfare. The upshot is that consuming animal products from factory farms cannot be wrong for welfare-based reasons, because no important difference in welfare effects follows from

---

most mistakes could be eliminated by slowing the processing line speed at slaughterhouses to a reasonable level – which would also save countless workers from disabling injuries each year. For a moving discussion of this last issue, see "The Most Dangerous Job" in Eric Schlosser, *Fast Food Nation*.

[94] I distinguish *traditional consequentialism* from other views that could technically be expressed using the conceptual structure of consequentialism, but that would thereby be forced to endorse a theory of the good that is radically disunified, or at least of a kind never encountered in consequentialist theories until the late twentieth century, such as theories of the good that imply that there is something especially bad about performing particular types of acts oneself, etc.

an individual consuming factory-farmed products from a store versus a dumpster, and there is nothing wrong with consuming factory-farmed products from a dumpster, at least from a traditional consequentialist perspective.[95]

The key empirical point is that, as described in the previous chapter, animal products from factory farms, like many other products we consume, are delivered by a massive and complex supply chain extended over significant amounts of time in which there is significant waste and inefficiency at each link, where that waste and inefficiency serve as a buffer to absorb any would-be effects from the links before. As a result, it is doubtful that even a lifetime of decisions to consume animal products by an individual would make a difference to the number of animals produced and harmed over that time, and so there are no significant welfare-based reasons for an individual not to consume those products.

A common response to these arguments is to insist that they do not raise any interesting issue beyond the familiar voting paradox, which asserts that individuals are not required to vote in elections because there is virtually no chance that an individual's vote will matter. This response is misguided for several reasons. For one thing, no one has ever found a plausible utilitarian solution to the voting paradox, and so even on the assumption that the underlying philosophical issue here is exactly the same as in the voting paradox, that does not undermine the claim that familiar utilitarian arguments about eating meat are unsound.

[95] Peter Singer and Jim Mason endorse the permissibility of eating meat acquired through dumpster diving in *The Ethics of What We Eat*, pp. 260-269 (US paperback edition), on the grounds that such a strategy is "impeccably consequentialist" (pg. 268).

More importantly, any plausible argument that individuals are required to vote seems to depend on one or more of the following:

- the small possibility that an individual's vote will trigger a *dramatic threshold effect*,[96]
- the fact that voters collectively *cause* the outcome of the election in an ethically important way,[97]
- the fact that voters have a *strong personal preference* to vote,[98]
- the fact that voters have *non-welfare-based* reasons to vote.[99]

On reflection, it should be clear that theorists cannot easily appeal to any analogous claims to explain why it is wrong to eat meat from factory farms. That is because theorists cannot appeal to dramatic threshold effects, because, as noted above, waste and other forms of inefficiency ensure that significant-enough threshold effects cannot arise from an individual's consumption decisions. Furthermore, theorists also cannot claim that

[96] For example, see Brian Barry "Comment" in Stanley Bern (ed.) *Political Participation*, pg. 39, and Derek Parfit, *Reasons and Persons*, pp. 72-74.

[97] For example, see Alvin Goldman, "Why Citizens Should Vote: A Causal Responsibility Approach", and Dan Moller, "The Morality of Overdetermination: Wrongdoing When it Makes No Difference", in in *Abortion, Killing, and Overdetermination – Three Essays*, PhD dissertation, Princeton University.

[98] For example, see Dennis Mueller, *Public Choice III*, pg. 306.

[99] For a summary and critical examination of such ideas, see Jason Brennan, *The Ethics of Voting*.

consumers of factory-farmed meat *cause* animal suffering in the ethically important way that voters cause the outcome of an election, because there is no similar causal connection between consumption of animal products and the horrible mistreatment of animals on factory farms. To see why, note that in Australia, New Zealand, and many other large nations consumers have essentially the same animal consumption behavior as in the United States, but such behavior does not cause animals to be mistreated on factory farms rather than treated humanely, because factory farms do not exist in those nations in the pervasive way that they exist in the United States. The explanation is that the horrible mistreatment of animals on factory farms does not have its proximate cause on the 'demand side' in consumer behavior, but instead on the 'supply side' in the decisions of producers, and also in perverse incentives created by irrational government policies. As a result, it is false to claim that animal consumption *causes* animals to be mistreated rather than treated humanely in anything like the way that voting *causes* one candidate to win rather than another. (Of course, there is a causal connection between individuals consuming meat and *meat being produced* – but that is perfectly consistent with the current point, which is that there is no causal connection analogous to a voting case between individuals consuming meat and *meat being produced in a way that involves horrible mistreatment of animals rather than humane treatment of animals*.[100]) Theorists also cannot appeal to personal preferences, because most individuals do not have a personal preference not to eat meat. Finally, theorists also cannot easily appeal to non-welfare-based-reasons not to consume animals, because, as noted above, an individual's

[100] For general discussion of the kind of claims about contrastive causation made here, see Jonathan Schaffer, "Contrastive Causation".

consumption of animals does not *cause harm* or make a difference to the amount of harm that animals suffer, and also does not *benefit those who cause such harm*, because although an individual's consumption of animal products does have a tangible effect on the revenues of *supermarkets and restaurants*, it does not made a difference to the revenues of *factory farms* for reasons similar to the reasons it does not make a difference to the number of animals produced on such farms – and of course utilitarians cannot appeal to non-welfare-based reasons even if there are such reasons, because it is an essential part of the utilitarian view that only welfare-based reasons are relevant to determining what is right and wrong. Taken together, all of this shows that even if it is wrong not to vote in an election, that does not provide any clear reason for thinking that it is wrong to consume animal products from factory farms. Instead, explaining why it is wrong not to vote in an election is an easy problem compared to the challenge of explaining how it could be wrong to consume animal products from factory farms.[101]

Another common response is to insist that, contrary to my arguments above, there is a decisive argument that shows that the possibility of dramatic threshold effects makes it the case that the expected marginal effect on animal welfare of one individual becoming a vegetarian is substantially positive on balance.[102] For example, here is Alastair Norcross's version of such an argument, which follows an argument by Peter Singer, which is also endorsed by Shelly Kagan:

[101] To say it is an easy problem compared to another problem is not to suggest that anyone has ever offered a plausible argument that citizens are required to vote in elections, especially from a utilitarian perspective.

[102] Thanks to Richard Yetter Chappell, Ryan Jenkins, Alastair Norcross, and Peter Singer for encouraging me to consider this kind of objection.

Suppose that there are 250 million chicken eaters in the US, and that each one consumes, on average, 25 chickens per year… Clearly, if only one of those chicken eaters gave up eating chicken, the industry would not respond. Equally clearly, if they all gave up eating chicken, billions of chickens (approximately 6.25 billion per year) would not be bred, tortured, and killed. But there must also be some number of consumers, far short of 250 million, whose renunciation of chicken would cause the industry to reduce the number of chickens bred in factory farms. The industry may not be able to respond to each individual's behavior, but it must respond to the behavior of fairly large numbers. Suppose that the industry is sensitive to a reduction in demand for chicken equivalent to 10,000 people becoming vegetarians. (This seems like a reasonable guess, but I have no idea what the actual numbers are, nor is it important.) For each group of 10,000 who give up chicken, a quarter of a million fewer chickens are bred per year. It appears, then, that if you give up eating chicken, you have only a one in ten thousand chance of making any difference to the lives of chickens, unless it is certain that fewer than 10,000 people will ever give up eating chicken, in which case you have no chance. Isn't a one in ten thousand chance small enough to render your continued consumption of

chicken blameless? Not at all. While the chance that your behavior is harmful may be small, the harm that is risked is enormous.[103]

[103] "Puppies, Pigs, and People", pp. 232-233. Compare Peter Singer's argument in "Utilitarianism and Vegetarianism": "Perhaps for every 10,000 vegetarians there is one fewer 20,000 bird chicken unit than there would otherwise be. Perhaps not: this is merely an example and I have no idea what the true figure would be; but there must be some point at which the number of vegetarians makes a difference to the size of the poultry industry. There must be a series of thresholds, hidden by the market system of distribution, which determine how many factory farms will be in existence. In this case one more person becoming a vegetarian will make no difference at all, unless that individual, added to the others who are already vegetarians, reduces demand below the threshold level at which a new factory farm would have started up (or an existing one would have remained in production, if the industry is declining). Looking at one's own decision to be a vegetarian, it may seem frustrating that one cannot be sure that one has saved even a single animal from a miserable life on a factory farm; but from a utilitarian perspective it really makes no difference whether each vegetarian is personally responsible for saving ten chickens a year from this fate, or one vegetarian in 10,000 makes the difference that will save 100,000 birds. Utilitarianism judges actions by their likely consequences, and so it ranks the certainty of saving ten chickens equally with the 1 in 10,000 chance of saving 100,000. As long as I have no idea whether or not my own decision to go vegetarian is the decision that takes the demand for chickens below the threshold, the strength of this reason for being a vegetarian is unaffected" (pp. 335-336). Compare also Shelly Kagan's argument in "Do I Make a Difference?": "…we know that there is some triggering number, T (more or less), such that every Tth purchase (more or less) triggers the order of another T chickens (more or less). I don't have any idea what that number is, but I do know that whatever it is, I have a 1 in T chance (more or less) of triggering the suffering of another T chickens (more or less). And so in terms of chicken suffering, my act of purchasing a chicken still has an expected disutility equivalent to one chicken's suffering. And since, by hypothesis, this is greater than the pleasure I will get from eating the chicken, the net expected utility of my purchase remains negative. As I walk to the butcher counter, then, not only don't I know whether my act will have bad results, I don't even know what the chances are that my act is a triggering act. But I do know, for all that, that the net expected results of my act are bad. So I should not buy a chicken" (pg. 124). Compare also Allan Gibbard in *Utilitarianism and Coordination*: "I do not accept that in cases of diffuse benefits, act-utilitarianism prescribes non-cooperation… [For example,] the net value of what n gas cheats accomplish is the sum of the values of n effects individual gas cheats could have. It is the sum of the net benefits from one gas cheat in a world with no other, the net benefit from one gas cheat in a world with two others, and so on up to a world with n-1 others. If the effect of n gas cheats is calamitous, at least one of these net benefits from an individual gas cheat must be negative. Hence it is possible for an individual to produce a bad result by helping to strain the gas system, no matter how uncertain and diffuse that result may be. If the system is likely to be under strain even with everyone cooperating, an act-utilitarian will cooperate. He will calculate the average expectable net benefit from an act of gas-cheating by dividing the likely effect of a large number of gas cheats by n" (pp. 26-27).

One thing to note about this passage is that it begins by explicitly acknowledging that the expected marginal effect of one individual giving up chicken is essentially zero, before going on to argue for the opposite conclusion: in particular, it begins by acknowledging that "Clearly, if only one of those chicken eaters gave up eating chicken, the industry would not respond", which is just a way of saying: "Clearly, the marginal effect of one individual giving up chicken is zero". For our purposes, we can simply ignore this slip and examine the subsequent argument.

The main problem with the Singer/Norcross/Kagan argument is that even insofar as threshold effects are possible at the margin, those effects are not sufficiently likely and are not of sufficient magnitude to drive the expected marginal effect anywhere close to the average effect, because buffers of waste and inefficiency ensure that the real expected marginal effect of an individual's animal consumption is essentially zero.

To illustrate how this undermines the argument, consider an analogous case: Richard makes paper T-shirts in his basement that say 'HOORAY FOR ENVIRONMENTALISM!', which he then sells online. Because Richard is very busy, he only produces the paper T-shirts at the beginning of each month, and because the T-shirts are made out of very thin paper, the entire stock disintegrates after one month, and so there is no inventory carry-over from month to month (to make the case simple). The T-shirts are incredibly cheap to produce and very profitable to sell and Richard doesn't care about waste per se (he is not an environmentalist himself – he is just an online entrepreneur), and so he produces far more T-shirts than he is likely to need each month, and then sells the excess at a nearly break-even amount at the end of each month to his

hippie neighbor, who burns them in his wood-burning stove. For many years Richard has always sold between 14,000 and 16,000 T-shirts each month, and he's always printed 25,000 T-shirts at the beginning of each month. Nonetheless, there is a *conceivable* increase in sales that would cause him to produce more T-shirts – in particular, if he sells over 20,000 this month, he'll produce 30,000 T-shirts at the beginning of next month; otherwise he'll produce 25,000 like he always does. So, the system is genuinely sensitive to a precise tipping point – in particular, the difference between 20,000 purchases and the 'magic number' of 20,001.

Suppose that a consumer knows all of these facts about Richard's business, and is considering buying a T-shirt for himself. What is the expected effect on the number of T-shirts produced of that consumer purchasing a T-shirt? The correct answer is essentially zero, because given what is known about the history of demand for Richard's T-shirts and how production quantities are determined, there is virtually no chance that exactly 20,001 people are going to buy Richard's T-shirts this month and trigger a dramatic threshold effect exactly on the margin – which, of course, is not to claim that there is *zero* chance of that happening, but rather that the odds of that happening – of exactly 20,001 of Richard's T-shirts being sold – is certainly dramatically lower than 1/5,000 or any other number that would drive the expected marginal effect of an individual buying one T-shirt anywhere near the consequence that 1 additional T-shirt is produced. This shows that the sort of general principles behind arguments like the Singer/Norcross/Kagan argument have to be mistaken, because insofar as they imply that consuming meat should be expected to have significant bad effects for animal welfare, they also imply that

buying one T-shirt in the story above should be expected to result in approximately 1 additional T-shirt being produced, which is the wrong result.

In response, it might be insisted that there is some crucial disanalogy between the T-shirt case just described and our actual situation with respect to animal products. One obvious difference might seem to be the amount of waste: in particular, in the T-shirt case, almost half the product is 'wasted'. However, upon further reflection there is no crucial disanalogy between the cases with respect to waste. For example, consider the meat that goes out of date in a wholesaler's meat locker or on a supermarket shelf, and is then sold to a dog food plant or 'rendered' into feed for other animals. Is that meat 'wasted'? What is the difference between that meat and the 'wasted' T-shirts in the story above that Richard sells to his neighbor? For current purposes it doesn't really matter whether such things are labeled as 'wasted' or not – what matters is that there are large amounts of 'wasted'-like meat at each stage of the supply chain that serve as buffers to prevent an individual's decision to purchase meat from making any difference to the number of animals that are produced at the far other end of the supply chain, just as there are large amounts of 'wasted'-like T-shirts in the story above that serve as a buffer to prevent a single individual's purchase of a T-shirt from having any real effect on the number of T-shirts produced. In the case of animal products, the relevant buffer is even more impenetrable, because even if an individual's decision somehow made a difference between the bottom links in the supply chain, that effect would almost certainly be absorbed by the buffer at next link up, and so on up the supply chain.

Furthermore, even if individual purchases really did succeed in making the price of, say, cattle at a feedlot $0.01 /lb. higher than it otherwise would have been, that would not make the dramatic difference to the number of cows that are brought into existence that it would have to make in order for the possibility of such a threshold effect to drive the expected marginal effect toward the average effect, in part because the number of cows that are brought into existence is surprisingly insensitive to very small changes in the wholesale price of beef for the reasons outlined above. As a result, a very small change in the wholesale price wouldn't have anywhere near the effect it would need to have on the number of animals produced in order to give rise to a threshold effect that would justify equating the expected marginal effect with anything like the average effect of animal consumption decisions.

As further confirmation of all of this, consider how the Singer/Norcross/Kagan reasoning would apply to the everyday example of power consumption:

> Suppose that there are 250 thousand power consumers in your region, and that each one consumes, on average, 25 units of power per year. Clearly, if only one of those power consumers stopped consuming power, the industry would not respond. Equally clearly, if they all gave up consuming power, vast amounts of power (approximately 6.25 million units per year) would not be produced. But there must also be some number of consumers, far short of 250 thousand, whose renunciation of power would cause the industry to reduce the amount of power produced. The industry

> may not be able to respond to each individual's behavior, but it must respond to the behavior of fairly large numbers. Suppose that the industry is sensitive to a reduction in demand for power equivalent to 10,000 people giving up power consumption. For each group of 10,000 who give up power, a quarter of a million fewer units of power are produced per year. It appears, then, that if you give up consuming power, you have only a one in ten thousand chance of making any difference to the amount of power produced, unless it is certain that fewer than 10,000 people will ever give up consuming power, in which case you have no chance. Isn't a one in ten thousand chance small enough to render your continued consumption of power inefficacious? Not at all. While the chance that your behavior will have an effect may be small, the effect that is risked is enormous.

As in the original argument, the only accurate claim here is the initial observation that the expected effect on production of an individual's consumption is essentially zero. The subsequent argument for the opposite conclusion has to be mistaken, because in the case of power production, as in the case of animal production, we know enough about the supply chain to know that threshold effects are not sufficiently likely at the margin and are not of sufficient magnitude to drive the expected marginal effect of consumption anywhere close to the average effect. For example, although it is easy to imagine that an individual is saving power by turning off the lights in her house, upon reflection it should

be clear that the expected effect of such an act on the quantity of power *produced* is essentially zero, because the marginal effect on a large power grid of a single individual turning off power in her home is merely to intangibly reduce the voltage in the neighborhood of that individual in a way that is not measurable from the point of view of power producers, and to increase the reading on her meter in a way that amounts only to a miniscule change in insignificant digits when meter readings are aggregated. As a result, the probability we should assign to an individual having any effect on the quantity of power produced by power plants is vanishingly small in a way that ensures that the expected effect on the quantity produced is closer to zero than it is to the average effect on the quantity produced of all similar actual acts of consumption.

As these examples illustrate, general arguments like the following have to be mistaken:

Invalid Argument

Premise 1: You know that if some number n of additional people were to perform action A, effects E would ensue.

Premise 2: You don't know exactly how many additional people will perform A.

Conclusion: Therefore, the expected marginal E-effect of you A-ing is (1/n)*E, if you know that there are no other thresholds for E-effects.

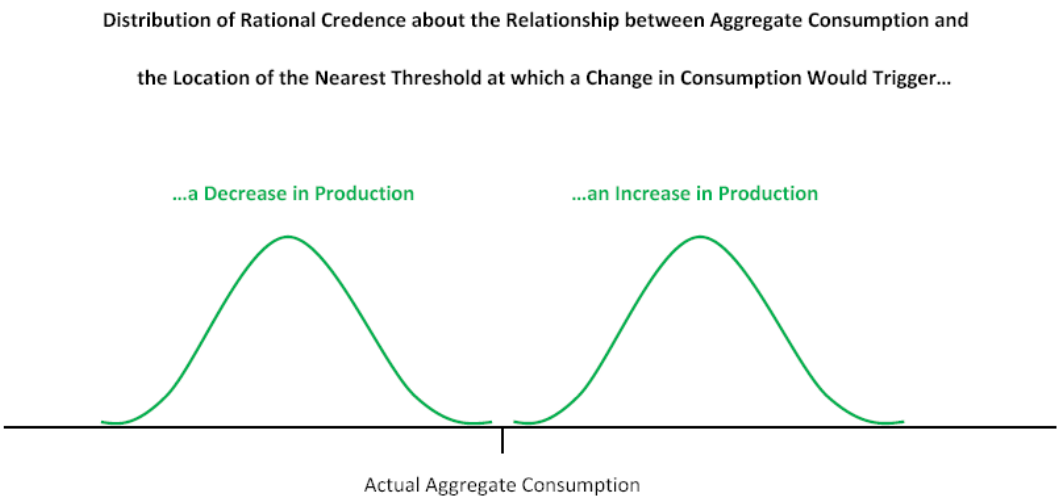
The argument above is invalid and instances of its conclusion are typically false because we usually have additional knowledge about the mechanisms at play within a collective action situation that rationally requires assigning a very different probability to a threshold landing at the margin than the simple idea that the probability is 1/n. As a result, if arguments such as the one above are to be made plausible, occurrences of '1/n' must be replaced with 'the probability of the relevant threshold landing at the margin'. However, once that substitution is made, the result does not support anything like the conclusion that the expected marginal effect on animal welfare of one individual becoming a vegetarian is substantially positive on balance – instead, it supports the initial thought that the expected effect is closer to zero than to the average effect, because given what is known about waste and inefficiency, it would be unreasonable to assign a probability to the nearest thresholds landing at the margin that was sufficiently high to vindicate the Singer/Norcross/Kagan argument. Instead, the actual probability of such thresholds landing at the margin in such cases is often vanishingly small given what we know – certainly dramatically less than 1/n for any value of n equal to or lesser than the total number of consumers.

In general, when evaluating consequentialist reasons for individual action in collective action situations, our knowledge about the mechanisms at play in such situations matters greatly, and it is a mistake to think that there is a simple argument that that shows that individuals must always 'cooperate' for consequentialist reasons in all or most such situations, even when the stakes are high, and even when consequentialists themselves

agree that 'cooperation' is required.[104] The crucial probabilistic point is that the probability of triggering effects that would not otherwise result is a function not only of the number of other actors whose collective action is necessary for such effects, but more importantly also depends on the expected difference between the margin and tipping points for those effects: and while the probability of triggering such effects is somewhat sensitive to the number of other relevant actors, it is in general much more sensitive to the expected difference between the margin and tipping points.[105] The probabilistic

[104] Here and elsewhere I am focusing on collective action situations in which there is a natural distinction between a 'cooperative' collectively-desirable option and a 'non-cooperative' collectively-undesirable option. Simple arguments that individuals are required to cooperate in such situations are often driven by what might be called the *Average Effects Fallacy*, which is the fallacy of equating the ethically relevant effects of an act with the average effects of actual acts of that type. Although such an equation is a mistake in general, it is plausible to maintain that average effects sometimes (but not always) track the ethical truth for more indirect reasons. For example, the analysis I ultimately defend below is consistent with this idea. I discuss the Average Effects Fallacy in more detail in the previous chapter.

[105] More precisely, the epistemic probability of making a difference is highly sensitive to the distribution of rational credence regarding the likely difference between the margin and nearest tipping points. In many such cases, a simplistic graph such as the following suffices to illustrate why the probability of making a difference is dramatically lower than the Singer/Norcross/Kagan argument assumes:



Of course, such a simplistic graph is inappropriate with respect to some situations, and there are also some cases in which the Singer/Norcross/Kagan reasoning delivers the correct verdict – because it is, in general, uncontroversial that there are *some* cases in which traditional consequentialism delivers the correct verdict.

mistake behind the Singer/Norcross/Kagan argument is to ignore completely the impact of such expected differences, and to assume instead that we should reason about the probability of making a difference as if we had no information at all about the difference between the margin and tipping points – which is almost always the opposite of the truth, as reflection on real-world examples such as power production and animal production illustrates.[106]

The upshot is that it is impossible for traditional consequentialist views to deliver the plausible verdicts that they themselves endorse in many real-world collective action situations, because by the reasoning above, in large real-world collective action situations individuals often have a probability of making a difference that is sufficiently small to ensure that 'non-cooperation' is the option with the greatest expected value, even when the stakes are high, and even when consequentialists themselves agree that 'cooperation' is required. When applied to the collective action problem of deciding whether to eat meat, in which the probability of an individual's consumption decisions causing anything significant to happen at the other end of the supply chain is vanishingly small, this means that true principles do not support anything like the conclusion that the expected marginal effect on animal welfare of one individual becoming a vegetarian is

---

The important point is that the Singer/Norcross/Kagan reasoning fails dramatically with respect to very many real-world cases in which even consequentialists agree that 'cooperation' is required.

[106] In other words: it is plausible that if we didn't know anything about a supply chain but yet did somehow know the average effect of a unit of consumption, then some 'principle of indifference' would tell us to equate the marginal effect and the average effect; however, in the real world we almost generally have additional evidence that makes it empirically indefensible to equate those things, and that makes it indefensible to assign a probability to making a difference that would be sufficiently high to vindicate the conclusions of the Singer/Norcross/Kagan argument.

substantially positive on balance, but instead support the opposite conclusion in a decisive way.[107]

In response to all of this, a theorist might abandon the idea that an action is to be judged by its *marginal effects*, and insist instead that an action is to be judged by the *average effects* of actions of that type, or in some more complicated way that falls under the rubric of 'rule consequentialism', 'cooperative consequentialism', or 'universalizability theory'. The main problem with such theories is that either they lack clear content, or else, insofar as they have clear content and are inconsistent with the practical conclusions of this paper, they are subject to clear counterexamples. For example, consider the following:

Stampede Case

> We find ourselves in an enormous stampede. Unless everyone stops stampeding, it is clear that an increasing number of innocent people will be seriously harmed and killed. However, it is also clear that there is

[107] As a result, I disagree with Shelly Kagan's analysis of collective action problems in "Do I Make a Difference?": After giving the same argument regarding chicken consumption as Norcross (the text of which appears several footnotes above), Kagan writes "I have discussed the example of purchasing a chicken at considerable length, because I take it to be a fairly representative case of the situation we often find ourselves in with regard to collective action problems. … But if my discussion of this sort of case is correct, then the consequentialist can handle such cases using the familiar appeal to expected utility. Admittedly, in such cases, I may not be able to know whether or not, if I act, I will be part of a cohort of the relevant size for triggering the bad results. But no matter. I can still know that the expected utility of my act is negative. And that will be enough to allow the consequentialist to condemn my act" (pg. 129). The problem with Kagan's discussion is that it mistakenly assumes that 1/n is the probability of an act being decisive in real-world collective action problems with n relevant actors and a single threshold effect, and that analogous probabilities should be used in cases with multiple threshold effects because we should reason as if we had no information about the expected difference between the margin and nearest tipping points. In particular, see Kagan's discussion on pg. 124 and the formula displayed on pg. 120.

virtually no chance that everyone will stop stampeding soon, and so anyone who stops stampeding will almost certainly be seriously harmed or killed in a way that does no good for anyone else.

Charity Case

You live in a different world in which everyone is required to make all of their charitable contributions on the last day of each year. There are only two charities to which contributions can be made, and the only options are to give $1,000 to one of those charities, or to give no money to any charity at all. Furthermore, there is a website where you can track everyone's decisions throughout the last day of the year – the website is highly accurate, updated every 5 seconds, etc. This system has been in place for several decades, and the trend is for almost everyone who is going to contribute to a charity to make their decision early in the day, and in fact very few people (less than one million) ever decide to give to charity after 10pm. Usually, about 100 million people decide not to give to any charity at all, which is the 'default' option. Knowing all of this, at 11:45pm you are deciding how to make your personal charitable contribution, and by looking at the website you see that 111 million people have given to charity A, 11 million have given to charity B, and 100 million people have not made any decision. It is well known that the good done by charity A is not a linear function of the number of contributors, because there are

> dramatic thresholds involved; in particular, for each increment of exactly 100 million contributors, charity A saves exactly 200 million people from death and ensures that they live happily ever after, but charity A does not help anyone in any other way – so, because of this threshold effect, exactly 200 million people will be saved by charity A regardless of whether 111 million people, 121 million people, or 199 people give to that charity. In contrast, for each and every contributor to charity B, one person is saved from death and lives happily ever after.

These cases are analogous to an interesting range of real-world collective action problems in which ethical actors are 'thrown into' a non-ideal situation in which a regrettable outcome looms in the background through no fault of their own in a way that makes non-universalizable action required in their circumstances. In particular, in the Stampede Case, the thing to do is to continue stampeding, despite the fact that a version of the 'categorical imperative' says to stop stampeding because everyone [in the same relevant circumstances] [acting on a maxim that implies] continuing stampeding would be dramatically worse than everyone [in the same relevant circumstances] [acting on a maxim that implies] stopping stampeding. In the Charity Case, the thing to do is to give to charity B, despite the fact that the average effect of giving to charity B is significantly worse than the average effect of giving to charity A [in the relevant circumstances], and despite the fact that the 'categorical imperative' says to give to charity A because everyone [in the same relevant circumstances] [always acting on a maxim that implies]

giving to charity A would be significantly better than everyone [in the same relevant circumstances] [always acting on a maxim that implies] performing any of the relevant alternative acts. As these cases and countless others illustrate, the alternative views mentioned above seem doomed to deliver mistaken verdicts, at least insofar as they have clear content and are inconsistent with the practical conclusions of this paper.[108]

When defenders of such views attempt to respond to such counterexamples, the typical strategy is to provide a more fine-grained specification of what is meant by the 'same relevant circumstances' or the 'same type of act';[109] however, although such views are then able to avoid an increasing number of counterexamples as such a specification becomes increasingly fine-grained, the only way to avoid all such counterexamples is by following such a strategy to the limit and adopting a specification that is maximally fine-grained; but, taken to that limit, such a view is then equivalent to the view that acts are to be judged by their marginal effects. As a result, there is no plausible way of developing such views while rejecting the idea that an action is to be judged by its marginal effects.

[108] The brackets are intended to draw attention to the various alternative formulations of 'average-effect consequentialism', 'rule consequentialism', 'cooperative consequentialism', and 'universalizability theory', and to show how each formulation is subject to counterexample. Although the details of Kantian ethics are notoriously unclear, reflection on these cases also clarifies that Kantian theories cannot be used to undermine the main arguments of this paper, because insofar as they avoid mistaken verdicts on the cases above, such theories seem perfectly consistent with the main arguments of this paper.

[109] On such views, there is usually no clear explanation of what it means for acts to be in 'relevantly similar circumstances' or of a 'relevantly similar type'. For discussion of these and other problems with such views, see Allan Gibbard, *Utilitarianism and Coordination*, Chapter 1. The problem is not merely that a 'generality problem' arises, but more importantly that it is impossible for there to be a solution to that problem that escapes decisive counterexamples. I provide further arguments against such theories in the preceding and following chapters.

Furthermore, there is a clear argument that non-marginalist consequentialism is not even conceptually possible, given a standard conception of consequentialism. That is because on a standard understanding of consequentialism it is a conceptual truth that consequentialism never requires acts that (are known to) lead to worse consequences than alternatives – but non-marginalist consequentialism would imply that there are such cases. For example, consider a ranking of outcomes via some set of values V, and consider a case in which V-consequentialism requires some action A on a marginalist interpretation. In such a case, by the definition of 'marginal', action A would lead to better V consequences than any of the alternatives – otherwise, A would not be required by the marginalist interpretation of V-consequentialism. But that means that any view that requires some action other than A sometimes requires acts that lead to worse V consequences than the alternatives. Clearly then, such a view cannot be a version of V-consequentialism, because such a view sometimes requires acts that (are known to) have worse V consequences than other alternative acts.

The upshot is that welfare-based views are unable to explain why it could be objectionable to eat meat from factory farms, and no attractive alternative explanation is offered by views that fall under the rubric of 'rule consequentialism', 'cooperative consequentialism', or 'universalizability theory'. (I argue against such theories at greater length in the following chapter.)

A different strategy is to argue that eating meat from factory farms is objectionable for other reasons. As noted in the previous chapter, such an explanation cannot be as straightforward as it initially appears, because an individual's decision to consume meat

does not actually result in any additional animals being harmed or killed, and so does not obviously violate deontological constraints or any other ethical principles; furthermore, initial appearances aside, your purchases at supermarkets and restaurants do not really have any tangible effect on the revenues of factory farms, and therefore do not *support* factory farms in any clear sense, because although an individual's consumption of animal products does have a tangible effect on the revenues of *supermarkets* and *restaurants*, it does not made a difference to the revenues of *factory farms* for reasons similar to the reasons it does not make a difference to the number of animals produced on such farms. Finally, straightforward deontological explanations and explanations in terms of complicity in evil tend to overgeneralize and imply that you are almost never permitted to consume anything at all, because petroleum companies routinely violate significant constraints,[110] and almost every possible consumption activity depends on and supports such companies to a much greater extent than buying animals at supermarkets and restaurants depends on and supports factory farms, because we often purchase gasoline directly from the petroleum companies themselves, or are at least only one step in the supply chain removed from such companies, and in the petroleum supply chain there is much less of a buffer caused by waste and inefficiency than in the supply chain for animal products.

In light of all this, as well as the considerations discussed in the previous chapter, the most plausible account of what we are required to do when products are produced in a morally objectionable way depends on the distinction developed in the previous chapter

[110] For examples of routine abuses, see Peter Maass, *Crude World*, especially chapters 2, 3, and 4.

between activities that are *essential* to a product or to the actual production of that product, and activities that are not.

Why Morality and All Other Forms of Normativity are Sometimes Dramatically Directly Collectively Self-Defeating

In a prisoner's dilemma, if everyone follows the strategy of self-interest, then everyone is certain to be worse off from the perspective of self-interest than they would have been if everyone had not followed self-interest instead. This shows that self-interest is sometimes *directly collectively self-defeating*.

In *Reasons and Persons* and *On What Matters*, Derek Parfit argues that all plausible moral theories – including the most plausible versions of consequentialism, contractualism, and Kantian ethics – imply that morality is never directly collectively self-defeating:

> [The assumption that morality is never directly collectively self-defeating] is either made or implied by most of the many different theories [of morality].[111]

[111] *Reasons and Persons*, pg. 113.

> …moral principles or theories are intended to answer questions about what *all* of us ought to do. So such principles or theories clearly fail, and condemn themselves, when they are directly self-defeating at the collective level.[112]

Some theorists believe not only that morality can never be directly collectively self-defeating (DCSD), but also that rationality and other forms of normativity can never be DCSD. For example, the Kantian idea that our acts or principles must be willable as universal law seems to be an obscure way of suggesting that both morality and rationality can never be DCSD. And even theorists who grant that there is a narrow form of rationality that is sometimes DCSD often insist that there is a broader and more important form of rationality, sometimes called 'enlightened self-interest', that is never DCSD.

These theorists are all mistaken, because morality and all other interesting forms of normativity are sometimes directly collectively self-defeating. To see why, consider cases like the following:

Stampede Case

> We find ourselves in an enormous stampede. Unless everyone immediately stops stampeding, it is clear that some of us will be

[112] *On What Matters*, Volume One, pg. 306, italics in the original.

moderately harmed. However, it is also clear that everyone will not immediately stop stampeding, and so anyone who does stop stampeding will be severely harmed in a way that does no good for anyone else.

In this case, from every interesting normative perspective – self-interest, enlightened self-interest, morality, pure benevolence, or whatever – each person is *required* to continue stampeding, despite the fact that the outcome would be *better* in every sense if each did not continue stampeding instead. This shows that morality and all other forms of normativity are sometimes directly collectively self-defeating, because everyone can be sure that if each person does what is required, the result will be worse from the perspective of each than if each had not done what is required instead.

Here is another example:

Units of Good Case

1,000 people are put into isolation booths. It is common knowledge that each must choose between Options A and B, with the following outcomes: If everyone chooses A, then each receives 99 additional units of good; if everyone chooses B, then each receives 100 additional units of good; otherwise, each person who chooses A receives 10 additional units of good and each person who chooses B loses a catastrophic 100,000 units of good.

In this case, from every interesting normative perspective, each person is *required* to choose A, despite the fact that it is clear that the outcome would be *better* from each person's perspective if everyone did not choose A instead. Once again, this shows that morality and all other forms of normativity are sometimes directly collectively self-defeating, because everyone can be sure that if each does what is required, the result will be worse than if each had not done what is required instead. Importantly, these conclusions follow even on the assumption that it is common knowledge that everyone knows the relevant facts and will satisfy their normative requirements.[113]

Why is it an essential part of normativity that it is sometimes directly collectively self-defeating? The answer is that even when it is common knowledge that everyone will

[113] On the intended and most straightforward understanding of these cases, each knows enough about what the others will do to know which choice will be objectively best, and thus each knows which choice is objectively required, and thus each knows of a particular option that s/he will successfully follow morality only if s/he chooses that option. If it is not immediately clear how agents could have such knowledge in these cases, it may help to consider the following: Imagine that you are one of 1,000 subjects in a case that is identical to the Units of Good Case except that in this case you have attached a device to the heads of each of the other subjects, where this device will ensure that each selects option A as the result of a pre-programmed chain of reasoning that is forced upon each of their minds. As a result, as you sit in your isolation booth, you know that the devices will ensure that most of the others will choose A, and so you know that it would be objectively best for you to choose A. Initial appearances aside, there is no difference relevant to knowledge between this case and the actual Units of the Good Case. That's because in the actual Units of Good Case subjects know that the others understand the situation and will act morally; as a result, each knows that most of the others will choose A as a result of a predictable chain of reasoning – in fact, we can assume it is the very same chain of reasoning that the devices force on the others in the modified case just described. So, the assumption that a subject knows that the others understand the situation and will act morally has the same result regarding knowledge as the assumption that a subject knows that there is a device attached to the others' heads that will force a particular choice: just as knowing that a device is attached to each allows one to know that most will choose A, so too knowing that each understands the situation and will act morally allows one to know that most will choose A. The result is common knowledge in the Units of Good Case that most of the others will choose A, which means knowledge that choosing A will be objectively best, which means knowledge that one successfully follows morality only if one chooses A. Similar remarks apply to the other cases discussed in this paper.

satisfy their normative requirements and that everyone knows the relevant facts, it is sometimes also clear that the option that would lead to the best outcome if universally chosen is associated in a way that is salient to everyone with great risks without compensating rewards, and in some such cases each person can, by this very reasoning, know that others will coordinate on a risk-averse option instead, thereby ensuring that each person is required to choose that risk-averse option themselves, even if it is clear that everyone choosing that risk-averse option guarantees a worse outcome from the perspective of each than if everyone did not choose that option instead.

In addition to showing that all forms of normativity are sometimes DCSD, this also shows that an important research program on morality and coordination problems is misguided, because the essential and guiding assumption of that research program is that following morality always guarantees optimal cooperation in coordination problems where it is common knowledge that: everyone has full information about the choices facing everyone, will act freely, will satisfy their normative requirements, and knows of a unique option that the outcome would be best if that option were chosen by everyone.[114]

In light of these results, we can draw some more general conclusions about normative theories.

First, consider the Kantian idea that an act is permissible only if the maxim behind that act is willable as universal law. What does this mean? Suppose that we do not know what

[114] For an brief description of this research program, see Derek Parfit, "Comments", *Ethics* 1986, pg. 867; for more detail, see Donald Regan, *Utilitarianism and Cooperation*, especially pp. ix –xi, and pp. 4-5, and Allan Gibbard, *Utilitarianism and Coordination*, especially pp. 6-9.

this means. Nonetheless, if we know that it implies that morality is never DCSD, then we can know that it is false, even if we do not know what it means.

More generally, consider versions of 'Kantian ethics', 'rule utilitarianism', 'utilitarian generalization', 'cooperative utilitarianism', or any other view on which a notion of 'universalizability' seems to play a guiding role, and suppose that defenders of such views, like most honest theorists, are unable to articulate the precise details of their views. As a result, purported counterexamples are always seen as merely "cases to be dealt with by a more complete analysis". How could we ever make progress with such views? Because we can show that morality is sometimes DCSD, we can show that such views are false if they imply that morality is never DCSD, even if the details of the views remain unclear. This might lead theorists to abandon such views, to rethink their motivations, or at least to state their commitments and non-commitments more clearly.

Recognizing that morality is sometimes directly collectively self-defeating should also lead us to reexamine our beliefs about what individuals are required to do in real-world collective action problems. For example, consider the following case:

Pollution Case

> Each of us will do better by not reducing emissions than by reducing emissions; however, at the same time, each of us will do substantially worse if no one reduces emissions than we would if everyone reduced emissions.

Many would say that each of us is required to reduce emissions in this case because the alternative is directly collectively self-defeating. However, that is a bad argument, because morality and all other forms of normativity are sometimes DCSD. So, if individuals are required to reduce emissions in such a case, it must be for some other reason.[115]

In response, it might be claimed that although morality is sometimes *mildly* DCSD as in the Stampede Case and the Units of Good Case above, it is never *dramatically* DCSD as it would have to be if, for example, not reducing emissions was permissible in our actual situation with respect to climate change.

At first glance, this response might seem promising. However, it does not succeed, because morality and all other forms of normativity are sometimes *dramatically* directly collectively self-defeating. To see why, consider cases like the following:

Dramatic Stampede Case

> We find ourselves in an enormous stampede. Unless everyone stops stampeding, it is clear that an increasing number of people will be seriously harmed and killed. However, it is also clear that everyone will not stop stampeding, and so anyone who does stop stampeding will be

[115] Such bad reasoning lies behind the question, "But what if everyone did that?".

> severely harmed or killed in a way that does no good for anyone else and simply adds to the ultimate aggregate harm caused by the stampede.

This case is representative of real-life stampedes. In such cases, individuals are not required to stop stampeding, even if continuing is dramatically directly collectively self-defeating.

Here is an infinitely dramatic example:

> **One Million Dollars Case**
>
> Everyone on the planet is isolated and instructed to choose a number, and a neon sign reading 'One Million' is lowered in front of each person. It is common knowledge that a qualitatively identical sign is lowered in front of each person, and it is common knowledge that if everyone chooses the same number, then the standard of living of each person in the world will be increased by an amount equivalent to one one-millionth of that number of dollars; otherwise, if everyone fails to choose the same number, each person's standard of living will be substantially reduced.

What number should each choose in this case? Each should choose one million, because it is common knowledge that one million is uniquely salient to everyone, which makes it

common knowledge that one million is the only number that has any chance of being chosen by everyone, which makes it the case that each should choose that number. However, if everyone chooses one million, the standard of living of each person in the world will remain the same rather than rising by, say, $1 billion each, which it is clear that everyone could bring about by simply by choosing the number one quadrillion instead of one million (and so on for any amount whatsoever). As this shows, morality and all other forms of normativity are sometimes *catastrophically directly collectively suboptimal*, because they sometimes direct everyone to choose an option that is certain to lead to a catastrophically worse outcome than an antecedently identifiable option that they could have directed everyone to choose instead. This is truly catastrophic, because instead of solving all of the world's material problems, following morality and other forms of normativity in such a case would not do anyone any good at all.[116]

Here is another dramatic example:

End of the World Case

Aliens come to Earth and force each family on the planet to choose between 'cooperating' and 'defecting', which are known to have the following consequences: If all choose to cooperate, the aliens will leave and everyone's life will go on the same as before – but if even one family

[116] This example is based on a case discussed by Thomas Schelling in "Bargaining, Communication, and Limited War", who also introduces the relevant notion of *salience*. For ease of exposition, I am pretending that one dollar would not make a difference to anyone on the planet.

> chooses to defect, in one year the aliens will destroy the Earth and every living thing on the Earth, and in the meantime will ensure that each family that chooses to cooperate has a miserable life of intense suffering, while each family that chooses to defect has a wonderful and flourishing final year on Earth.

A glib philosopher might insist that every family would be required to cooperate in this case. But upon reflection, it is clear that if a billion families were actually in this situation, then some of them, by this very reasoning and without making any moral mistake, would choose to defect, thereby ensuring the end of the world in one year; as a result, cooperation would mean a futile sacrifice of one's own family in a way that was impermissible. This illustrates the way in which morality and other forms of normativity are sometimes *catastrophically* directly collectively self-defeating.

In response to all of this, some might insist that it is simply *absurd* to think that morality is sometimes dramatically directly collectively self-defeating. Such a thought is true in an important sense – it is true in the same sense that it is true that we sometimes find ourselves in situations that are *absurd*. But *absurdity* in that sense does not give rise to a reductio – just as finding ourselves in an *absurd* situation such as the End of the World Case would not show that we were not in that situation.

What does contractualism say about all of this? It is unclear. Contractualism is, roughly, the view that an act is required if it is required by principles that we would agree upon in

some special scenario, or, alternatively, if it is required by principles that we could not reasonably reject. One problem for contractualism is that everyone would want everyone to cooperate in the End of the World Case, and everyone would agree to cooperate if such agreement was possible and binding; but if it is supposed to follow from these facts that contractualism implies that individuals are required to cooperate in the End of the World Case, then the view is false, and it is false because it ignores every interesting aspect of collective action problems. A problem for Scanlonian contractualism seems to be that for any principle that requires cooperation in the End of the World Case, that principle could be reasonably rejected, and for any principle that permits defection, that principle could be reasonably rejected; so, the view seems to have no clear content when applied to such collective action problems.

In response, contractualists might insist that their view does not mistakenly imply that 'cooperation' is required in the Dramatic Stampede Case and the End of the World Case. But if that is correct, then it is unclear how their view could require us to reduce our emissions in our actual situation with respect to climate change, especially in light of cases such as the following, which seem analogous in to our actual situation in all important respects, including responsibility, costs/benefit considerations, and general non-cooperation as a foreseeable outcome:

Asteroid Case

> A giant asteroid will kill us all if we do nothing. The only way we can save our planet is by gathering in Kansas and firing our ray guns at the asteroid in unison: if we were almost all there, the collected beam would be just powerful enough to vaporize the asteroid. Unfortunately, it is clear that not nearly all of us will cooperate in this way, and so our planet is doomed. Nonetheless, the destruction of the Earth will be delayed based on the number of people who do show up – in particular, for every person who shows up in Kansas and fires their gun, the destruction of the Earth will be delayed by 1/1,000 of a second. An additional detail is that the only reason we are threatened by the asteroid is that one of our mineral-extracting expeditions to the asteroid knocked it into its threatening orbit, and we do not have time to deflect it using any other means.[117]

A glib philosopher might insist that everyone would be required to make the trip to Kansas in this case. But upon reflection, it is clear that if we were actually in this situation, then most people, by this very reasoning and without making any moral mistake, would choose to not to make the trip to Kansas, thereby ensuring a catastrophic outcome; as a result, taking the trip to Kansas would mean a futile sacrifice of one's own

[117] This case is not intended as a clear example of direct collective self-defeat. As the analogy properly assumes, a genuine solution to anthropogenic climate change would have significant costs but would not be catastrophically costly even in the short run.

interests in a way that was not required, given that almost everyone else was not willing to make such a sacrifice themselves.

As these cases collectively illustrate, contractualist theories are either false because they imply that morality is never DCSD, or else they do not provide any clear reason for thinking that we are required to reduce our emissions, and do not give us any clear guidance about how to think correctly about genuinely challenging collective action problems. Similar remarks apply to universalization theories: either those theories are false because they imply that morality is never DCSD, or else they do not provide any clear reason for thinking that we are required to reduce our emissions, and do not give us any clear guidance in how to think correctly about genuinely challenging collective action problems.[118]

The general upshot is that morality and all other forms of normativity are sometimes dramatically directly collectively self-defeating, which means that many influential normative theories are either false, or at least don't have the consequences that their

[118] The cases above are counterexamples even to sophisticated universalizability theories that are intended to apply to real-world, non-ideal situations, while at the same time remaining never directly collectively self-defeating. For example, consider Parfit's principle "Follow the rules whose being followed by everyone would make things go best, unless some other people have not followed these rules, in which case do whatever, given the acts of others, would make things go best" (*On What Matters*, Volume One, pg. 317). Taken literally, this principle seems to deliver the mistaken verdict that everyone is required to choose B in the Units of Good Case, because at the moment that everyone chooses in that case, no one has yet failed to follow optimific principles, and so the principle seems to imply that each must choose B. Similar remarks apply regarding at least most of the other cases above. An alternative formulation might be: 'Follow the rules whose being followed by everyone would make things go best, unless it is *known* that some other people won't follow these rules, in which case do whatever, given the acts of others, would make things go best'. This principle does not deliver the wrong verdict on the cases above; however, this principle is sometimes dramatically directly collectively self-defeating, as such cases illustrate.

adherents take them to have.[119] One important consequence is that morality and other forms of normativity cannot be relied upon to solve collective action problems even in a world of normatively flawless agents. In particular, even if a disaster will ensue if everyone acts in a particular way or on a particular principle, that does not settle the question of whether individuals are permitted to act in that way or on that principle. And because many of the most important questions about modern moral life are essentially questions about what individuals are required to do in such situations – for example, what individuals are required to do about climate change, what individuals are required to do when products are produced in morally objectionable ways – an important practical upshot is that such questions cannot be answered by asking 'But what if everyone did that?', or by more sophisticated appeals to 'universalizability'.[120]

Deontologists might object that this entire discussion depends on a sense of *betterness* that is foreign to their view, because (they might say) their view is concerned with *acts* rather than *outcomes*.[121] However, such an objection is misguided. If we all continue stampeding in the Stampede Case, it is certain that we will cause harm, whereas if we all stop stampeding, it is certain that we will do no harm at all. As a result, it is perfectly

[119] It is worth noting that the arguments here do not depend on any controversial premises. In particular, the arguments here do not depend on the controversial premise that the *better than* relation is intransitive; compare Stuart Rachels, "Counterexamples to the Transitivity of Better Than", and Larry Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Even if the *better than* relation were intransitive, that would not show that morality is sometimes *dramatically* DCSD, as the arguments here reveal.

[120] Here it may be useful to note that in the preceding chapters I argued that there are often on balance no consequentialist reasons to cooperate in such situations, and that there are sometimes no weighty reasons of any other kind to cooperate.

[121] For such an objection, see Robert Adams, "Should Ethics be More Impersonal?", pg. 259 in Jonathan Dancy (ed.) *Reading Parfit*.

sensible and correct to say that, collectively, continuing stampeding is deontologically worse than stopping stampeding, but that nonetheless each of us individually is required to continue stampeding, because if an individual were to stop, s/he would do something (namely, severely harm an innocent person – him or herself) that is deontologically worse than what s/he would do by continuing stampeding. That is why, in the Stampede Case, deontology is DCSD.[122]

A more subtle objection comes from agent-neutral consequentialists, some of whom believe that it is a clear and important virtue of their view that it is never directly collectively self-defeating. For example, Derek Parfit argues:

> [Agent-neutral consequentialist theories] cannot be directly self-defeating, since [they are] *agent-neutral*: giving to all agents *common* moral aims.[123]
>
> …Common-Sense Morality is often directly collectively self-defeating. [But] a moral theory must be collectively successful. [Those who believe in Common-Sense Morality] must therefore revise their beliefs, moving from [Common-Sense Morality to a form of agent-neutral consequentialism].[124]

---

[122] Another example from Parfit: "Suppose that each could either (1) carry out some of his own duties or (2) enable others to carry out more of theirs. If all rather than none give priority to their own duties, each may be able to carry out fewer. Deontologists can face [situations in which their theory is DCSD]" (*Reasons and Persons*, pg. 98).

[123] *Reasons and Persons*, pp. 54-55, italics in the original.

[124] *Reasons and Persons*, pg. 111. See also *On What Matters*, Volume One, pg. 306.

This is Parfit's main argument for the sort of moral theory he favors in *Reasons and Persons*. Unfortunately, this argument is unsound, because agent-neutral consequentialism is sometimes directly collectively self-defeating. For the clearest possible example, consider the following modification of the Units of Good Case, and the graph that follows, which represents the possible outcomes in this more complicated case:

Complicated Units of Good Case

1,000 people are put into isolation booths. It is common knowledge that each must choose between Options A and B, with the following outcomes: If everyone chooses A, then each receives 99 additional units of good; if everyone chooses B, then each receives 100 additional units of good; otherwise, every person who chooses A receives 10 additional units of good and every person who chooses B loses a catastrophic 100,000 units of good, unless one and only one person chooses A, in which case every person receives 101 additional units of good.

The graph below represents the possible outcomes in this case:

If each chooses A (bringing about the yellow dot outcome on the graph), then for each it is true that any alternative act would have led to less good, and so each satisfies agent-neutral consequentialism; however, if each chooses B (bringing about the red dot outcome), then for each it is true that there is something else s/he could have done (namely, choose A) that would have led to more good (by bringing about the blue dot outcome), and so no one satisfies agent-neutral consequentialism. At the same time, it is certain that the outcome is better if everyone chooses B and does not successfully follow agent-neutral consequentialism than if everyone chooses A and successfully follows agent-neutral consequentialism. Finally, because of what each knows about how the others will choose, each knows that choosing A will make the outcome objectively better

than choosing B, and thus agent-neutral consequentialism requires each person to choose A; at the same time, it is certain that if everyone were to disobey agent-neutral consequentialism and chose B instead, the outcome would be better than if everyone chooses A, even though no one would then satisfy agent-neutral consequentialism. As a result, agent-neutral consequentialism is DCSD in this case, because it requires each to choose in a way that satisfies agent-neutral consequentialism, but that is certain to lead to a worse outcome than if each did not do what agent-neutral consequentialism actually requires and did not satisfy agent-neutral consequentialism instead. In other words, in this case each person can be certain that: *if we all successfully follow agent-neutral consequentialism, we will thereby cause our agent-neutral consequentialism-given aims to be worse achieved than they would have been if none of us had successfully followed agent-neutral consequentialism*.[125]

Is this a bad result for agent-neutral consequentialism? No. It would be a bad result for agent-neutral consequentialism if it were never DCSD, because we've seen that all plausible normative theories are sometimes DCSD.

Why then does Parfit think that agent-neutral consequentialism is never DCSD? Parfit offers the following sufficient conditions for direct collective self-defeat:

[125] Compare (i) on pg. 54 of Parfit, *Reasons and Persons*.

> A theory T is directly collectively self-defeating when:
>
> (i) it is *certain* that, if we all successfully follow T, we will thereby cause our T-given aims to be worse achieved than they would have been if none of us had successfully followed T, or
>
> (ii) our acts will cause our T-given aims to be best achieved only if we do not successfully follow T.[126]

Based on these conditions, Parfit offers the following argument that agent-neutral consequentialism is never DCSD:

> [Agent-neutral consequentialism] cannot be directly self-defeating, since it is agent-neutral: giving to all agents common moral aims. If we cause these common aims to be best achieved, we must be successfully following this theory. Since this is so, it cannot be true that we will cause these aims to be best achieved only if we do not follow this theory.[127]

[126] *Reasons and Persons*, pp. 54. Parfit notes that when considering agent-relative theories, we must focus only on (i), understanding it as: "it is certain that, if we all successfully follow T, we will thereby cause the T-given aims *of each* to be worse achieved than they would have been if none of us had successfully followed T" (pg. 55).

[127] *Reasons and Persons*, pp. 54-55.

In the last sentence of the preceding quote, Parfit concludes that it is necessarily false that: *our acts will cause our agent-neutral consequentialist aims to be best achieved only if we do not successfully follow agent-neutral consequentialism*, which is an instance of (ii), where 'agent-neutral consequentialism' replaces 'T'. From this, it is supposed to follow that agent-neutral consequentialism is never DCSD.

At this point, someone might object to Parfit's argument as follows: "On the analysis above, a theory can be DCSD in either way (i) or way (ii), and Parfit has shown only that agent-neutral consequentialism cannot be DCSD in way (ii); so, it doesn't follow from Parfit's premises that agent-neutral consequentialism cannot be DCSD in way (i), and so it doesn't follow that agent-neutral consequentialism is never DCSD."

In reply to this objection, Parfit would presumably insist that (i) is to be understood in such a way that (i) implies (ii). If that's right, then Parfit's demonstration that agent-neutral consequentialism can never be DCSD in sense (ii) also shows that it can never be DCSD in sense (i).[128]

---

[128] Along these lines, Parfit would presumably insist that (i) is to be understood in essentially the following way:

> (i*) for every possible combination of actions open to us x: if as a result of x we would each satisfy T, then our T-given aims would be worse achieved as a result x than they would have been if none of us had satisfied T. An individual satisfies T if and only if it is true that her T-given aims would have been worse achieved if she had acted differently.

This statement is complicated and difficult to parse. For present purposes, there are three things to note: first, (i*) makes clear what Parfit means by "it is certain that" in (i) (See Parfit's discussion on pp. 53-54 of *Reasons and Persons*); second, (i*) makes clear what Parfit means by "successfully follow T" in (i) (See pg. 106 of *Reasons and Persons*: "Why is it true [in a particular case] that we successfully follow [a theory] M? Because each is doing what, of the acts that are possible for him, best achieves his M-given aims."; see also pg. 53); finally, (i*) implies (ii), which vindicates the reply to the objection outlined above, because (i*) is essentially what Parfit means by (i).

However, Parfit's argument still faces a decisive objection. To see the problem, first note that even if (i) implies (ii), Parfit's argument is still invalid as stated:

> If C is ever (i) or (ii), then C is sometimes DCSD.
>
> C is never (ii).
>
> Therefore, C is never (i), since (i) implies (ii).
>
> Therefore, C is never DCSD.

If the problem is not immediately apparent, it might help to combine the two middle claims:

> If C is ever (i) or (ii), then C is sometimes DCSD.
>
> It is not the case that C is ever (i) or (ii).
>
> Therefore, it is not the case that C is sometimes DCSD.

This argument is invalid because it denies the antecedent. To get a valid argument, we would have to understand the first premise as a biconditional, and thus we would have to interpret (i) and (ii) as together yielding a full analysis of direct collective self-defeat.

However, Parfit explicitly claims that (i) and (ii) provide only sufficient conditions for direct collective self-defeat, and not a full analysis.[129] As a result, Parfit's argument is invalid, because it has the invalid form above.

Of course, this raises the question of whether (i) and (ii) can in fact yield a full analysis of direct collective self-defeat – in other words, it raises the question of whether the following is true:

> A theory T is DCSD when *and only when* either (i) is true or (ii) is true, where (i) and (ii) are understood in the way that Parfit intends.

This *Implicit Analysis* is false, because it does not capture the essence of direct collective self-defeat, including the essential idea that a theory is DCSD when it *directs us toward outcomes that are certain to be worse*.[130] In particular, the Implicit Analysis fails to deliver the correct verdict on the cases discussed above in which:

> (iii) it is common knowledge that: everyone knows the relevant facts, will act freely, will satisfy their normative requirements, and everyone also

[129] Parfit makes this explicit in the following passage, where he explains how he intends the phrase "[A theory T is] directly collectively self-defeating when…" to be understood: "By 'when' I do not mean 'only when'" (*Reasons and Persons*, pg. 54).

[130] *Reasons and Persons*, pg. 54.

> knows that: if each does what T actually requires, the T-given aims of each will be worse achieved than they would have been if none had done what T actually requires.

At the very least, such cases show that (iii) is an additional sufficient condition for direct collective self-defeat, which means that Parfit's argument that agent-neutral consequentialism is never DCSD cannot be salvaged, because if (iii) is sufficient for direct collective self-defeat, then it follows from the units of good cases that agent-neutral consequentialism is sometimes DCSD.

In response, a defender of the Implicit Analysis might say "But consider the possibility that in the Complicated Units of Good Case one and only one player chooses Option A; then, each person successfully follows agent-neutral consequentialism and brings about the best outcome; this shows that even in the Complicated Units of Good Case agent-neutral consequentialism does not direct us toward outcomes that are certain to be worse."

This reply gives the phrases 'direct us toward' and 'successfully follows' a perverse and non-standard meaning that is very different from their meaning in the intuitive thought that a theory is DCSD when it directs us toward outcomes that are certain to be worse, or when it is certain that the outcome would be worse if each successfully followed the theory than if each did not. In particular, this reply involves a backward-looking conception of directing an agent toward an outcome and successfully following a theory

that is irrelevant to any interesting normative concept. To see why, return to the players in the Units of Good Case and assume that all the players evaluate options and make their decisions simultaneously as well as independently. Now consider the point in time as they are about to make their decisions. At that point in time, does agent-neutral consequentialism direct the players toward a particular outcome? It does – namely, the outcome in which everyone chooses Option A: after all, even before anyone chooses Option A, choosing Option A has greater expected good than choosing Option B from the perspective of each player (indeed, each player *knows* that choosing Option A will lead to an objectively better outcome), and thus agent-neutral consequentialism directs each player to choose Option A, and thus each player successfully follows agent-neutral consequentialism only if that player chooses Option A.

The Implicit Analysis denies all of this. Instead, on that analysis agent-neutral consequentialism gives the players no direction at all before their decisions are made, on the grounds that there are multiple combinations of choices that would result in satisfaction of agent-neutral consequentialism. That is how the Implicit Analysis insists that agent-neutral consequentialism does not direct the players away from the best outcome: according to the analysis, there are no facts about what agent-neutral consequentialism directs the players to do until after everyone has made their decision, at which point the theory 'directs' everyone to have chosen in such a way that they now each satisfy agent-neutral consequentialism. However, that is not only a revisionary account of how agent-neutral consequentialism directs agents toward outcomes – because it entails that what agents know about the consequences of their choices has no relevance to what consequentialism directs them to do – it is also an unacceptable account, because

any interesting normative theory must provide direction for our decisions, and not only after they are made.

In response, a defender of the Implicit Analysis could attempt to bite the bullet and simply insist that agent-neutral consequentialism offers no direction in such cases until after decisions are made. However, the costs of such a stance prove unacceptably high when applied to other cases, especially cases that involve physical indeterminacy with no residual epistemic uncertainty. For example, consider a case that is somewhat similar to the Units of Good Case, but where the uncertainty of the outcomes derives entirely from physical indeterminacy:

One-Player Units of Good Case

You know that you alone must choose between the following two options, and that your choices will have the following consequences for yourself and 999 other innocent people:

Option A: 99% chance that everyone receives 99 additional units of good; 1% chance that everyone receives 10 additional units of good.

Option B: 1% chance that everyone receives 100 additional units of good; 99% chance that everyone receives negative 100,000 units of good.

Suppose that the chances in this case are purely physical and that there is no residual epistemic uncertainty. (For example, suppose that physicists have designed a non-deterministic pleasure and pain dispensing device to have these properties; you will simply choose whether to press the 'Option A' or 'Option B' button.)

On any sensible interpretation, agent-neutral consequentialism directs you to choose Option A in this case, which means that you successfully follow agent-neutral consequentialism only if you choose Option A. Would defenders of the Implicit Analysis agree? If they do not, then they are committed to the view that agent-neutral consequentialism (almost) never provides any guidance to our decisions at all, partly because physical indeterminacy always underlies all of our decisions. So, to avoid this result, they would presumably agree that agent-neutral consequentialism directs you to choose Option A in this case.

But if that is right, then there is a powerful argument that agent-neutral consequentialism directs each player to choose Option A in the original Units of Good Case. For consider that, for each player in that original case, there is some distribution of credences that that player ought to have, given his or her evidence, about how the other players will choose. Given that distribution of rational credences, we can imagine a one-player game with the same outcomes and probabilistic structure, but where the probabilities arise from physical indeterminacy with no residual epistemic uncertainty as in the One-Player Units of Good Case. If, as we are assuming, agent-neutral consequentialism directs you to choose Option A in the One-Player Units of Good Case, then it also directs each player to choose Option A in the one-player game that is derived in such a way from his or her rational

credences in the original Units of Good Case. But if agent-neutral consequentialism directs each player to choose Option A in the one-player games that are derived from their rational credences, then it also directs each player to choose Option A in the original Units of Good Case itself, because there is no normatively relevant difference between the choices that each individual would face in those one-player games and the corresponding choices that they face in the original Units of Good Case. As a result, initial defenders of the Implicit Analysis are forced either to abandon that analysis by admitting that agent-neutral consequentialism directs players to choose Option A in the original Units of Good Case, or else to bite an unacceptable bullet and insist that agent-neutral consequentialism almost never provides any guidance to our decisions at all, because physical indeterminacy always underlies our decisions.

The preceding discussion shows the importance of distinguishing between a normative theory's theory of objective value and its theory of choice. As we have just seen, a theory of objective value never directs us toward any outcomes itself – it is only in conjunction with a theory of choice that we are directed to make particular choices, and thereby directed toward particular outcomes. This shows that the analysis of direct collective self-defeat under consideration must be inadequate, because that analysis focuses only on satisfaction of a theory's theory of objective value, and not on satisfaction of its theory of choice. In other words, that analysis must be inadequate because the notion of direct collective self-defeat is about what a theory directs us toward, and without a theory of choice a theory never directs us toward anything at all.

For these reasons, an adequate analysis of the conditions under which a normative theory is directly collectively self-defeating must be tied to that normative theory's theory of choice. Here is a proposal:

> A theory T is directly collectively self-defeating (DCSD) when: (from the perspective of each of us) it is certain that, if each of us successfully follows T's theory of choice, we will thereby cause our T-given aims to be worse achieved than they would have been if none of us successfully followed T's theory of choice.[131]

[131] If successfully following T's theory of choice is the same as doing what T requires, then this is equivalent to the claim that: *A theory T is directly collectively self-defeating (DCSD) when: (from the perspective of each of us) it is certain that, if each of us does what T requires, we will thereby cause our T-given aims to be worse achieved than they would have been if none of us did what T requires*. On the assumption that individuals are always required to maximize expected good, a different way of putting the argument of this chapter is that even when it is common knowledge that everyone has full information about the choices facing everyone and that everyone is disposed to follow their normative requirements, in some cases it is clear that the option that has the highest expectation of good from the perspective of every individual is an option that would lead to a worse outcome than some other antecedently identifiable option if universally chosen. I do not put the argument this way here because I do not believe that there is any sense in which individuals are always required to maximize expected good. To see why, imagine that you have to choose between Option A and Option B, and that the expected value of Option A is 100 and the expected value of Option B is 99, because if you choose Option B you will bring about 99 additional units of good for sure, whereas if you choose Option A there is a 99% chance of bringing about 105 additional units of good, and a 1% chance of bringing about an extra-bad negative 395 units of good. In such a case, a risk-averse choice of Option B is often required, especially when the good that is at stake is the good of risk-averse others and not merely your own good. Although this shows that individuals should not always act to maximize expected good, it is consistent with the truth of consequentialism, and is consistent with the idea that how individuals should choose is a function of their evidence – the point is merely that a different theory of choice is needed to replace the theory that individuals must always choose in a way that maximizes expected good.

This New Analysis provides an analysis of direct collective self-defeat for cases in which everyone has two options. To test this analysis, we can consult our judgments about cases, and our judgments about the concept of direct collective self-defeat. Upon reflection, the New Analysis delivers the correct verdict on all of the cases that theorists have discussed in connection with direct collective self-defeat, and, unlike the Implicit Analysis discussed above, also fits our concept of direct collective self-defeat, according to which a theory is DCSD when it directs each of us to act in a way that is certain to be worse than if everyone did not follow the theory's directions instead.

A defender of the Implicit Analysis might raise the following objection: "Perhaps the New Analysis and/or (iii) captures the idea that a theory is DCSD when it directs us toward outcomes that are certain to be worse. But that idea is inconsistent with other more firmly held beliefs that we have about self-defeat, and so the notion of direct collective self-defeat must be regimented in a different way – most likely, in the way the Implicit Analysis suggests. That is because Donald Regan and others have provided cases that show that normative theories sometimes direct us away from the best outcomes, but are not thereby self-defeating." What the objector has in mind are cases like the following:

Miners Case

Suppose that several miners are trapped, with floodwaters rising. Before we can find out where these men are, we must decide which floodgate to close. The outcomes would be these:

		The men are in	
		Shaft A	Shaft B
	Gate 1	We save ten	All die
We close	Gate 2	All die	We save ten
	Gate 3	We save nine	We save nine

Assume that, on the evidence, the men are equally likely to be in either shaft.[132]

In this case, we are required to close Gate 3, even though it is certain that we will thereby bring about an outcome that isn't best; nonetheless, this does not show that normativity is directly collectively self-defeating. Does this undermine the idea that (iii) is a sufficient for direct collective self-defeat?

[132] This example is taken verbatim from Parfit, "What We Together Do", pp. 2-3, who follows Donald Regan, *Utilitarianism and Cooperation*, pg. 265.

It does not. What the Miners Case shows is that there is a crucial distinction between, on the one hand, *choosing an option that is certain to lead to an outcome that is not best* and, on the other hand, *choosing an option that is certain to lead to a worse outcome than an antecedently identifiable option*, and that a theory is DCSD when it directs us to choose in the latter way, but not when, as in the Miners Case, it merely directs us to choose an option that is certain to be not best. In particular, if we close Gate 3 we bring about an outcome that is certain to be not best, but we do not bring about an outcome that is certain to be worse than the outcome of an antecedently identifiable option, because there is no other option that is antecedently *certain* to lead to a better outcome than closing Gate 3. This is in perfect tune with conditions (i), (ii), and (iii) above, because the natural way of extending those conditions to many-option cases such as the Miners Case is by claiming that a theory is DCSD when it directs everyone to choose an option that is *certain* to lead to a worse outcome than an antecedently identifiable alternative option that it could have directed everyone to choose instead – but not when, as in the Miners Case, the theory merely directs everyone to choose an option that is certain to lead to an outcome that is not best.[133] As a result, the Miners Case does not ultimately raise a problem for the intuitive notion of direct collective self-defeat, and does not raise a problem for the view that conditions (i), (ii), and (iii) are each sufficient for direct collective self-defeat.

[133] Such an extension presumably must be restricted to cases in which everyone chooses between the 'same' options, where those options are individuated in a 'natural' way – in other cases the notion of direct collective self-defeat seems to have no clear application. In any event, the Miners Case could also be described as a two-option case, where Option One is to close Gate 3, and Option Two is to close one of the other gates. The New Analysis also delivers the correct verdict given that description, because choosing Option One is *not certain* to lead to a worse outcome than choosing Option Two, and therefore the New Analysis does not imply that morality is DCSD in the Miners Case.

Theorists might also be interested in whether a theory can be collectively self-defeating when it is *totally-successfully followed*, where a theory T is totally-successfully followed if and only if: everyone successfully follows T's theory of choice by doing what is actually required, and everyone satisfies T's theory of objective value by performing an act such that any other act would have made their T-given aims worse achieved. The Complicated Units of Good Case shows that for any plausible normative theory T (including agent-neutral consequentialist theories), it is sometimes clear that by totally-successfully following T, we would thereby cause our T-given aims to be worse achieved than they would be if no one successfully followed T's theory of choice and no one satisfied T's theory of objective value. This shows that all plausible normative theories are sometimes collectively self-defeating even when each individual does what is (known to be) both "subjectively" and "objectively right".

In response to all of this, it might be claimed that agent-neutral consequentialism is still the only normative theory on which it is always metaphysically possible to bring about the outcome that is best without anyone acting in a way that is wrong. However, the units of good cases show that this is false unless the word 'wrong' is given a stipulative definition that has no connection to our actual normative concepts, because the only way everyone could bring about the outcome that is best in those cases is by each acting in a way that each actually knows would lead to an objectively worse outcome. This shows that agent-neutral consequentialism has no interesting advantage over other types of ethical theories with respect to collective self-defeat.

For related reasons, the arguments above cannot be dismissed by simply insisting on a definition of direct collective self-defeat on which (iii) is not a sufficient condition for direct collective self-defeat. In part, this is because direct collective self-defeat, like knowledge, is a notion that we track and have an interest in prior to seeing any stipulative definition, as is illustrated by our interest in prisoner's dilemmas and other situations in which self-interest is DCSD – and so direct collective self-defeat is not a notion that we are free to define however we like. More specifically, insofar as we should care whether a theory is sometimes directly collectively self-defeating, that is because having that property means that regrettable consequences are assured even in cases like those described in (iii) in which it is common knowledge that everyone knows the relevant facts and will successfully follow the theory. As a result, a definition on which satisfaction of (iii) is not sufficient for direct collective self-defeat has no practical or theoretical interest, not only because it does not track the notion of 'directing us toward outcomes that are certain to be worse', but more importantly because it does not track the kind of collective self-defeat that it is regrettable for a theory to imply – because the most regrettable form of collective self-defeat is when a theory is collectively self-defeating in the sense of (iii), when it is collectively self-defeating even though it is common knowledge that everyone knows the relevant facts and will successfully follow the theory, and that regrettability is not mitigated in any interesting way when it is also true that if individuals had failed to do what they actually know they are required to do, the outcome could have been better. As a result, any discussion that rejects (iii) as a sufficient condition for direct collective self-defeat is doomed to reduce to a mere definitional exercise that has no connection to any property that we should care whether a

normative theory has – whereas endorsing (iii) is essential to capturing the kind of collective self-defeat that is of central interest from both a practical and theoretical perspective.[134]

The preceding discussion suggests the following evaluation of Parfit's main argument for never-directly-collectively-self-defeating moral theories:

Parfit's Main Argument

To be plausible, a moral theory must be never DCSD.

So, we must reject common-sense morality and other theories that are sometimes DCSD, and instead endorse a version of never-DCSD moral theory.

The first premise is false, because morality is sometimes DCSD. As a result, not only is it consistent to deny the conclusion, but there is decisive reason for thinking that it is false, because a moral theory is false if it is never DCSD. If we follow Parfit in thinking that consequentialists, contractualists, Kantian theorists, and most others have been "climbing the same mountain" toward the goal of developing the most plausible version of never-

[134] An additional consideration is that many theorists, including Derek Parfit, take facts about what would be wrong when agents know the relevant facts as explanatorily fundamental – which provides decisive reason to think that the sense of direct collective self-defeat captured by (iii) is the sense that must have the greatest theoretical importance, because (iii) is explicitly concerned with whether a theory is collectively self-defeating when everyone knows the relevant facts. (See Parfit, *On What Matters*, Volume One, Section 21.)

DCSD moral theory, then this means that those theorists have all been climbing the wrong mountain.[135]

This is not to denigrate Parfit's work. Every aspect of this discussion was made possible by Parfit's work, and his work has the highest virtues of clarity, testability, originality, and importance. Because Parfit's work has such virtue, identifying a clear objection to his arguments leads to important progress in normative theory.

In sum, morality and all other interesting forms of normativity are sometimes dramatically directly collectively self-defeating, which means that many influential normative theories are either false, or at least don't have the consequences that their adherents take them to have. In particular, morality and other forms of normativity cannot be relied upon to solve collective action problems even in a world of normatively flawless agents. A practical upshot is that many of the most important questions about modern moral life cannot be answered by asking 'But what if everyone did that?', or by more sophisticated appeals to 'universalizability'.

[135] For this metaphor and a summary of Parfit's arguments that the most plausible versions of consequentialism, contractualism, and Kantian ethics all imply that morality is never DCSD, see *On What Matters*, Volume One, pp. 25-26. Parfit endorses Parfit's Main Argument in *On What Matters*, Volume One, pg. 306: "In [each-we dilemmas], in acting on common sense moral principles, we are acting in ways that are directly collectively self-defeating. If we were Rational Egoists, that would be no objection to our view, since this form of Egoism is a theory about individual rationality and reasons. But moral principles or theories are intended to answer questions about what all of us ought to do. So such principles or theories clearly fail, and condemn themselves, when they are directly self-defeating at the collective level". See *Reasons and Persons*, pg. 111 and pg. 113 for an earlier discussion and more explicit presentation of the argument.

The Equilibrium Objection

In "Group Morality", Frank Jackson uses an example that bears some similarity to the Stampede Case to argue that it is possible to "have a group action which is wrong, yet every constituent act is right; and a group action which is right yet every constituent act is wrong" (pg. 102). Parfit accepts Jackson's conclusions in later work, but neither Parfit nor Jackson take these conclusions to show that morality is sometimes directly collectively self-defeating.[136] In what follows I show that Jackson's conclusions do not clearly follow from the example he discusses, and that his discussion cannot be extended to show that morality is sometimes directly collectively self-defeating (DCSD) – but that such conclusions are vindicated by the examples discussed above, despite an important objection that is suggested by reflection on Jackson's discussion.

Here is Jackson's example:

> [Suppose that] There is a steady stream of traffic going to work. Everyone is driving at 80 kilometres per hour. It would be safer if everyone was

[136] Jackson does not claim that his conclusions show that morality is sometimes DCSD, and in later work Parfit continues to rely on the premise that morality is never DCSD despite Parfit's endorsement of Jackson's conclusions in "What We Together Do".

> driving at 60. The right group action is for everyone together to drive at 60. But what about each person, should he or she drive at 60? The answer may well be no; for it may well be the case that if he or she were to drive at 60, everyone else would still drive at 80, and so a lot of dangerous overtaking would result. For each individual the right action is to keep driving at 80, so avoid dangerously disrupting the traffic flow; yet the right group action is for everyone to drive at 60. Thus, we have in this example a right group action – everyone together driving at 60 – with each and every constituent individual action – each action of a person driving at 60 – wrong. And also we have a wrong group action – everyone together driving at 80 – with each and every constituent action – each action of a person driving at 80 – right. We see, therefore, that not even the attractive-sounding principle that if a group action is right, at least one of its constituent acts is right, is valid.

This case presupposes some initial wrongdoing by some individuals – in particular, the initial drivers who break the speed limit – which means that the case does not show that a morally suboptimal outcome would result if *everyone* followed morality, which means that the case does not show that morality is sometimes DCSD. Furthermore, even if we imagine a group of morally flawless agents somehow 'thrown into' the case Jackson describes as in a stampede, the case still does not clearly show that morality is sometimes DCSD, and for similar reasons does not support Jackson's own conclusions.

The problem is that, contrary to what Jackson tacitly assumes, each individual driver can choose among a wide range of possible speeds. This detail undermines Jackson's argument, because although no individual driver is required to reduce his or her speed instantaneously to the morally ideal speed of 60, nonetheless at each moment each individual is required to reduce his or her speed *slightly* – which means that if everyone in the group follows morality, the morally ideal speed of 60 will be reached by the group *in the morally optimal way given the group's starting point*. (Upon reflection, it is clear that this is exactly what morality would require in such a case, on the assumption that it is common knowledge that morality will be universally followed.) As a result, if everyone follows morality, this leads to the morally optimal outcome of everyone driving 60, and it leads to that outcome along a path that is also morally optimal given the relevant starting point – which arguably means that if each person does follow morality along that ideal path, then the group itself also acts rightly at each moment along that path, given its suboptimal starting point. As a result, Jackson's case does not clearly support his conclusions that it is possible to "have a group action which is wrong, yet every constituent act is right; and a group action which is right yet every constituent act is wrong". Furthermore, even if 'is wrong' is stipulated to mean 'has a suboptimal instantaneous outcome' (as Jackson intends),[137] Jackson's case is still consistent with the idea, and might even seem to illustrate the truth of the idea, that the optimal course of action for a group is in perfect harmony with the optimal course of action for each of its constituent individuals *whenever a stable equilibrium develops as a result of every individual following morality*.

[137] See Jackson's discussion of 'objectively right' on pg. 92.

It could be claimed that this *equilibrium objection* also undermines the force of the Stampede Case discussed above. However, a crucial difference is that in a stampede, in contrast to highway traffic, individuals have only two real options: continue stampeding at the dictated rate, or else be trampled – and if everyone continues stampeding at the dictated rate, then all individuals will continue to have only those two options, ensuring that the ultimate outcome never tends toward an equilibrium that is morally desirable, given realistic assumptions.[138]

More importantly, even if such an equilibrium explanation were available for the stampede cases, such an explanation is not available regarding the units of good cases discussed above, because those latter cases involve a 'one-shot' decision situation in which it is simply impossible for a desirable equilibrium to develop in the way the equilibrium objection assumes. As a result, those cases provide a decisive demonstration that morality and all other forms of normativity are sometimes dramatically DCSD, and a decisive demonstration that the best course of action for a group can come radically apart from the best course of action for each of its constituent individuals, even when a stable equilibrium develops as a result of each individual following morality.

[138] Another crucial difference is that stampedes arise without any wrongdoing by any individual, unlike Jackson's example involving high-speed highway traffic.

Conclusions

In the first part of this dissertation, I focused on how to secure global cooperation on a comprehensive emissions-reduction scheme, which is the most important and the most intractable problem that stands in the way of an effective global response to climate change. I provided a straightforward, realistic, and non-coercive solution that is superior to all competitors along both ethical and practical dimensions, growing in superiority as assumptions about the dispositions of powerful nations are made increasingly pessimistic. In particular, I explained how a simple game theoretic result shows that even when our existent tools for solving collective action problems are useless, we can often escape a social dilemma by turning the force of such a dilemma against itself, by intentionally engineering and placing ourselves into a new dilemma, the predictable outcome of which is a solution to the initial, otherwise insoluble dilemma. A climate treaty can use this strategy by intentionally creating something like a multi-player prisoner's dilemma at the level of nations, the predictable outcome of which is universal ratification and long-run compliance to an effective climate treaty, even if compliance makes the current citizens of many nations worse off than the no-treaty status-quo, and even if nations are disposed not to comply whenever it is in the interest of their current citizens not to comply. I also explained why such a treaty is the key to engineering the most *ethical* climate treaty

possible, and to making progress beyond the unpromising *incrementalist*, *idealist*, and *hard-headed realist* approaches to climate treaties that dominate the literature in philosophy, politics, economics, and law. The result is a 'meta-architecture' for securing agreement and compliance to a chosen global response to climate change that remains relatively agnostic about the details and architecture of the global response that ought to be chosen. More generally, the result is a framework for constructing economically and ethically optimal solutions to many previously intractable collective action problems.

In the second part, I examined what individuals are required to do in large real-world collective action situations, especially when collective cooperation is implausible, and especially when collective non-cooperation would lead to catastrophe. I argued that even given worst-case-scenario assumptions about the effects of climate change, individual citizens are not required to reduce their emissions by a significant amount, because without an effective global response such reductions are both costly and futile in a way that makes such reductions not required. In fact, I argued that such futility and lack of requirement at the individual level explains why individuals are required to favor a coercive intergovernmental response to climate change, and why nations are justified in adopting such a response even if it is unfair to their citizens and contrary to their interests. In general, I showed that nations can be justified in acting in ways that make their citizens worse off, are unfair to their citizens, and that would be reasonably rejected by their citizens – which provides a counterexample to many political theories.

Along the way, I made progress on a number of issues at the foundations of normative ethics and political philosophy. I analyzed the ethical significance of futility,

distinguishing between, on the one hand, cases in which actions are required despite being futile in some intuitive sense, and, on the other hand, cases in which actions are not required because they are futile in a more complete sense that makes actions genuinely not required. I showed that existent normative theories are unable to offer a plausible account of what individuals are required to do in the kind of collective action situations that are common in a market-based society, and are therefore unable to explain many of the most important facts about modern moral life. In particular, I noted that straightforward mathematical and empirical considerations show that an appeal to expected consequences cannot possibly deliver the verdicts on such cases that consequentialists themselves assume, and I showed that alternative theories that appeal to 'universalizability', 'direct harm', and other notions also cannot deliver plausible verdicts on such cases. For these and other reasons, I argued that a plausible account of what individuals are required to do in a large market-based society must invoke a distinction between activities that are *essential* to a product or to the actual production of a product, and activities that are not.

In the final chapters, I showed that morality and other forms of normativity are sometimes *dramatically directly collectively self-defeating*, where a normative system has that property when it requires everyone to act in a way that everyone can see is certain to be dramatically worse along the relevant normative dimension than if everyone did not follow the requirements of that normative system instead. This means that a wide range of normative theories are either false, or at least don't have the consequences that their adherents take them to have, including consequentialist theories, contractualist theories, Kantian theories, universalization theories, enlightened self-interest theories, and many

other normative theories. It also means that morality and other forms of normativity cannot be relied upon to solve collective action problems even in a world of normatively flawless agents. One practical upshot is that even when a disaster will ensue if everyone acts in a particular way or on a particular principle, that doesn't settle the question of whether individuals are permitted to act in that way or on that principle.

In sum, I have provided answers to many of the most pressing questions about what we should do about difficult collective action problems, especially when non-cooperation would lead to catastrophe.

Acknowledgements

First and foremost, thanks to Gideon Rosen, who served as my primary advisor for this project. Any attempt to describe his great kindness, generosity, and integrity would be inadequate.

Thanks also to Stewart Cohen and John Devlin for being similarly great advisors for an earlier project, and to Joe Rachiele, George Rudebusch, Peter Singer, and Jane Willenbring for their advice and very generous help with this project.

Along with many of the people I have just thanked, Jack Spencer and Jack Woods provided encouragement for this project when I needed it and in a way that mattered. Special thanks also to Tom Blackson, Greg Fitch, Chris Griffin, Gil Harman, Tom Kelly, Nick Kroll, Mike Malone, Tristram McPherson, Steve Nelson, Ryan Robinson, Dave Schmidtz, David Sherry, Heidi Speck, and all of my other colleagues at Northern Arizona University, Arizona State University, the University of Arizona, and Princeton University. McPherson's work on animal ethics was the catalyst for this entire project; I am also especially indebted to the work of Tom Schelling, Peter Singer, and Judith Jarvis Thomson.

Many years ago, at Moma's Pizza in Tempe, Arizona, Charlie Kurth explained to me how the widespread phenomenon of market failure shows that laissez-faire capitalism is indefensible. At the time, it may have seemed that I wasn't paying attention. I owe similar thanks to many others for correcting many other embarrassingly confused views that I have held, and for helping me see the limits of my powers of judgment.

Thanks also to my mom and dad, Jan Bryant and Garry Budolfson, and to everyone else in my family.

In the preceding I have tried to thank everyone who it would be wrong not to thank in this context. In what follows I will try to thank everyone who I should thank for their specific help with this dissertation. I apologize for any omissions.

Thanks to Gideon Rosen, Peter Singer, and Jane Willenbring for incredible comments on the entire dissertation. Thanks also to John Devlin, Chris Griffin, Derek Parfit, Joe Rachiele, Ryan Robinson, and George Rudebusch for especially generous discussions. In addition, thanks to Sameer Bajaj, John Basl, Heather Berginc, Tom Blackson, Cheshire Calhoun, Richard Yetter Chappell, Tim Clark, Stewart Cohen, Christian Coons, Michael Cox, Jeff Downard, David Faraci, Ada Fee, Brian Fiala, Zach Forman, Liz Harman, Travis Hoffman, Pat Lambrecht Hould, Stan Hould, Andy Hultgren, Bruce Hultgren, Ryan Jenkins, Robert Keohane, Victor Kumar, Joel MacClellan, Sarah McGrath, Tristram McPherson, Nathan Meyer, Michael Oppenheimer, Guido Pincione, Ángel Pinillos, Dan Russell, Carolina Sartorio, Dave Schmidtz, Dan Shahar, Daniel Silvermint, John Thrasher, Hannah Tierney, Ian Vandeventer, Chad Van Schoelandt, Steve Wall, Brandon Warmke, Michael Weber, Sara Worley, and audiences at the University of

Arizona Center for the Philosophy of Freedom (2011 and 2012), Bowling Green State University (2012), the 2011 Colorado State University Animal Ethics Conference, and Northern Arizona University, where I first presented the arguments regarding climate change and individual requirements in 2009.

There is an interesting analogy between writing acknowledgments and deciding who to invite to a wedding. In both cases, you don't want to leave anyone out – but in both cases, you must leave some out, even if only unintentionally – and as the number who you don't leave out grows, so does the insult to those who you do leave out. Perhaps the best strategy is to 'elope', and thus include no one. But that also seems like a cowardly strategy – but, on the other hand, the best strategy often does seem cowardly, and so perhaps it ultimately requires great courage? Another alternative is to fret and wring hands until the problem consumes your entire life.

Finally, you might publicly announce some principle by which you will make your decision, where that principle will inevitably seem suboptimal to many – but, if properly designed, would at least seem reasonable to all as a principle by which to make the relevant decision if everyone fully understood the need for such a principle and if everyone fully understood the 'you can't please all of the people all of the time' nature of the problem – even if that principle could still be reasonably rejected by many on the grounds that they would have done much better under an equally justifiable alternative. That has been my goal here.

As a final note, I wish I could celebrate the completion of this dissertation by having a beer and going surfing with my friend Greg Fitch.

Bibliography

Adams, Robert. "Should Ethics be More Impersonal?" *Reading Parfit*. Ed. Jonathan Dancy. Oxford: Blackwell Publishers: 1997. 259.

Aldy, Joseph and Robert Stavins, Ed. *Architectures for Agreement: Addressing Global Climate Change in the Post-Kyoto World*. New York: Cambridge University Press, 2007.

---. *Post-Kyoto International Climate Policy: Implementing Architectures for Agreement*. New York: Cambridge University Press, 2009.

---. *Post-Kyoto International Climate Policy: Summary for Policymakers*. New York: Cambridge University Press, 2009.

Almeida, Michael, Ed. *Imperceptible Harms and Benefits*. Kluwer: Netherlands, 2000.

Archer, David. *The Global Carbon Cycle*. Princeton, NJ: Princeton University Press, 2010.

---. *The Long Thaw: How Humans Are Changing the Next 100,000 Years of Earth's Climate*. Princeton, NJ: Princeton University Press, 2008.

Archer, David and Stefan Rahmstorf. *The Climate Crisis: An Introductory Guide to Climate Change*. New York: Cambridge University Press, 2010.

Arntzenius, Frank and David McCarthy. "Self Torture and Group Beneficence." *Erkenntnis* 47.1 (1997): 129-144.

Barrett, Scott. *Environment and Statecraft: The Strategy of Environmental Treaty-Making.* New York: Oxford University Press, 2003.

---. *Why Cooperate? The Incentive to Supply Global Public Goods*. New York: Oxford University Press, 2007.

Barry, Brian. "Comment." *Political Participation*. Ed. Stanley Benn. Canberra: Australian National University Press, 1978. 47.

Beitz, Charles. *Political Theory and International Relations*. Revised ed. Princeton, NJ: Princeton University Press, 1999.

Biermann, Frank and Rainer Brohm. "Implementing the Kyoto Protocol Without the USA: The Strategic Role of Energy Tax Adjustments at the Border" *Climate Policy* 4 (2005): 289-302.

Bostrom, Nick and Milan Ćirković, Ed. *Global Catastrophic Risks*. New York: Oxford University Press, 2008.

Brennan, Jason. *The Ethics of Voting*. Princeton, NJ: Princeton University Press, 2011.

Broome, John. "The Ethics of Climate Change: Pay Now or Pay More Later?" *Scientific American* 298 (2008), 96-102.

Campbell, T. Colin and Thomas M. Campbell II. *The China Study: The Most Comprehensive Study of Nutrition Ever Conducted*. Dallas: BenBella Books, 2006.

Coase, Ronald. "The Problem of Social Cost." *Journal of Law and Economics* 3 (1960): 1-44.

Dancy, Jonathan, Ed. *Reading Parfit*. Oxford: Blackwell Publishers: 1997.

Feinberg, Joel. "The Rights of Animals and Future Generations." *Philosophy and Environmental Crisis*. Ed. William Blackstone. Athens, GA: University of Georgia Press, 1974. 43-68.

Frank, Robert. *Microeconomics and Behavior*. 8th ed. New York: McGraw-Hill Education, 2009.

Frankel, Jeffrey. "Formulas for Quantitative Emissions Targets." *Architectures for Agreement: Addressing Global Climate Change in the Post-Kyoto World*. Ed. Joseph Aldy and Robert Stavins. New York: Cambridge University Press, 2007. 31-56.

Frey, R.G. "Utilitarianism and Moral Vegetarianism Again: Protest or Effectiveness." *Food for Thought: The Debate Over Eating Meat*. Ed. Steve F. Sapontzis. Amherst, NY: Prometheus Books, 2004. 118-123. Contemporary Issues.

Friedman, Milton. *Capitalism and Freedom*. Chicago: University of Chicago Press, 1962.

Gardiner, Stephen, Simon Caney, Dale Jamieson and Henry Shue, Eds. *Climate Ethics: Essential Readings.* New York: Oxford University Press, 2010.

Gardiner, Stephen. "Ethics and Global Climate Change." *Ethics* 114 (2004): 555-600. Reprinted in *Climate Ethics: Essential Readings.* Ed. Stephen M. Gardiner, et al. New York: Oxford University Press, 2010. 3-38.

---. "Is 'Arming the Future' with Geoengineering Really the Lesser Evil?: Some Doubts about the Ethics of Intentionally Manipulating the Climate System." *Climate Ethics: Essential Readings.* Ed. Stephen M. Gardiner, et al. New York: Oxford University Press, 2010. 284-314.

---. *A Perfect Moral Storm: The Ethical Tragedy of Climate Change.* New York: Oxford University Press, 2011.

Gibbard, Allan. *Utilitarianism and Coordination*. Diss. Harvard University, 1971.

Glover, Jonathan. "It Makes No Difference Whether or Not I Do It." *Proceedings of the Aristotelian Society, Supplementary Volumes* 49 (1975): 171-209.

Goldman, Alvin. "Why Citizens Should Vote: A Causal Responsibility Approach." *Social Philosophy and Policy* 16.2 (1999): 201-217.

Hardin, Garrett. "The Tragedy of the Commons." *Science* 162 (1968): 1243-1248. Reprinted in *Environmental Ethics: What Really Matters, What Really Works*. Ed. David Schmidtz and Elizabeth Willott New York: Oxford University Press, 2002. 331-340.

Hart, H.L.A. "Are There Any Natural Rights?" *Philosophical Review* 64 (1955): 175-91.

Harvey, Hal and Sonia Aggarwal. "The Costs of Delay." San Francisco: ClimateWorks Foundation, 2011.

Hill, Thomas E. "Symbolic Protest and Calculated Silence." *Philosophy and Public Affairs* 9.1(1979): 83-102.

Harrod, Roy. "Utilitarianism Revised" *Mind* 45 (1936): 137-56.

Harrison, Jonathan. "Utilitarianism, Universalization, and our Duty to be Just." *Proceedings of the Aristotelian Society* 53 (1952-3): 105-34.

Intergovernmental Panel on Climate Change. *Climate Change 2007: The Physical Science Basis*. AKA: Solomon, Susan, Dahe Qin, Martin Manning, Zhenlin Chen, Melinda Marquis, Kristen B. Averyt, Melinda Tignor, Henry L. Miller, eds. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. New York: Cambridge University Press, 2007.

---. *Climate Change 2007: Mitigation of Climate Change*. AKA: Metz, Bert, Ogunlade R. Davidson, Peter R. Bosch, Rutu Dave, Leo A. Meyer, eds. *Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, New York: Cambridge University Press, 2007.

Ismer, Roland and Karsten Neuhoff. "Border Tax Adjustment: A Feasible Way to Support Stringent Emission Trading." *European Journal of Law and Economics* 24.2 (2007): 137-164.

Jackson, Frank. "Group Morality." *Metaphysics and Morality: Essays in Honour of J.J.C. Smart*. Ed. J. J. C. Smart, Philip Pettit, Richard Sylvan & Jean Norman. Oxford: Blackwell Publishers: 1987.

Kagan, Shelly. "Do I Make a Difference?" *Philosophy & Public Affairs* 39.2 (2011): 105–141.

Kennedy, John F. "American University Commencement Address" 10 June 1963. http://www.americanrhetoric.com. 15 January 2012.

Kutz, Christopher. *Complicity: Ethics and Law for a Collective Age*. New York: Cambridge University Press, 2000.

Levitt, Steven and Stephen Dubner. *SuperFreakonomics: Global Cooling, Patriotic Prostitutes, and Why Suicide Bombers Should Buy Life Insurance*. New York: HarperCollins, 2009.

Locke, John. *Second Treatise of Government*. Hackett Publishing, 1980.

Lomborg, Bjorn. *The Skeptical Environmentalist: Measuring the Real State of the World*. New York: Cambridge University Press, 2001.

Lomborg Bjorn, Ed. *Global Crises, Global Solutions: Costs and Benefits*. 2nd ed. New York: Cambridge University Press, 2009.

Maass, Peter. *Crude World: The Violent Twilight of Oil*. New York: Vintage Books, 2009.

MacClellan, Joel. "Size Matters: Animal Size, Contributory Causation, and Ethical Vegetarianism." *Journal of Animal Ethics*, forthcoming: Spring 2012.

McPherson, Tristram. "Why I Am A Vegan (and You Should Be One Too)." Unpublished MS. 1 January 2010.

Metz, Bert, Ogunlade R. Davidson, Peter R. Bosch, Rutu Dave, Leo A. Meyer, eds. *Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change,* New York: Cambridge University Press, 2007.

Michaelowa, Axel. "Graduation and Deepening." *Architectures for Agreement: Addressing Global Climate Change in the Post-Kyoto World*. Ed. Joseph Aldy and Robert Stavins. New York: Cambridge University Press, 2007. 81-104.

Moller, Dan. "The Morality of Overdetermination: Wrongdoing When it Makes No Difference." In *Abortion, Killing, and Overdetermination – Three Essays*. Diss. Princeton University, 2005.

Mueller, Dennis. *Public Choice III*. New York: Cambridge University Press, 2003.

Muller, Richard. *Physics and Technology for Future Presidents*. Princeton, NJ: Princeton University Press, 2010.

Norcross, Alastair. "Puppies, Pigs, and People: Eating Meat and Marginal Cases." *Philosophical Perspectives* 18 (2004): 229-245.

Nordhaus, William. *A Question of Balance: Weighing the Options on Global Warming Policies*. New Haven: Yale University Press, 2008.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.

Nussbaum, Martha. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, MA: Harvard University Press, 2007.

Olson, Mancur. *The Logic of Collective Action*. Cambridge, MA: Harvard University Press, 1965.

Ostrom, Elinor. *Governing the Commons*: *The Evolution of Institutions for Collective Action*. New York: Cambridge University Press, 1990.

---. "A diagnostic approach for going beyond panaceas." *Proceedings of the National Academy of Sciences* 104.39 (2007): 15181-15187.

---. "A General Framework for Analyzing Sustainability of Social-Ecological Systems." *Science* 325 (2009): 419-422.

Pauwelyn, Joost H. B. "U.S. Federal Climate Policy and Competitiveness Concerns: The Limits and Options of International Trade Law." Working paper, April 2007. http://www.nicholas.duke.edu/institute.

Parfit, Derek. “Comments.” *Ethics* 96.4 (1986): 832-872.

---. *On What Matters*. 2 vols. New York: Oxford University Press, 2011.

---. *Reasons and Persons*. New York: Oxford University Press, 1984.

---. “What We Together Do.” Unpublished MS. 29 March 1988.

Pinkert, Felix. “Coordination problems, Act Consequentialism, and the consequences of what we together do”. Unpublished MS. 15 December 2011.

Pogge, Thomas. *Realizing Rawls*. Ithaca, NY: Cornell University Press, 1989.

Pollan, Michael. *The Omnivore’s Dilemma: A Natural History of Four Meals*. New York: Penguin, 2006.

---. “Why Bother?” *The New York Times*. New York Times, 20 April 2008. Web. January 15, 2012.

Posner, Eric and Cass Sunstein. “Climate Change Justice.” *Georgetown Law Journal* 96(2008): 1565-1612.

Posner, Eric and David Weisbach. *Climate Change Justice*. Princeton, NJ: Princeton University Press, 2010.

Rachels, Stuart. “Counterexamples to the Transitivity of Better Than.” *Australasian Journal of Philosophy* 76.1 (1998): 71–83.

Railton, Peter. “Alienation, Consequentialism, and the Demands of Morality.” *Philosophy and Public Affairs* 13.2 (1984): 134-171. Reprinted in Peter Railton, *Facts, Values, and Norms*, New York: Cambridge University Press. 151-186.

Rawls, John. *The Law of Peoples*. Cambridge, MA: Harvard University Press, 2001.

---. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.

Regan, Donald. *Utilitarianism and Cooperation*. New York: Oxford University Press, 1980.

Regan, Tom. *The Case For Animal Rights, Updated with A New Preface*. Berkeley: University of California Press, 2004.

Robbins, John. *The Food Revolution: How Your Diet Can Help Save Your Life and Our World*. 10th Anniversary ed. San Francisco: Conari Press, 2011.

Scanlon, Thomas. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 2000.

Schafer-Landau, Russ. "Vegetarianism, Causation and Ethical Theory." *Public Affairs Quarterly* 8.1 (1994): 85-100.

Schaffer, Jonathan. "Contrastive Causation." *The Philosophical Review* 114.3 (2005): 297-328.

Schelling, Thomas C. "Bargaining, Communication, and Limited War." *Journal of Conflict Resolution,* 1.1 (1957): 19–36.

---. "An Essay on Bargaining." *The American Economic Review*, 46.3 (1956): 281-306.

---. "Some Economics of Global Warming." *American Economic Review* 82.1 (1992): 1-14.

---. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press, 1981.

---. "What Makes Greenhouse Sense? Time to Rethink the Kyoto Protocol." *Foreign Affairs*, 81.3 (2002): 2-9.

Schlosser, Eric. *Fast Food Nation: The Dark Side of the All-American Meal*. New York: HarperCollins,2002.

Schmidtz, David and Elizabeth Willott, Ed. *Environmental Ethics: What Really Matters, What Really Works*. New York: Oxford University Press, 2002.

Schmidtz, David. "The Institution of Property." *Social Philosophy and Policy* 11 (1994): 42-62. Reprinted in *Environmental Ethics: What Really Matters, What Really Works*. Ed. David Schmidtz and Elizabeth Willott. New York: Oxford University Press, 2002. 361-372.

Shahar, Dan. "Justice and Climate Change." *The Independent Review* 14.2 (2009): 219-237.

Shue, Henry. "Subsistence Emissions and Luxury Emissions." *Climate Ethics: Essential Readings.* Ed. Stephen Gardiner, et al. New York: Oxford University Press, 2010. 200-214.

---. "Subsistence Emissions and Luxury Emissions." *Law & Policy* 15.1 (1993): 39-59. Reprinted in *Climate Ethics: Essential Readings.* Ed. Stephen Gardiner, et al. New York: Oxford University Press, 2010. 200-214.

Singer, Peter. *Animal Liberation*. New York: Ecco Press, 2002.

---. “Famine, Affluence, and Morality.” *Philosophy and Public Affairs* 1 (1972): 229-243.

---. “Make Meat-Eaters Pay: Ethicist Proposes Radical Tax, Says They're Killing Themselves and the Planet.” New York Daily News, 25 October 2009. Web. 15 January 2012.

---. “One Atmosphere.” Reprinted in *Climate Ethics: Essential Readings.* Ed. Stephen Gardiner, et al. New York: Oxford University Press, 2010. 181-199.

---. *Practical Ethics*. New York: Cambridge University Press, 2011.

---. “Utilitarianism and Vegetarianism.” *Philosophy and Public Affairs* 9.4 (1980): 325-337.

---. *Writings on An Ethical Life*. New York: Ecco Press, 2000.

Singer, Peter and Jim Mason. *The Ethics of What We Eat: Why Our Food Choices Matter*. New York: Rodale Press, 2006.

Sinnott-Armstrong, Walter. “It’s Not My Fault: Global Warming and Individual Moral Obligations.” *Climate Ethics: Essential Readings.* Ed. Stephen Gardiner, et al. New York: Oxford University Press, 2010. 332-346.

Smart, J.J.C. “Extreme and Restricted Utilitarianism.” *Philosophical Quarterly* 6 (1956): 344-54.

Sober, Elliott. “Philosophical Problems for Environmentalism.” *The Preservation of Species: The Value of Biological Diversity*. Ed. Bryan Norton. Princeton, NJ: Princeton University Press, 1986. Reprinted in *Environmental Ethics: What Really Matters, What Really Works*. Ed. David Schmidtz and Elizabeth Willott New York: Oxford University Press, 2002. 145-157.

Solomon, Susan, Dahe Qin, Martin Manning, Zhenlin Chen, Melinda Marquis, Kristen B. Averyt, Melinda Tignor, Henry L. Miller, eds. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* New York: Cambridge University Press, 2007.

Stern, Nicholas. *The Economics of Climate Change: The Stern Review*. New York: Cambridge University Press, 2007.

Stewart, Richard and Jonathan Wiener. *Reconstructing Climate Policy: Beyond Kyoto*. Washington, D.C.: AEI Press, 2003.

Storr, Will. “The Rape of Men.” *The Guardian*. The Guardian, 16 July 2011. Web. January 15, 2012.

Sunstein, Cass. *Laws of Fear: Beyond the Precautionary Principle*. New York: Cambridge University Press, 2005.

---. *Risk and Reason: Safety, Law, and the Environment*. New York: Cambridge University Press, 2004.

---. *Worst-Case Scenarios*. Cambridge, MA: Harvard University Press, 2007.

Temkin, Larry. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. New York: Oxford University Press, 2012.

Thaler, Richard and Cass Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New York: Penguin Books, 2009.

Thomson, Judith Jarvis. “A Defense of Abortion.” *Rights, Restitution, and Risk: Essays in Moral Theory*. Ed. William Parent. Cambridge, MA: Harvard University Press, 1986. 1-19.

---. *The Realm of Rights*. Cambridge, MA: Harvard University Press, 1990.

---. *Rights, Restitution, and Risk: Essays in Moral Theory*. Ed. William Parent. Cambridge, MA: Harvard University Press, 1986.

---. “Some Ruminations on Rights.” *Rights, Restitution, and Risk: Essays in Moral Theory*. Ed. William Parent. Cambridge, MA: Harvard University Press, 1986. 49-65.

Unger, Peter. *Living High and Letting Die*. New York: Oxford University Press, 1996.

Victor, David G. and Keohane, Robert O. “The Regime Complex for Climate Change.” American Political Science Association Annual Meeting, 20 July 2010. Available at SSRN: http://ssrn.com/abstract=1643813.

Weiner, Rachel. “Obama: ‘We Can't Solve Global Warming Because I F---ing Changed Light Bulbs in My House.’” *The Huffington Post*. 5 November 2008. Web. 15 January 2012.

Wiener, Jonathan. “Incentives and Meta-Architecture.” *Architectures for Agreement: Addressing Global Climate Change in the Post-Kyoto World*. Ed. Joseph Aldy and Robert Stavins. New York: Cambridge University Press, 2007. 67-80.

Woodard, Christopher. *Reasons, Patterns, and Cooperation*. Routledge, 2008.

Zimmermann, Michael. *The Concept of Moral Obligation*. Cambridge: Cambridge University Press, 1996.