

Dimensions of Moral Agency

Edited by

David Boersema

**CAMBRIDGE
SCHOLARS**

P U B L I S H I N G

Dimensions of Moral Agency, Edited by David Boersema

This book first published 2014

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2014 by David Boersema and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-6692-X, ISBN (13): 978-1-4438-6692-7

TABLE OF CONTENTS

Introduction: Dimensions of Moral Agency 1
David Boersema, Pacific University

Part One: Moral Agency

Chaos and Constraints 14
Howard Nye, University of Alberta

The Myth of the Mental (Illness)..... 30
Sarah Vincent, University of Memphis

The Role of Love in Descartes' Meditations..... 39
Chiara Bandini, San Francisco State University

Emotions, Ethics, and Equality: Humanity (Ren) as "True Moral Feeling" 61
Mary I. Bockover, Humboldt State University

Epictetus' Serenity Meditation 97
Morgan Rempel, University of Southern Mississippi

Species Egalitarianism and Respect for Nature 108
Pierson Tse, University of Calgary

Berkeley on Human Freedom and Moral Responsibility..... 117
Brandon Bowen, University of Utah

Part Two: Moral Agency and Society

Aristotle, Virtue, and the Wrong Kind of Reasoning 126
Noell Birondo, Wichita State University

The Green Problem: How Green Consumption Creates Overconsumption..... 140
Colton Markham, Pacific University

CHAOS AND CONSTRAINTS

HOWARD NYE

Introduction

Chaos theory tells us that our world exhibits “sensitive dependence on initial conditions,” or that small changes at any point can lead to dramatically different outcomes. These have become known as “butterfly effects,” after Edward Lorenz’s vivid example of “the flap of a butterfly’s wings in Brazil set[ting] off a tornado in Texas.”¹ Laura Cannon suggested that the pervasiveness of these effects might be morally important when she “consider[ed] the plight of Lorenz’s butterfly,” and “wondered how a butterfly might feel if it had the mental capacities to comprehend Lorenz’s discovery. What sense of responsibility might it feel, knowing that its movement might be the cause of great suffering? Might some butterflies sit paralyzed on the branch in an attempt to avoid being the cause of such harm?”²

Cannon, following Samuel Scheffler (1995), claimed that the far-reaching effects of our economic decisions create trouble for commonsense notions of responsibility. But I believe that Cannon and Scheffler underappreciated the pervasiveness of the problem that butterfly effects pose for views according to which we have stronger moral responsibilities to avoid causing harm ourselves than we have to prevent harms that would occur in the absence of our interference. In this paper I argue that, given the harmful butterfly effects our actions are likely to have, all plausible theories of agent-centered constraints on harming entail that we *must* sit paralyzed—or kill ourselves—in order to avoid causing harm.³ I believe that this result is extremely important for ethical theory. Shelly Kagan, Frank Jackson, and Michael Smith have argued that *absolute* constraints against inflicting serious harms on innocents regardless of the benefits of doing so lead to paralysis under conditions of risk.⁴ But to many of us, the most plausible constraints on harming are not absolute. Even if inflicting harm is significantly harder to justify than failing to prevent harm, it seems that we should still be permitted to painlessly kill one individual to save the rest of the world’s billions from dying the most excruciating deaths imaginable. I shall argue, however, that the likelihood that our actions will

have dramatic butterfly effects undermines all plausible *non-absolutist* understandings of constraints, according to which it is harder but not impossible to justify inflicting serious harms.⁵

1. An Overview of the Argument

My argument begins with the observation that for any way we might sustain our lives, we will have to perform some set of actions, *A*, of which it is reasonable to believe that some members will have butterfly effects. As chaos theory shows, it is not just the consumption choices of westerners in a global economy that can lead to dramatically different outcomes. The most non-intrusive existence that could sustain our lives—say that of a hermit who expends minimal effort tending his garden before dutifully returning to the fetal position—will run a far greater risk of causing a dramatic cascade of events than Lorenz’s butterfly. The minimal life-sustaining actions our hermit must perform, repeated millions of times over the course of his life, will make it virtually certain that his actions will somewhere make things dramatically more different than they would have been had he not performed them. In fact, it is reasonable to believe that his actions over the course of his life will have many such effects.

Because butterfly effects result in such dramatic events as tornadoes, it is reasonable to believe that at least some of these effects of *A* will make some individuals worse off and others better off than they would have been had *A* not been performed. Call the former the “butterfly effect harms of *A*,” or *BH(A)*, and call the latter the “butterfly effect benefits of *A*,” or *BB(A)*. It is, in particular, reasonable to believe that for any acts *A* that could sustain our lives, there will be at least some deaths in *BH(A)* and some life-savings in *BB(A)*. Such are the results of causing and preventing such momentous events as tornadoes. There is, however, no reason to believe that *BB(A)* will tend to be either greater or less than *BH(A)*. As such, it is reasonable to expect that *BB(A)* will on average be about equal to *BH(A)*.

If the beneficial upshots of our conduct were able to justify the harmful upshots, so long as the benefits were equal to or greater than the harms, then the expected butterfly effect benefits in *BB(A)* would exactly justify the expected butterfly effect harms in *BH(A)*, and we could—as it seems we should—treat the unpredictable butterfly effects of our actions as something we can ignore for practical purposes. But agent-centered constraints on harming hold that some harmful upshots of our conduct cannot be justified by equal or somewhat greater benefits. These constraints hold, for instance, that the benefits of saving five individuals

from dying of organ failure cannot justify the harms we would cause to one healthy individual by removing her organs and transplanting them into the five.⁶ But if the harms our conduct inflicts on some cannot be justified by the equal or somewhat greater benefits it brings to others, there is a serious worry that for any course of action *A* that would be needed to sustain our lives, *BH(A)* cannot be justified by *BB(A)*.

One way to prevent butterfly effects from making trouble for agent-centered constraints on harming would be to claim that these constraints make it difficult to justify inflicting only those harms we *intend*. I will argue that this is implausible in Section 2, but for now I simply note that most proponents of agent-centered constraints hold—plausibly—that the benefits of saving five cannot justify certain ways of causing foreseen but unintended harms to one, like running her over if this is the only way to reach the five in time to save them from drowning.⁷ Another way to prevent butterfly effects from making trouble for constraints on harming would be to claim that these constraints make it difficult to justify actions only if they are “proximate” causes of harm, or if we can foresee who the victims of these actions will be. I will argue that this is implausible in Section 3, but for now I simply suggest that most proponents of constraints will hold—plausibly—that somewhat greater benefits cannot justify certain ways of causing harm distally or to unknown victims. These would presumably include saving five by pulling a trigger that sets off an elaborate Rube-Goldberg device that kills one, or sets off a device that fires thousands of rifles at thousands of victims, an unknown one of which is loaded with live ammunition.

If this is right, then any plausible theory of agent-centered constraints on harming will hold that the infliction of certain unintended distal harms cannot be justified by equal or somewhat greater benefits. I will argue in Section 4 that because of the drastic nature of butterfly effects, it is reasonable to believe that some of these difficult-to-justify harms will be among the butterfly effect harms in *BH(A)*. Because each benefit in *BB(A)* is needed to justify a corresponding harm of equal magnitude in *BH(A)*, this will mean that some deaths in *BH(A)* will remain unjustified by life-savings and other reasonably expected benefits in *BB(A)*. So all plausible theories of constraints on harming will hold that for any way we could sustain our lives, the butterfly effects of our actions can be expected to kill others in ways that cannot be justified by the lives they will save. Moreover, I observe in Section 5 that on any plausible theory of constraints on harming, the benefits of saving oneself and *N* individuals are insufficient to justify killing *N* other innocent individuals in ways that are difficult to justify. If this is right, then all otherwise plausible theories

of agent-centered constraints on harming entail that we are morally required either to allow ourselves to waste away or kill ourselves. This, I argue, undermines the plausibility of agent-centered constraints on harming.

2. Constraints Against Only Intended Harm?

Once we know about butterfly effects, we can foresee with reasonable certainty that if we act to sustain our lives, we will make some individuals worse off than they would have been. As I mentioned, one way to deny that this makes trouble for agent-centered constraints on harming is to insist that these constraints make it difficult to justify causing only those harms we intend, while harms we merely foresee can be justified by the equal or slightly greater benefits of causing them. Although a few authors have suggested something like this view of agent-centered constraints,⁸ I think they fail to appreciate how unattractive it really is. If we read “intending harm” literally, then you need to intend a victim to suffer the harm of death only if her dying plays a causal role in what you aim at. But then a prohibition against causing only intended harm would permit you to save five individuals dying of organ failure by harvesting the organs of one healthy individual while she is alive, since her dying as a result of your removing her organs would be a byproduct that plays no role in saving the five.⁹ This would seem to undermine the entire motivation for believing in agent-centered constraints on harming.

If, on the other hand, we interpret “intending harm” a bit less strictly, as something like intending a harmful effect on someone’s body or intending someone to instantiate a property that ends up harming her, then you must intend harm in harvesting the organs of one to save five, although you need not intend harm in driving over one individual trapped in the road to save five others from drowning.¹⁰ But because it seems about as abhorrent to knowingly run over one to save five as to harvest her organs to save them, this understanding of constraints against only causing intended harm, which prohibits the latter but permits the former, also seems to undermine the entire motivation for believing in agent-centered constraints on harming.

To appreciate the absurdity of such a theory of constraints, consider the following cases:

Less Harmful Transplant. You have two ways of saving five individuals from dying of organ failure: (1) remove the organs of one healthy individual and transplant them into the five, or (2) run over four healthy

individuals who are trapped in the road that you would need to drive over to get organ-failure-preventing drugs to the five.

Less Harmful Terrorism. The only way to save five from being killed by a cannon is to drop bombs that will have two effects: (a) destroy the cannon's ammunition and (b) demoralize the terrorists operating the cannon into surrendering by killing some of four innocent bystanders they care about, where both (a) and (b) would be sufficient by itself to save the five. Suppose that if you (1) drop your bombs with an intention of killing one bystander, a mind-reading demigod will shield the other three from your bombs and you will kill only one, but if you (2) drop your bombs with the intention merely of destroying the ammunition, the demigod will leave the three unshielded and you will kill all four.¹¹

A theory according to which there are constraints against inflicting only intended harms (which are strong enough to make it wrong to harvest one individual's organs to save five) would in these cases tell us to take option 2 and save the five by killing four individuals instead of only one.¹² But the mere fact that by killing the four we could avoid having a problematic intention towards the one is a preposterously narcissistic justification for killing three additional individuals. If we are not permitted in these cases to inflict lethal harm on the one with the intention of doing so, then we cannot be permitted to do what we foresee will certainly kill the four either.¹³

3. Constraints Against Only Proximal or Identifiable Harm?

I have thus argued that plausible theories of agent-centered constraints on harming must apply these constraints to some harms that are foreseen but unintended. As I mentioned, one might still deny that constraints on harming apply to harms caused by butterfly effects by holding that these constraints make it more difficult to justify actions only if they are "proximal" causes of harm, or one can foresee who the victims of the actions will be.¹⁴ But it seems quite implausible that how proximally an action causes harm, or whether one can identify its victims, matters *in itself*, quite independently of this indicating a greater risk of causing harm. Consider:

Greater Distal Harm. You have two ways of saving five from drowning: (1) drive straight, which will kill one when you drive over a platform, the

depression of which will crush her, or (2) take an alternate road, which will kill four when you drive over a different platform, the depression of which will crush them. But while the four are located directly under the second platform, the depression of its top half will kill them by setting off an elaborate Rube-Goldberg device with hundreds of causal intermediaries that will eventually cause the downward movement of its bottom half, which will crush them.

More Unknown Victims. You have two ways to prevent the sadistic dictator Pedro from shooting five innocents: (1) shoot one other innocent yourself, or (2) press a button that will select four other innocents from a databank of everyone in the world and send reliable kill-bots after them, which you know with certainty will kill the four.

A theory of constraints that applied only to proximally caused harms or harms with known victims (and was strong enough to make it wrong to drive over one to save five) would in these cases tell us to take option 2 and save the five by killing four individuals instead of only one.¹⁵ But the mere fact that we would kill the four by a longer sequence of causal intermediaries or that we don't know who they will be are ridiculous reasons to kill three additional individuals. If we are not permitted in these cases to proximally cause the death of the identifiable one, we cannot be permitted to distally cause the death of the possibly unidentifiable four either.

4. Plausible Constraints Against Distal Harm

I have thus argued that plausible theories of agent-centered constraints on harming must hold that they apply to causing some merely foreseen distal harms to unknown victims. But these theories are directly vulnerable to the butterfly effect argument. I should emphasize that there are many forms such theories can take. Some will hold that it is difficult to justify harms so long as they counterfactually depend on events that constitute our actions.¹⁶ These will make *all* the harms in *BH(A)* difficult to justify, and consequently unjustified by the roughly equal benefits in *BB(A)*.

Other theories of constraints against merely foreseen, distally caused harm will hold something more like the view that it is difficult to justify a harm if one's actions *produce* it, or there is a continuous, transitive chain of causal events linking one's action to the harm.¹⁷ Since some harms in *BH(A)* will not be produced by *A*, these theories may allow some harms in *BH(A)* to be justified by equally great benefits in *BB(A)*. But it is

reasonable to believe that for any act A that will sustain our lives, *some* lethal harms in $BH(A)$ will be produced by A —this, after all, is the way butterflies' wings kill the victims of the tornadoes they set off. Moreover, since the expected benefits in $BB(A)$ are equal to the expected harms in $BH(A)$, for $BB(A)$ to justify $BH(A)$, each harm in $BH(A)$ must be justified by a corresponding benefit of equal magnitude in $BB(A)$. In particular, each harm in $BH(A)$ that is produced by A must be justified by a corresponding benefit of equal magnitude in $BB(A)$ —since all other benefits in $BB(A)$ are already needed to justify the harms in $BH(A)$ that are *not* produced by A . So, since theories which posit constraints against producing distal harm entail that these harms in $BH(A)$ that are produced by A cannot be justified by equal or somewhat greater benefits, they too will entail that the benefits in $BB(A)$ cannot justify the harms in $BH(A)$.

The same logic applies to more elaborate theories of constraints against causing distal harm. Some, for instance, will have “distributive exemptions” for actions that cause harm to some individuals with the same materials or forces that would have caused harm to others had they not been performed.¹⁸ While some lethal harms in $BH(A)$ may be caused in this way, the pervasive nature of butterfly effects makes it reasonable to believe that there will be other lethal harms in $BH(A)$ that are *not* caused in this way (like deaths caused by tornadoes that wouldn't have formed had one not acted), making it impossible for $BH(A)$ to be entirely justified by $BB(A)$. Other theories may hold that benefits cannot easily justify harms if elaborate explanatory relations accumulate between them. For instance, Frances Kamm proposes that a benefit cannot easily justify causing a harm if “something—[the] means [to the benefit]—brings along with it causes [an effect on a victim that harms her] either directly or by overlapping with the direct cause of [this effect].”¹⁹ But *whatever* one takes the distally caused harms to be that are difficult to justify, the drastic nature of butterfly effects makes it reasonable to believe that some of these will be in $BH(A)$, making it impossible for $BB(A)$ to entirely justify $BH(A)$. For instance, it is reasonable to believe that some of your actions will cause lethal disasters, but that some aspects of these disasters will cause lives to be saved later on—which fits Kamm's criterion for actions, the lethal harms of which cannot be justified by their life-saving benefits.

5. Is this an Argument for Universal Suicide?

I have thus argued that, given the likelihood that our actions will have butterfly effects, all plausible theories of agent-centered constraints on harming entail that for any way we could sustain our lives, it is reasonable

to expect that it will involve at least $N \geq 1$ instances of difficult-to-justify killing, which are not justified by the corresponding N lives that it can be reasonably expected to save.²⁰ Now any theory of agent-centered constraints strong enough to make it wrong to kill one individual in the difficult-to-justify way to save five others will entail that we are not permitted to kill N individuals in the difficult-to-justify way to save $N+1$ others. Moreover, any plausible theory of constraints on harming will apply them to cases where we would be one of the beneficiaries of the harming—and consequently will not permit us to kill N individuals in the difficult-to-justify way to save N other individuals and ourselves. If, for instance, we are morally prohibited from harvesting the organs of one person to transplant into two others, then surely we remain morally prohibited from doing so if we are one of the two who need organs. So any theory of agent-centered constraints on harming that is strong enough to be plausible will not permit us to perform a set of acts that would sustain our lives, even though this would save our lives and the lives of N others, if it would in a difficult-to-justify way kill a different group of N individuals. Since, as I have argued, on any plausible theory of agent-centered constraints, the butterfly effects of our actions make it overwhelmingly likely that any set of acts that would sustain our lives *will* involve at least N instances of killing that are difficult-to-justify and consequently unjustified by the fact that it will save our lives and those of N others, any otherwise plausible theory of agent-centered constraints will not permit us to sustain our lives. It will require us either to let ourselves waste away or kill ourselves.²¹

If I am correct that, because of the likely butterfly effects of our actions, all otherwise plausible theories of agent-centered constraints on harming require us either to allow ourselves to waste away or kill ourselves, what should we conclude? Should we conclude that there are no agent-centered constraints on harming, or should we conclude that we are in fact morally required to let ourselves waste away or kill ourselves? A moral theory should not be dismissed simply because it entails that, given our contingent circumstances, we are all morally required to let ourselves die or kill ourselves.²² But there seems to be something absurd about the view that we must waste away or kill ourselves simply because of the harms that would be wrought by the unpredictable butterfly effects of our actions, when we can reasonably expect these butterfly effects to prevent comparable amounts of harm, and we can live our lives in ways that are predictably beneficial to others and can consequently be expected to do more good than harm on the whole.²³ The implausibility of the idea that the unpredictable harms our lives are likely to cause cannot be justified by

the fact that that our lives are likely to prevent even greater harms seems to illustrate the direct implausibility of the view embodied in otherwise plausible constraints on harming—namely that such factors as whether a harmful upshot of our conduct was produced by our actions or would have occurred in their absence make a significant intrinsic moral difference.²⁴ So I think we should continue to believe that we are *not* morally required to waste away or kill ourselves, and conclude from my argument that there are no agent-centered constraints on harming.

The great irony is that it is the view that killing is worse than letting die, rather than the view that letting die is just as bad as killing, that seems to make morality too demanding.

References

- Aboodi, Ron, Adi Borer, and David Enoch. 2008. "Deontology, Individualism, and Uncertainty: A Reply to Jackson and Smith." *The Journal of Philosophy*, 105: 259-272.
- Bennett, Jonathan. 1981. "Morality and Consequences." In S. McMurrin (ed.), *The Tanner Lectures on Human Values*, Salt Lake City: University of Utah Press, 46-116.
- . 1995. *The Act Itself*. Oxford: Clarendon Press.
- Cannon, Laura. 2003. "The Butterfly Effect and the Virtues of the American Dream." *Journal of Social Philosophy*, 34: 545-555.
- Draper, Kai. 2005. "Rights and the Doctrine of Doing and Allowing." *Philosophy and Public Affairs*, 33: 253-280.
- Donagan, Alan. 1977. *The Theory of Morality*. Chicago: University of Chicago Press.
- Finnis, John, Germain Grisez, and Joseph Boyle. 2001. "'Direct' and 'Indirect': A Reply to Critics of Our Action Theory." *The Thomist*, 65: 1-44.
- Fitzpatrick, William. 2006. "The Intend/Foresee Distinction and the Problem of 'Closeness.'" *Philosophical Studies*, 128: 585-617.
- Foot, Philippa. 1967. "The Problem of Abortion and the Doctrine of Double Effect." *Oxford Review*, 5: 5-15. Reprinted in P. Foot, *Virtues and Vices*, Oxford: Oxford University Press, 19-31.
- . 1984. "Killing and Letting Die." In J.L. Garfield and P. Hennessey (eds.), *Abortion: Moral and Legal Perspectives*, Amherst: University of Massachusetts Press, 177-185.
- Hall, Ned. 2004. "Two Concepts of Causation." In J. Collins, N. Hall, and L. Paul, *Causation and Counterfactuals*, Cambridge, Massachusetts: the MIT Press, 225-276.
- Hanser, Matthew. 1999. "Killing, Letting Die, and Preventing People from Being Saved." *Utilitas*, 11: 277-295.
- Hart, H.L.A. 1967. "Intention and Punishment." *Oxford Review*, 4. Reprinted in H.L.A. Hart, *Punishment and Responsibility*, Oxford: Clarendon Press.
- Hart, H.L.A., and Tony Honoré. 1985. *Causation in the Law*. 2nd ed. Oxford: Clarendon Press.
- Hilborn, Robert. 2001. *Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers*, 2nd edition. Oxford: Oxford University Press.
- Howard-Snyder, Frances. 2011. "Doing vs. Allowing Harm." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2011 Edition, URL = <http://plato.stanford.edu/archives/win2011/entries/doing-allowing/>
- Jackson, Frank and Michael Smith. 2006. "Absolutist Moral Theories and Uncertainty." *The Journal of Philosophy*, 103: 267-283.
- Kagan, Shelly. 1989. *The Limits of Morality*. Oxford: Oxford University Press.
- Kamm, Frances. 2007. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. New York: Oxford University Press.
- Lorenz, Edward. 1972. "Predictability; Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas?" *American Association for the Advancement of Science*, 139th meeting.
- Mack, Eric. 1985. "Three Ways to Kill Innocent Bystanders: Some Conundrums Concerning the Morality of War." *Social Philosophy and Policy*, 3: 1-26.
- McMahan, Jeff. 1993. "Killing, Letting Die, and Withdrawing Aid." *Ethics*, 103: 250-279.
- . 2009. "Intention, Permissibility, Terrorism, and War." *Philosophical Perspectives*, 23: 345-372.
- Mikhail, John. 2007. "Universal Moral Grammar: Theory, Evidence, and the Future." *Trends in Cognitive Science*, 11: 143-152.
- . 2011. *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. New York: Cambridge University Press.
- Moore, Michael. 2009. *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*. Oxford: Oxford University Press.
- Norcross, Alastair. 2003. "Killing and Letting Die." In R.G. Frey and C.H. Wellman (eds.), *A Companion to Applied Ethics*, Oxford and Boston: Basil Blackwell, 451-463.

- Pruss, Alexander. 2013. "The Accomplishment of Plans: a New Version of the Principle of Double Effect." *Philosophical Studies*, 165: 49-69.
- Quinn, Warren. 1989a. "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing." *Philosophical Review*, 98: 287-312.
- . 1989b. "Actions, Intentions, and Consequences: The Doctrine of Double Effect." *Philosophy and Public Affairs*, 18: 334-351.
- Ross, W.D. 1930. *The Right and the Good*. Oxford: Clarendon Press.
- Rickless, Samuel. 2011. "The Moral Status of Enabling Harm." *Pacific Philosophical Quarterly*, 92: 66-86.
- Russell, Bruce. 1977. "On the Relative Strictness of Negative and Positive Duties." *American Philosophical Quarterly*, 14: 87-97.
- Scanlon, T.M. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge: Harvard University Press.
- Scheffler, Samuel. 1995. "Individual Responsibility in a Global Age." *Social Philosophy and Policy*, 12: 219-236.
- Shaw, Joseph. 2006. "Intentions and Trolleys." *Philosophical Quarterly*, 56: 63-83.
- Thomson, Judith. 1976. "Killing, Letting Die, and the Trolley Problem." *The Monist*, 59: 204-217.
- . 1985. "The Trolley Problem." *Yale Law Journal*, 94: 1395-1415.
- . 2008. "Turning the Trolley." *Philosophy and Public Affairs*, 36: 359-374.
- Vihvelin, Kadri and Terrance Tomkow. 2005. "The Dif." *The Journal of Philosophy*, 102: 183-205.

Notes

¹ The example is from Lorenz 1972. For a systematic introduction to chaos theory, see Hilborn 2001.

² Cannon 2003, 145.

³ The basic idea of this argument was suggested to me by Allan Gibbard. I am also grateful to John Ku for many extremely helpful discussions of it. But any problems with the argument as I develop it here should be attributed entirely to me.

⁴ That is, under conditions where we can reasonably assign probabilities to the various possible outcomes of our conduct, but we cannot know these outcomes with certainty. See Kagan 1989, 87-91 and Jackson and Smith 2006.

⁵ Besides applying only to absolute constraints on harming, another limitation of Kagan, Jackson, and Smith's arguments is that they allow defenders of constraints to avoid the conclusion that constraints require paralysis by applying constraints only in cases where the probability of causing harm exceeds a certain threshold (see Kagan 1989, 90n5; Jackson and Smith 275-278; and Aboodi et al. 2008. Jackson and Smith argue that this response is problematic, to which Aboodi et al.

respond). But because keeping ourselves alive seems overwhelmingly likely to cause harmful butterfly effects somewhere down the line, the probability of causing harms, to which my argument appeals, seems to surpass any thresholds for the application of constraints to which their defenders might appeal.

⁶ See Foot 1967.

⁷ See Foot (1967, 1984), as well as (among others) Thomson (1976, 2008), Donagan (1977), Hanser (1999), Vihvelin and Tomkow (2005), Draper (2005), Kamm (2007), and Rickless (2011). Even authors like Quinn (1989b) and McMahan (2009), who believe that there are *stronger* constraints against intended harming than foreseen harming, still accept that there are constraints against certain ways of inflicting foreseen but unintended harms, which entail that the infliction of these harms cannot be justified simply by the equal or somewhat greater benefits of inflicting them (see Quinn 1989a and McMahan 1993).

⁸ See Mack (1985); Aboodi et al. (2008, 267); and Pruss (2013, 50).

⁹ To see this, observe that in such a case if the one were somehow to survive the removal of her organs, you would have cause to rejoice. This point has been noted in the literature at least since Hart 1967, and was well discussed by Foot (1967, 21-22), Russell (1977, 95-96), and Bennett (1981, 110-111; 1995, 210-213). For a version of the Doctrine of Double Effect that seems to embrace this conclusion, see Finnis, Grisez, and Boyle 2001. (Note, moreover, that it would be implausible to think that there is a serious constraint against causing intended "harms" like mere organ removal, independent of their lethal consequences. If the only way to save five from dying was to remove the organs of one, but the procedure was guaranteed to be painless, intrude in no way into the life of the one, and result in the one having better organs that would cause her to live longer than she would have with the organs we removed, I think that there could be no serious moral objection to removing her organs to save the five. At the very least, the objection could be nowhere near as strong as the one there seems to be to lethally removing her organs).

¹⁰ This is because any effect on the individual's body plays no causal role in your saving the five—you would have saved them just as well if she failed to instantiate any properties at all. For this kind of interpretation of the constraint against causing intended harm see Mack 1985; Quinn 1989b; Shaw 2006, 69-71; and Mikhail 2007, 145; 2011, 133-136, 148-152. Quinn at least was well aware that such constraints would not by themselves prohibit running over one to save five, but in addition to them he defended constraints against causing certain unintended harms (1989a), which would not permit us to run over one to save five.

¹¹ One might worry that the intentions with which you drop your bombs are not within your voluntary control, so (1) and (2) are not legitimately distinct options in Less Harmful Terrorism (see e.g. Ross 1930, 4-6; Bennett 1981, 96-98; 1995, 194-196; and Scanlon 2008, chapter 2). This actually constitutes an important objection to the view that there are constraints against inflicting only intended harms, but it is not the objection I am presenting here. We can, for my purposes, assume that your intentions are under your voluntary control in Less Harmful Terrorism—say,

because you have pills that you can take that you know will induce the relevant intentions.

¹² Such a theory would forbid us from taking option 1 (since it would involve inflicting intended harm) and permit us to take option 2 (since the harm it inflicts is not intended and it does more good than harm). It would, moreover, treat option 2 as more strongly favored by moral reasons than simply not saving the five in the same way that we would commonsensically treat saving five as more strongly favored by moral reasons than saving four instead.

¹³ Fitzpatrick (2006) might try saying that in these cases, the events of your car moving where it does or your bombs exploding where they do—which you intend—“constitute” the death or lethal injury of the four, so the choice is actually between more or less intended harm rather than between more foreseen harm or less intended harm. It is, however, preposterous to say that the event of the four dying is identical to, or constituted by, the event of your car moving where it does or your bombs exploding where they do. The former could take place several minutes or hours after the latter and at completely different locations (if the four were rushed to hospital). Moreover, it does not matter whether an intended event “constitutes lethal injury” so long as it is equally sure to kill its victims. Suppose that instead of driving over the four or blowing up the cannon’s ammo, you could hire a giant to get the medicine to the five or destroy the cannon’s ammo by telling him the hiding place of the four, who he would very much like to eat. Obviously, you need not here intend any injury to the four, but if telling the giant about their hiding place is just as sure to kill them as driving over them or exploding bombs in their vicinity, it is no easier to justify.

¹⁴ For suggestions along these lines, see the ordinary causal claims described in Hart and Honore (1985), Moore (2009), and Pruss (2013, 61-63).

¹⁵ Such theories would forbid us from taking option 1 (since it would involve inflicting proximal harm, or harm to identifiable victims) and permit us to take option 2 (since the harm it inflicts is not proximal or to known victims, and it does more good than harm). These theories would also treat option 2 as more strongly favored by moral reasons than simply not saving the five in the same way that we would commonsensically treat saving five as more strongly favored by moral reasons than saving four instead.

¹⁶ In contrast to harms—like those that result from our simply failing to save others—that counterfactually depend only on those features of our conduct that constitute omissions, or non-occurrences of events that would have constituted actions on our parts. Something like this is the most natural way of understanding Donagan’s (1977) view, Quinn’s (1989a, 294) initial suggestion for clarifying the DDA (before he incorporated elements concerning wrongful omissions), and Vihvelin and Tomkow’s (2005) view.

¹⁷ On the distinction between production and other kinds of causation, see Hall 2004. Theories of agent-centered constraints that apply them primarily to something like harms produced by one’s actions include those of Hanser (1999), Draper (2005), and Rickless (2011) (although Rickless (79-81) seems to describe some events—like removing a trap-door that is preventing someone from being

hanged (but *not* removing a laser-beam that is preventing someone from being hanged)—as “initiating causal sequences” that lead to harm even when they are not producers of harm; the basis of his categorization is unclear and seems to have been influenced by salience).

¹⁸ See Thomson 1976, 1985—although Thomson (2008) has since rejected this idea.

¹⁹ Kamm 2007, 149. I do not mean to suggest that this has even the slightest shred of plausibility as a theory of agent-centered constraints. I mention it only to illustrate the point that whatever one takes to be the principled distinction between the distal harms that are (as opposed to the distal harms that are not) difficult to justify, the butterfly effect argument will still apply.

²⁰ Note that N is not necessarily the number of lethal harms in $BH(A)$, since as I discussed a theory of constraints might hold that some of these harms are not of the kind that cannot be justified by equally important life-savings in $BB(A)$. If so, then the expected number of lethal harms in $BH(A)$ is $N+M$, where M is the expected number of easy-to-justify deaths A causes, which are justified by a corresponding number M of expected life savings in $BB(A)$. Because the M lives saved exactly justify the M deaths caused, I omit them from discussion below for the sake of simplicity.

²¹ Whether the theory of constraints requires us to stay as still as we can until we waste away, requires us to actively kill ourselves, or permits us to do either, depends upon exactly which harmful upshots of our bodily processes it counts as difficult-to-justify. For instance, a theory like Donagan’s (1977, 42-43) might count only harms that counterfactually depend on actions produced in the right way by our deliberate decisions. Such a theory would seem to require us to stay as still as we can until we waste away, since any deliberate bodily movements (including those involved in actively hastening our deaths) would risk causing difficult-to-justify harms via butterfly effects, while staying as still as we can is guaranteed not to do so. But as Alastair Norcross (2003, 455-456) observes, it seems implausible to treat harms as easier to justify if they are caused by our own (easily controllable) reflexes, and as Frances Howard-Snyder (2011, §6) observes, it can seem implausible to treat certain harms as easier to justify even if they are caused simply by the position of our own (easily movable) bodies.

On a view of constraints that counts harms caused by our automatic bodily processes or even the positions of our own bodies as difficult-to-justify, simply staying as still as we can until we waste away would not be guaranteed to avoid all risks of causing difficult-to-justify harms as a result of the butterfly effects of our conduct. Since the expected harms that will result from the butterfly effects of our bodily processes will be much less if we stay as still as we can until we waste away than if we actually sustain our lives, these views of constraints will hold that sustaining our lives is more deeply morally wrong than staying as still as we can until we waste away—in the same way that killing five individuals, each of whom has a different organ you need to survive, is more deeply wrong than killing only one of the five. But the expected harms that will result from the butterfly effects of our actively killing ourselves may be even less than the expected harms that will

result from the butterfly effects of our bodily processes if we stay as still as we can until we waste away. If (as seems plausible) this is so, then theories of constraints that count harms caused by our bodily processes as difficult to justify will presumably require us to actively kill ourselves so as to minimize the amount of harm we can be expected to inflict on others. (Just because running some risk of inflicting harm is unavoidable or permissible does not mean that we cannot be required to minimize the harm we can be expected to inflict by sacrificing our own lives. For instance, as Philippa Foot (1967) and Judith Thomson (2008) would presumably argue, if the only way to prevent ourselves from killing five individuals with the trolley we are driving is to ram it into a wall that will certainly kill ourselves and run some small chance of killing one other innocent, we would be morally required to do so. Similarly, if the only way to save ten billion from dying is to either (1) kill five individuals or (2) kill ourselves and one other individual, it seems plausible that we would be required to take option (2) and kill ourselves and the one other.)

I am grateful to Octavian Ion, Bradley Strawser, Julie Tannenbaum, and Justin Weinberg for very helpful discussion of this issue.

²² For instance, our contingent circumstances might be so bad that the only way for us to save billions of others from even more painful deaths was to kill ourselves, in which case a moral theory *should* require us to kill ourselves.

²³ For simplicity I omitted the benefits we can confer on others by acting in predictably more beneficial ways in my argument that, because of the likely butterfly effects of our actions, any plausible theory of agent-centered constraints will require us to let ourselves waste away or kill ourselves. To see why this conclusion still follows, note that theories of constraints according to which killing is only marginally harder to justify than letting die do not seem plausible. If killing is harder to justify, it is *much* harder—although presumably not impossible—to justify. As such, plausible theories of constraints will permit us to remain alive only if the number of others we save by doing so is much greater than the number we can be expected to kill in difficult-to-justify ways through the butterfly effects of our actions (*after* removing the life-savings needed to justify any *easy*-to-justify harms that we inflict over the course of our lives). The most beneficial lives most of us could lead—say working optimally hard in optimally lucrative careers and giving our earnings to organizations like Oxfam—might be reasonably predicted to save some hundreds or thousands of individuals. But it would seem reasonable to expect that, given the sensitive dependence of future events on initial conditions, such lives will over the course of all future history cause unpredictable butterfly effects that will kill much greater numbers of individuals in what plausible theories of constraints will regard as difficult-to-justify ways (while of course saving similar numbers of individuals too). As such, it does not seem that the number of individuals one could save by living even the most beneficial life possible would (after removing the life-savings needed to justify any *easy*-to-justify harming) be much greater than the number of individuals one would kill in ways that plausible theories of constraints will regard as difficult-to-justify.

²⁴ That is to say, the implausible requirement to forgo living lives that would on balance do more good than harm simply because of the harms caused by the butterfly effects of our actions seems to be more than a mere counter-intuitive consequence of otherwise plausible agent-centered constraints on harming. The implausibility of this requirement seems to illustrate a generally implausible aspect of these constraints, considered in themselves and independent of what else they entail—namely that they place a great deal of weight on such factors as whether a harmful upshot of our conduct was produced by our actions or would have occurred in their absence, which seem on reflection to be rather arbitrary or devoid of significant intrinsic moral importance.