

UCLA

UCLA Electronic Theses and Dissertations

Title

The Justification of Antagonistic Response to Wrongdoing

Permalink

<https://escholarship.org/uc/item/8sd4z245>

Author

Goldman, David Michael

Publication Date

2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

The Justification of
Antagonistic Response to Wrongdoing

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

by

David Michael Goldman

2012

© Copyright by
David Michael Goldman
2012

ABSTRACT OF THE DISSERTATION

The Justification of Antagonistic Response to Wrongdoing

by

David Michael Goldman

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2012

Professor Pamela Hieronymi, Chair

There is a strong Western tradition of opposing angry, hostile, or antagonistic reactions to wrongdoing. In the twentieth century, leaders like Mahatma Gandhi and Dr. Martin Luther King, Jr. counseled responding to wrongdoing with forgiveness and love rather than anger, hate, or vindictiveness.

This ideal has taken on an exalted status in Western culture. Gandhi and King are widely regarded as moral saints. And yet sometimes antagonism seems deeply appropriate. Consider a very serious wrong: suppose, for instance, that a driver viciously and deliberately runs down my small child on the sidewalk. I *ought* to be resentful and angry at someone who does such a thing. It is troubling if I am not.

In light of such worries, some philosophers have defended antagonistic response to wrongdoing. In the dissertation I explore these defenses.

Chapter 1 identifies my target. These responses, which I call *antagonistic*, involve taking the wrongdoer's suffering to be intrinsically, non-instrumentally valuable. Chapters 2 and 3 explore attempts to establish that such response is required by important interpersonal

attachments. In chapter 2, I discuss P. F. Strawson's well-known defense of antagonistic resentment in "Freedom and Resentment;" in chapter 3, I discuss attempts to tie antagonistic response to attachment to the victim of wrongdoing. In both cases, I argue, there are non-antagonistic alternative reactions—centrally, a form of moral sadness or disappointment—that can serve the same interpersonal roles. Antagonistic response is not, in fact, required by these important interpersonal attachments.

Chapters 4 and 5 discuss restorative and nullification defenses of response to wrongdoing. Restoration involves restoring some valuable state of affairs that has been destroyed by wrongdoing; annulment involves retroactively altering the wrongdoing itself in some way. Retroactively altering the past may sound farfetched; I argue that we can make sense of it, and appeal to such annulment to defend some specific ways of responding to wrongdoing. In the end, though, these responses do not involve the antagonism that is condemned by King and Gandhi. I conclude with a qualified endorsement of the King-Gandhi ideal: none of the attempts to justify antagonistic response that I consider succeed.

The dissertation of David Michael Goldman is approved.

David Howard Dolinko

Barbara Herman

Gavin Lawrence

Calvin G. Normore

Pamela Hieronymi, Committee Chair

University of California, Los Angeles

2012

For Tracy

TABLE OF CONTENTS

Abstract of the Dissertation	ii
Acknowledgements.....	vii
Vita.....	x
Chapter 1: Antagonism	1
Chapter 2: Modification of the Reactive Attitudes.....	13
Chapter 3: Anger and the Victim.....	39
Chapter 4: Restoration of Relationships	68
Chapter 5: Nullification, not Restoration.....	90
Conclusion	137
Bibliography	139

ACKNOWLEDGEMENTS

I owe thanks first and foremost to my advisor, Pamela Hieronymi, whose guidance, criticism, and encouragement has made this a far better piece of philosophy than it otherwise would have been, and made me into a far better philosopher than I otherwise would have been.

I first started thinking about the themes of this dissertation six years ago, while reading Strawson's "Freedom and Resentment" for a course taught by Pam. Since then, she has patiently and supportively directed me through every major project I have undertaken in graduate school. She has every quality I could have looked for in an advisor: she is demanding and critical when I need it; she identifies what's valuable in my writing long before I can see clearly what I'm doing; and—last but certainly not least—doing philosophy with her is a genuine joy.

Barbara Herman, Gavin Lawrence, and Calvin Normore have all helped me in ways too numerous and wide-ranging to do justice to here. Even when the project was undergoing tectonic shifts in direction (the word 'talionic' does not appear once in what follows!), they have helped me keep my bearings. For that I am grateful.

I have been particularly lucky in my choice of an outside member of my committee, David Dolinko. I did not know when I began work on this dissertation that I would wind up working so directly on the justification of criminal punishment; nor did I know that David is among the most incisive philosophers working on criminal punishment today. I have learned a tremendous amount from his work, from his feedback on my work, and from my conversations with him; and I am painfully aware of how far short my arguments must still fall in his eyes.

Chapter 2 has been circulated and commented on especially widely; I owe special thanks to Pamela, my committee members, Tiffany Teeman Blase, Zac Cogley, Daniel Haggerty, Andrew Khoury, James Rocha, and two anonymous referees for their detailed and insightful

comments on it during its development. Thanks also to audiences at the Central States Philosophical Association, the Albritton Society, the Rocky Mountain Ethics Congress, the Eastern APA, the UCLA Ethics Writing Workshop, and the 2011 Job Market Seminar for wide-ranging, productive discussions of this material. Large portions of chapter 3 are informed (no doubt imperfectly) by those discussions.

Too many people have assisted me with comments, feedback, and conversation to recount them all. In addition to those listed above, I want to thank Jorah Dannenberg, Janelle DeWitt, Sonny Elizondo, Jeff Helmreich, Adam Masters, Ari Meadow, Herbert Morris, Jesse Summers, Steve White, and Robert Paul Wolff. This list omits many. I blame my atrocious memory for that, and thank everyone who has been omitted.

I am grateful for the support of the staff at the UCLA Philosophy Department. They have made every complex, bureaucratic aspect of graduate study perplexingly easy.

My students at UCLA and especially at UCLA Extension have done more for my philosophical development than they know. Is a teacher supposed to learn so much from his students? I always thought it was the other way around.

Thanks to my parents, Tom and Diane, and my sister, Rachel, for their support and encouragement. Thanks also to my many friends for their support throughout graduate school. I owe particular thanks to Robert Richter, who directed me to *On Killing* and thus gave my thinking about the important and difficult issues discussed in section V of chapter 3 concrete shape. More generally, his conversation and friendship have contributed to this dissertation in ways great and small.

My wife Tracy is my support, my inspiration, and the love of my life. My children, Miles and June, bring me more joy than I ever expected. Being able to come home to them each night gave me the strength to complete this project.

Thanks finally to Bob Turansky, who introduced me to philosophy 14 years ago and remains one of the greatest teachers I have ever known.

VITA

Education

University of California, Los Angeles	M.A., Philosophy, 2007
University of California, Berkeley	B. A. With High Distinction, Philosophy, 2003

Employment

UCLA Extension	Instructor, 1/2011-6/2012
UCLA Philosophy Department	Instructor, Summer 2011 Teaching Fellow, 2006-2011

Presentations

- Eastern APA, December 2011 (refereed), "Modification of the Reactive Attitudes"
- Rocky Mountain Ethics Congress IV, August 2011 (refereed), "Modification of the Reactive Attitudes"
- Albritton Society, May 2011, "Modification of the Reactive Attitudes"
- UCLA Ethics Workshop, October 2010, "Modification of the Reactive Attitudes"
- Central States Philosophical Association, September 2010 (refereed), "Modification of the Reactive Attitudes"
- UCLA Ethics Workshop, February 2010, "Moral Education"
- UCLA Ethics Workshop, February 2009, "Relationships Gone Wrong"
- UCLA Ethics Workshop, April 2008, "Retribution and Ill Will"

CHAPTER 1: ANTAGONISM

There is a strong Western tradition of opposing angry, hostile, or antagonistic reactions to wrongdoing. In the twentieth century, leaders like Mahatma Gandhi and Dr. Martin Luther King, Jr. counseled responding to wrongdoing with forgiveness and love rather than anger, hate, or vindictiveness. “If [peace] is to be achieved,” King said in his Nobel Peace Prize Acceptance Speech in 1964, “man must evolve for all human conflict a method which rejects revenge, aggression, and retaliation. The foundation of such a method is love.”¹

Call what is being suggested here the Gandhi-King ideal. On this ideal, it would be better to eliminate a broad swath of angry, hostile, or antagonistic responses to wrongdoing if we could. Even if we cannot eliminate such reactions altogether, one might think, we should condemn and disavow them wherever they do appear.

This ideal, and its proponents, have taken on an exalted status in Western culture. Gandhi and King are widely regarded as moral saints. And their rejection of antagonistic response is commonly viewed as a sort of moral ideal—perhaps not something that ordinary people like us can hope to accomplish, but something admirable and praiseworthy.

And yet loving, conciliatory, or otherwise non-antagonistic responses can sometimes seem deeply inadequate. Consider a very serious wrong: suppose, for instance, that a driver viciously and deliberately runs down my small child on the sidewalk. A loving response to such an action seems problematic. It is worrisome if I do not experience anger or hostility toward this person. That is not just because we doubt that I would be unable to control my antagonistic or

¹ Note that King delivered two different speeches for his Nobel Peace Prize, a shorter acceptance speech and a longer Nobel Lecture. I cite both in what follows.

retaliatory impulses in such a situation. Those impulses also seem deeply appropriate. I *ought* to be angry at a person who runs down my child.

This dissertation explores the justification of antagonistic response to wrongdoing. In it I examine several strategies for developing the intuition that antagonistic response is sometimes deeply appropriate. I want to see if antagonistic responses can be defended.

By stipulation, I use the term ‘antagonistic’ to label the class of responses that are rejected by the Gandhi-King ideal. Although I will give an account of these responses below, it is important to recognize that they vary widely. That is, I think, an accurate reflection of the wide-ranging and ambitious nature of the Gandhi-King ideal. King is particularly fond of open-ended lists: in the Nobel Peace Prize Acceptance Speech quoted above, he decries “revenge, aggression, and retaliation,” as well as “resorting to violence and oppression” to fight violent oppression, while in the *Letter from Birmingham Jail* he singles out “bitterness and hatred” (844).

In what follows, I discuss several attempts to justify different members of this broad class. None of these justifications are applicable to the entire class; instead, they apply to one or another specific form of antagonistic response. I want to know if *any* members of the broad class of antagonistic responses can be justified.

Although I take an inclusive view of antagonistic response, it is possible to be too inclusive. One might think that any treatment that makes someone worse off, or creates the subjective experience of being made worse off (I will use the term “suffering” to capture these broad categories) is antagonistic. But it is not.

Reflecting on the actual activities of Gandhi and King helps here. (I should say, though, that I am not attempting to capture the precise views of Gandhi, King, or others involved in these movements. Rather, I am trying to capture a familiar ideal within the Western tradition, one that Gandhi and King grappled with and put into action.) Gandhi was instrumental in the destruction of a significant part of the British Empire. This surely made the British worse off in a variety of ways, but Gandhi did not view his actions as immoral for that reason; instead, he worked for the end of British colonial rule persistently and, in the end, effectively. George Orwell captured this when he described Gandhi's strategy as "a sort of non-violent warfare, a way of defeating the enemy without hurting him and without feeling or arousing hatred" (467).

Defeating another is compatible with the Gandhi-King ideal; *hurting* him, or feeling or arousing *hatred*, is not. But what does the difference between these consist in?

The defeat that is compatible with the Gandhi-King ideal might involve suffering in several ways. First, Gandhi and his followers aimed to frustrate the British in their pursuit of the unjust end of oppressing and exploiting India. It seems likely that the British (or, at least, those British citizens who endorsed these ends) would experience this frustration as being made worse off. Second, accomplishing this aim had the side-effect of making the British worse off in a number of ways: they lost a source of tax income, easy access to cheap labor, access to inexpensive tea, and the like.

Why doesn't the Gandhi-King ideal condemn bringing about these forms of suffering?

One might think that this just turns on the amount of suffering inflicted: losing income usually involves much less suffering than being the object of physical violence. But even if this were always true, which it is not, it would not account for another aspect of the Gandhi-King ideal, one that Orwell recognizes: the rejection of hatred. Hating someone often leads to

inflicting suffering on her; but it can be concealed and controlled, so that it causes no suffering for its object. Even this contained hatred is rejected by the ideal.

Here is a proposal: the antagonistic reactions to which Gandhi and King object involve taking the suffering of another to be intrinsically valuable—valuable and worth promoting in itself, apart from its connections to other valuable ends or purposes. Hatred seems to involve taking this sort of attitude towards someone's suffering. By contrast, Gandhi need not have viewed the frustration the British experienced, or their financial losses, as itself valuable: he might have viewed those as disvalued but unavoidable features of the attempt to end oppression and secure Indian autonomy.²

One might, antagonistically, take someone's suffering to be intrinsically valuable; or, nonantagonistically, one might take it to be a disvalued but unavoidable feature of some valuable action or end. There is a third possibility as well: one might take someone's suffering to be *instrumentally* valuable. One might view the suffering as a valuable means to some further, distinct end. It seems clear that the Gandhi-King ideal rejects taking someone's suffering to be intrinsically valuable. Should it also reject taking someone's suffering to be instrumentally valuable?

If we think about Gandhi's and King's resolute opposition to violence, even violence that could be instrumentally valuable in this way, it might seem that the answer is yes: that the Gandhi-King ideal should condemn *any* way of valuing someone's suffering.

I think that is a mistake. Consider first the fact that nonviolent resistance often involved treating suffering as instrumentally valuable in just this sort of way. That becomes clear in

² This echoes the doctrine of double effect: the suffering of the British is a foreseen but unintended consequence of Gandhi's actions. Notice, however, that I am not claiming that bringing about foreseen but unintended suffering is permissible. Rather, I claim that doing so is not antagonistic.

Martin Luther King's *Letter from Birmingham Jail*. In the letter, a reply to published criticism of his actions in Birmingham, King writes:

Nonviolent direct action seeks to create such a crisis and foster such a tension that a community which has constantly refused to negotiate is forced to confront the issue....My citing the creation of tension as part of the work of the nonviolent-resister may sound rather shocking. But I must confess that I am not afraid of the word "tension." I have earnestly opposed violent tension, but there is a type of constructive, nonviolent tension which is necessary for growth....The purpose of our direct action program is to create a situation so crisis-packed that it will inevitably open the door to negotiation. (837-8)

The protests organized and led by King create tension and crisis. That at least involves inflicting economic harm, and provoking social and interpersonal conflict. These are all forms of suffering inflicted on the communities targeted by nonviolent direct action. King and his fellow-protesters go beyond just accepting that some suffering will be a side-effect of their actions; they use this suffering as a means to the end of forcing negotiation and growth.

King acknowledges that this may sound shocking. But I think he is correct to view his strategy in Birmingham as compatible with the Gandhi-King ideal. There is an important divide between being willing to use suffering to bring about some good end, and viewing suffering as itself valuable. The latter will be my focus in this dissertation. (Careful readers will have noted that King contrasts his strategy with *violence* rather than hatred or other ways of viewing suffering as intrinsically valuable. I will have more to say about violence shortly.)

That is not to say that it would be easy or uninteresting to justify inflicting suffering by appeal to its instrumental value. It is obvious that suffering can be instrumentally valuable—bombing Tripoli in 2011 hindered Moammar Qaddafi's suppression of a resistance movement and helped to bring about the end of his dictatorial regime, for instance—but that does not suffice to show that we would be justified in inflicting such suffering. Justifying inflicting suffering by appeal to its instrumental value takes more than simply identifying that instrumental

value. Doing so in any individual case would be an accomplishment. It is just not the accomplishment that I am after.

The justificatory challenge facing viewing suffering as intrinsically valuable is different. Just viewing someone's suffering as intrinsically valuable—independent of whether one acts on that or not—seems morally problematic. Think about the ways one might take such a stance. One might enjoy someone's suffering, without an eye to its being useful for any further end; or one might deliberately aim to make someone suffer, without having any further goal in mind. Regardless of whether or how one acts, there is something morally repellant about taking this sort of stance towards someone's suffering.

Valuing suffering for itself is, I think, the central focus of the Gandhi-King ideal's condemnation. But what about violence? Gandhi and King were quite clear about their categorical condemnation of overt violence. And yet it does not seem like overt violence must be antagonistic in the way I have set out. What if Indian freedom fighters had engaged in overt violence, while viewing the suffering of British soldiers that they brought about as a merely instrumentally valuable means to some end? That would not involve viewing the suffering of the British soldiers as intrinsically valuable, and so it does not seem to be antagonistic, on my account. My account of antagonism—of the responses that are rejected by the Gandhi-King ideal—seems to leave out the violent responses that were most clearly and resoundingly rejected by Gandhi and King.

One might think of the rejection of violence as a strategic move: given the circumstances in which Gandhi and King operated, violence might have been less effective, or introduced

greater uncertainty about success, than other strategies.³ White American public opinion in the mid twentieth century might not have responded with sympathy to violently assertive African-Americans, for instance. I don't deny that Gandhi and King were both shrewd, strategic leaders. But their rejection of violence did have a principled ground, which should not be overlooked or dismissed in favor of purely strategic explanations.

There are two possible explanations for the principled rejection of violence. First, consider a connection that King makes in his Nobel Lecture. Violence, he writes, "is immoral because it seeks to humiliate the opponent rather than win his understanding: it seeks to annihilate rather than convert. Violence is immoral because it thrives on hatred rather than love." King seems to think that violently inflicting suffering cannot be a way of promoting the end of increased understanding. Instead he connects violence with annihilation and hatred. This suggests that violence and antagonism are quite closely connected. Perhaps there is some necessary connection between violence and antagonism, such that it is not possible to engage in violence without coming to view the suffering of one's opponent as intrinsically valuable.

Alternatively, the Gandhi-King ideal might separately condemn violence, without assimilating it to antagonism. In the same passage in which he connects violence and hatred, King also argues that "...in spite of temporary victories, violence never brings permanent peace. It solves no social problem: it merely creates new and more complicated ones. Violence is impractical because it is a descending spiral ending in destruction for all." This attacks the instrumental, strategic value of violence, without relying (at least, not explicitly) on a connection between violence and antagonism.

³ Cf. Jean Hampton's remark in *Forgiveness and Mercy* (which I discuss in greater detail below): "Still, one wonders whether King and others chose their non-violent method of fighting...because the superior, punitive method was denied them" (n. 24, p. 128).

This discussion of violence raises a number of disparate questions: what those involved in the nonviolent movement actually thought about violence; whether violence is instrumentally useful at all; whether an instrumental justification of violence is successful on its own terms. These are all interesting and important questions, but I set them aside to focus squarely on antagonism: on valuing suffering for its own sake. This discussion also raises the question whether (and how) antagonism and violence are connected. That does bear on my subject; in chapter 3, I examine the connections between violence and antagonism in greater depth.

The antagonism I focus on, which involves viewing the wrongdoer's suffering as intrinsically valuable or valuable for its own sake, seems particularly morally problematic; but I want to reiterate that it is not rare, or without intuitive appeal. Taking an antagonistic stance towards someone need not involve adopting the end of making her suffer; it may manifest in other ways, for instance as a disposition to be pleased at its object's suffering. Keeping this in mind, we can see that a wide range of reactions to wrongdoing are antagonistic. Importantly for the next two chapters, *resentful anger* is antagonistic: it involves being motivated to inflict suffering on a wrongdoer, or being disposed to take pleasure in her suffering, without attention to any further good to which the suffering might contribute.⁴ The retaliatory impulse to strike back at someone who has wronged one, or done one some injury, often seems to involve viewing that retaliatory injury as good in itself, and so is often antagonistic in the sense I am after. In the

⁴ There might be some basic or minimal form of anger which does not involve any of these antagonistic elements; perhaps it just involves a negative appraisal of an object, coupled with anger's characteristic affect. In any case, in chapters 2 and 3, below, I focus on angry resentment in particular, which does involve antagonistic elements. I also note that even the affect characteristic of anger can seem antagonistic: it seems to involve some urge to strike at the object of the anger, or some pleasure at seeing that object made worse off.

aftermath of serious wrongdoing like a vicious, deliberate attack on my child, hateful, angry, or retaliatory responses like these do seem deeply appropriate.

Antagonism can also seem deeply appropriate in the context of criminal punishment. Many retributive theories of punishment make punishment out to be antagonistic in some way. One strain of retributive thinking about criminal punishment holds that it is a primitive, unanalyzable, intrinsic good for wrongdoers to suffer.⁵ If an individual, group, or institution punishes criminals, and takes itself to be justified in doing so by the intrinsic goodness of the criminals' suffering, then that punishment is antagonistic: it involves taking that suffering to be intrinsically good.⁶ Other retributive accounts of punishment also seem to involve antagonism, albeit less directly and flat-footedly; I draw on some such accounts in chapters 4 and 5 below.

A caveat about the way that I will approach attempts to justify antagonistic response. I do not examine the grounds for Gandhi's and King's suspicions about antagonism. My hunt for a justification for antagonism proceeds as if the burden of proof rests on those who would defend antagonism: as if antagonistic response to wrongdoing is plainly morally problematic, unless some compelling justification for it is forthcoming.

That in turn suggests that I do not think the problems with antagonistic hatred, anger, and the like need further examination. But they do. The rejection of antagonism is a serious, substantive claim that is far from fully fleshed out. Gandhi and King both suggest that the

⁵ See, e.g., Michael Moore, "The Moral Worth of Retribution."

⁶ I am deliberately vague about what would be involved in an institution or group "taking itself to be justified by the intrinsic goodness of a wrongdoer's suffering." That might be a matter of the beliefs of the individuals involved, or a matter of what is codified in the laws or rules of the institution, or something else. Criminal punishment is not my primary focus in what follows, and the discussion of it that I do undertake will not require settling these questions; so I will not attempt to do so. I do, however, want to register my confidence that we can make sense of antagonism in legal settings, even though the paradigmatic forms of antagonism involve emotions like anger or hatred which might not be experienced by legal institutions in any straightforward way.

antagonism they condemn is incompatible with deeply morally important forms of love.⁷ That is a serious but underdeveloped concern. It deserves to be developed and examined at length. That does not happen here.

Instead, I search for a way of showing that some form of antagonism is morally necessary: that, no matter how problematic antagonism might seem, it is required by some deeply important part of human life. In chapters 2 and 3, I examine strategies that aim to connect resentment—which I, following many other philosophers, think is a form of anger and clearly antagonistic—to important interpersonal attachments. In chapter 4, I examine attempts to show that some antagonistic attitudes—not necessarily resentful anger—are required to restore a morally important state of affairs that has been destroyed or damaged by wrongdoing.⁸

Antagonistic response involves taking the wrongdoer’s suffering to be intrinsically valuable. I explore attempts to show this stance to be a valuable or important one to take. But why not just argue, directly, that the wrongdoer’s suffering is in fact intrinsically valuable? Why not argue that this view of the wrongdoer’s suffering would be *correct*?

Many of the antagonistic responses I discuss are emotional reactions. As Justin D’Arms and Daniel Jacobson have argued, showing that an emotion accurately presents its object does not entail that the emotion is morally appropriate. They give the example of being envious of a recently-tenured colleague who is also a close friend: the envy might present the colleague’s accomplishment accurately, but be the sort of thing that is inappropriate to feel about the

⁷ See, e.g., “Letter from Birmingham Jail,” p. 845.

⁸ This latter strategy is an attempt to show that *antagonistic* response—hatred or malice, for instance—is required to accomplish an important end. That must be carefully distinguished from showing that the wrongdoer’s *suffering* is required to accomplish some important end. That would give the wrongdoer’s suffering an instrumental value, which, again, is not antagonistic in the way that I am after.

accomplishment of a close friend.⁹ If D'Arms and Jacobson are correct about this, the fact that antagonistic emotions accurately present a victim's suffering as intrinsically valuable would not suffice to show that such emotions are morally acceptable.¹⁰

Of course even if some form of antagonistic moral response is shown to be necessary in one of these ways, we would not have a complete justification for it. Maintaining a morally important interpersonal attachment, or accomplishing an important end, would have to be considered in light of the moral problems with antagonism that ground the King-Gandhi ideal. It is an open possibility that the moral problems with antagonism are more serious than whatever attachment or benefit requires antagonism.

In any event, I argue that all these attempts to justify antagonistic response fail before they ever confront this issue. They fail to get off the ground in the first place: antagonism is not in fact required by our morally important attachments, and it does not contribute to bringing about an important form of restoration.

In chapter 5 I turn to another, more successful justificatory strategy. This strategy involves an appeal to annulment: to the prospect of retroactively altering past wrongdoing. There is something paradoxical about talk of altering the past; I think we can make sense of it, and appeal to it to justify hard treatment of a wrongdoer.

But when this strategy is examined closely, it is not so clear that it requires antagonism of the sort I have focused on. The annulment I describe requires that the wrongdoer be made to

⁹ See D'Arms and Jacobson, "The Moralistic Fallacy: On the 'Appropriateness' of Emotions," p. 82 for discussion of the envy example.

¹⁰ I should also admit that I am not even certain how to set about arguing that someone's suffering is intrinsically valuable, because I am not certain how to argue for anything's being intrinsically valuable. I am not certain how to argue about final ends. When I try, I find myself identifying some further goods that the object of my argument contributes to—which, of course, makes the object out to be instrumentally, not intrinsically, valuable.

suffer; but that need not involve antagonistic anger, or hatred, or malice. If this is correct, annulment may provide a well-worked-out, but instrumental, justification for inflicting some (perhaps quite extensive; perhaps even violent) suffering. That would be an interesting result—one that might push us away from a central tenet of nonviolent resistance—even if it did not amount to a justification of antagonism.

But that is the end of a long road. Onwards into the attempts to justify antagonism. I turn first to P. F. Strawson.

CHAPTER 2: MODIFICATION OF THE REACTIVE ATTITUDES

In this chapter I consider P. F. Strawson's defense of our moral "concepts and practices" in his immensely influential essay "Freedom and Resentment."¹¹ Strawson argues that antagonistic responses to wrongdoing are an important and ineliminable part of our moral lives.

Strawson situates our moral concepts and practices within *engaged interpersonal relationships*. It is an essential part of being in such relationships, he thinks, that we make certain broadly moral *demands* on each other. It is, in turn, essential to making a demand that we be susceptible to certain *attitudinal responses* if the demand should be violated. If Strawson is correct about this, relationships, demands, and responses stand or fall together: eliminating our susceptibility to certain attitudes in response to wrongdoing amounts to dropping the moral demand, and thereby exiting an interpersonal relationship.

"Exiting an interpersonal relationship" might call to mind severing one's contact with someone. Strawson has in mind a much deeper and more costly change to the way one relates to a person. Exiting an interpersonal relationship involves adopting what Strawson calls the "objective attitude" towards the person: one comes to view her as an object to be "managed or handled or cured or trained" (p. 79ff) and ceases to relate to her as a fellow human being.

Resentment is Strawson's central example of the attitudinal responses that are essential to participation in interpersonal relationships. It is antagonistic: it involves valuing another's

¹¹ Unless indicated otherwise, all citations in the text are to "Freedom and Resentment" as reprinted in Watson, *Free Will*. There are other defenders of antagonistic response to wrongdoing, most notably Jeffrie Murphy (see, for instance, *Getting Even: Forgiveness and Its Limits*). See also Hampton and Murphy, *Forgiveness and Mercy*; William Ian Miller, *Eye for An Eye*, an account of the role historically played by vengeance; and Peter French, *The Virtues of Vengeance*.

Others who write in Strawson's shadow have expressed hesitation about Strawson's claims in defense of antagonistic response. See for instance Gary Watson, "Responsibility and the Limits of Evil," p. 140; George Sher, *In Praise of Blame*, p. 138; and Derk Pereboom, *Living Without Free Will* pp. 212-213 ff.

suffering for itself, though sometimes in an attenuated way. Resentment entails, Strawson says, a “modification...of the general demand that another should, if possible, be spared suffering,” which modification can give rise to a “preparedness to acquiesce in the infliction of suffering”¹² on a wrongdoer.

In what follows, I treat resentment as a form of anger. Strawson himself does not claim that reactive attitudes like resentment are emotions, much less forms of anger; but that does seem (to myself, and to many of Strawson’s followers) like a plausible way of understanding his view. At the very least, resentment often does involve an affective element resembling the affect of anger.

I think that the connections Strawson sees between relationships, demands, and responses are real and important. But, I will argue, specifically antagonistic attitudes need not be a part of this close-knit group. I will argue that a form of sadness can keep us within engaged, demanding interpersonal relationships without being antagonistic in any of the ways I have just sketched. If that is correct, eliminating antagonistic response need not come at the high price of ceasing to demand goodwill and exiting an interpersonal relationship.

I

I will begin by describing the broad outlines of Strawson’s argument. Strawson re-poses questions about responsibility and blame in terms of certain attitudes we have in response to wrongdoing: personal, reactive attitudes, when we are wronged ourselves, and impersonal,

¹² See p. 90. There Strawson is speaking of indignation and disapprobation, the third-personal analogues of resentment; but he makes clear that his claims extend to resentment as well.

vicarious analogues of the reactive attitudes, when we consider wrongs directed at others.¹³

When I am wronged, I will naturally feel *resentful*; when I see someone else wronged, I will feel *indignant*.

Strawson introduces these attitudes by describing the broader context within which they arise. He begins by pointing to what he calls a “commonplace”: “the very great importance that we attach to the attitudes and intentions towards us of other human beings, and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about those attitudes and intentions” (75). This concern about the attitudes of others arises, Strawson says, within a wide variety of interpersonal relationships:

We should think of the many different kinds of relationship which we can have with other people—as sharers of a common interest; as members of the same family; as colleagues; as friends; as lovers; as chance parties to an enormous range of transactions and encounters. (76)

Within these relationships, Strawson continues, we demand goodwill or interpersonal regard of each other:

Then we should think, in each of these connections in turn, and in others, of the kind of importance we attach to the attitudes and intentions towards us of those who stand in these relationships to us... In general, we demand some degree of goodwill or regard on the part of those who stand in these relationships to us, though the forms we require it to take vary widely in different connections. (76)

Within any human interpersonal relationship, Strawson thinks, we demand goodwill or interpersonal regard. Resentment arises in reaction to violation of that demand.

Strawson expresses hope that focus on resentment and these related phenomena will generate new insight into old problems about responsibility. In particular, questions about whether determinism would, or should, undermine blame or responsibility can be re-posed as

¹³ Strawson discusses resentment and its third-personal, vicarious analogue separately; his argument applies to both, and I draw equally on his parallel discussions of the two in what follows.

questions about whether determinism would, or should, undermine the resentment that arises within this context, or the attitudes which are its impersonal analogues.

Strawson focuses on determinism's impact on these reactive attitudes; but, he says, he does not know what the thesis of determinism is. He only knows that if the thesis of determinism is true, it is universally true: all behavior is determined.¹⁴ If determinism undermines resentment and its interpersonal analogues, then, it does so in *every* case. Strawson aims to show that resentment could not be undermined universally. I am interested in Strawson's strong conclusion about resentment's resilience, not in the threat posed by determinism in particular; so in what follows I set that aspect of Strawson's discussion to one side.

Strawson first maps out the ways in which resentment and the like are ordinarily undermined. He asks "what sorts of special considerations might be expected to modify or mollify this feeling [of resentment] or remove it altogether" (77).

Strawson's answer to this question draws on his understanding of resentment as a reaction to a violation of the demand for goodwill. Two sorts of considerations will inhibit resentment, so understood.

One type of inhibiting consideration shows that, contrary to initial appearances, the demand for interpersonal regard has been met. In the face of the "appearance of this demand's being flouted or unfulfilled," considerations of this type "show this appearance to be mere appearance, and hence inhibit resentment" (85). This sort of consideration inhibits a reaction to the violation of a demand, by showing that the demand was *not violated in the first place*. Strawson lists considerations like "He didn't mean to" and "He couldn't help it" (77) as examples of this type of consideration.

¹⁴ See, e.g., p. 80.

Another type of inhibiting consideration undermines the demand for interpersonal regard itself. Such a consideration “tends to inhibit resentment because it tends to inhibit ordinary interpersonal attitudes in general, and the kind of demand and expectation [i.e., for inter-personal regard] which those attitudes involve” (86). This sort of consideration inhibits a reaction to a violation of a demand by *undermining the demand*. These considerations often indicate that the purported agent was not in fact an agent with a will, capable of showing interpersonal regard: “He’s a hopeless schizophrenic,” for instance. Or these considerations show that, in the circumstances, the demand was not in effect: “He has been under very great strain lately” (78). For ease of reference, call this second way of inhibiting resentment an *exemption*.¹⁵

I want to emphasize that this taxonomy follows from Strawson’s analysis of resentment and the like as, most fundamentally, reactions to a *demand’s* being *violated*. A reaction to a violation of a demand may be inhibited if the demand was not in fact violated or if the demand itself is undermined.¹⁶

¹⁵ Following R. Jay Wallace’s use of the term “exemption” in *Responsibility and the Moral Sentiments*. This term is not used by Strawson, but it has gained wide currency in the literature.

¹⁶ (This note and notes 10, 11, and 13 below attempt to strengthen my claims about “Freedom and Resentment,” by working through interpretive questions about the text in greater detail. Those not interested in the details of Strawson’s essay may ignore them.) Strawson’s discussion of the category of resentment-inhibitions that I have labeled ‘exemptions’ in fact proceeds (on pp. 78-79, and again on p. 81) without any mention of the demand for goodwill. Strawson instead characterizes exemptions in terms of the “objective attitude (or range of attitudes)” towards someone. This attitude, he says, “cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships; it cannot include resentment, gratitude, forgiveness, anger...” (79). This way of proceeding creates the impression that Strawson’s exemptions involve undermining relationships, instead of undermining demands.

Exemptions from the demand for goodwill do involve exiting an interpersonal relationship with the person exempted; but it is important to recognize that we exit the interpersonal relationship *in virtue of the fact that we cease demanding goodwill*. I think the most compelling evidence for giving the demand, not the relationship, the main explanatory role here is the way that doing so makes clear sense of Strawson’s separation of inhibiting considerations into two kinds (into considerations that undermine the demand, on the one hand, and ones that show it to be satisfied, on the other). In addition, even though it is not foregrounded in Strawson’s first discussion of exemptions, the role of the demand is apparent elsewhere.

Just before discussing exemptions, Strawson describes the other class of considerations that inhibit the reactive attitudes (namely, the considerations that show that the demand for goodwill was satisfied). Contrasting those considerations with exemptions, Strawson says that non-exemptions “do not suggest that the agent is in any way an inappropriate object of that kind of demand for goodwill or regard which is reflected in our ordinary reactive

Strawson argues that neither sort of consideration operates universally. It is, first of all, simply unlikely that a consideration of the first sort would operate universally: it is simply unlikely that everyone always satisfies the demand for goodwill.¹⁷

The more pressing question is whether an exemption might operate universally: whether the demand for goodwill, and hence *all* reactive attitudes—including not only resentment and indignation but also positive reactive attitudes, like gratitude, that we might have when the demand is satisfied—might be lifted altogether.

Recall that demanding goodwill of someone is an essential part of being in an interpersonal relationship with her. If I cease to demand goodwill of someone, to that extent I withdraw from my interpersonal relationship with her. I view her objectively, as an object to be manipulated or controlled rather than as a person to be engaged with. Exempting everyone from the demand for goodwill all the time, then, entails ceasing to participate in any engaged interpersonal relationships whatsoever.

Strawson argues that an across-the-board withdrawal from interpersonal relationships on these grounds is not possible for us. Just why this is impossible is not straightforwardly clear,

attitudes” (78). This strongly suggests that the contrast class, the exemptions I am interested in, *do* indicate that the demand for goodwill is inappropriate.

Later, Strawson repeats his argument in the context of the vicarious, third-personal analogues of resentment and the like. Strawson makes clear, throughout this second discussion, that his reasoning proceeds along the exact same course as it does in the first-personal case. When he comes to exemptions, Strawson writes:

Again the parallel holds for those generalized or moral attitudes towards the agent which we are now concerned with. The same abnormal light which shows the agent to us as one in respect of whom the the personal attitudes, *the personal demand*, are to be suspended, shows him to us also as one in respect of whom the impersonal attitudes, *the generalized demand*, are to be suspended. ... We may say: to the extent to which the agent is seen in this light, he is not seen as one on whom *demands and expectations* lie in that particular way in which we think of them as lying when we speak of moral obligation; he is not, to that extent, seen as a morally responsible agent, as a term of moral relationships, as a member of the moral community. (86, emphasis added)

The presence of the demand in the account is much clearer here.

¹⁷ Strawson also points out that eliminating resentment in this way would not amount to undermining responsibility altogether, in the way that concerns incompatibilists like his “pessimist” character. This way of universally eliminating resentment would simply mean that everyone—perhaps quite responsibly, and praise-worthily—bears everyone else the appropriate level of goodwill.

because Strawson seems to make several different moves at this point. Because the main argument of this paper will be consistent with any reading of Strawson's claims on this point, I will only sketch the several distinct claims that Strawson appears to make.

First, appealing to "the human commitment to participation in ordinary inter-personal relationships," which is "thoroughgoing and deeply rooted" (81), Strawson seems to say that, as a matter of natural fact, humans are incapable of withdrawing from all our interpersonal relationships. Relatedly, speaking of "the human isolation" (81) which this would entail, and appealing to "an assessment of the gains and losses to human life" (83), Strawson also seems to say that doing so would simply be unpleasant. Finally, some of Strawson's remarks also suggest that, as a *conceptual* matter, we cannot imagine withdrawing from these relationships altogether on the grounds of a consideration like the truth of determinism. Interpersonal relationships (and the phenomena essentially connected to them) are, Strawson says, "part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general framework" (83).¹⁸ The existence of interpersonal

¹⁸ In the text, Strawson says that "our natural human commitment to ordinary interpersonal *attitudes*"—not *relationships*—is "a part of the general framework..." (83, emphasis added). When this suggestion recurs near the end of "Freedom and Resentment," it is once again stated in terms of attitudes:

Inside the general structure or web of human *attitudes and feelings* of which I have been speaking, there is endless room for modification, redirection, criticism, and justification. But questions of justification are internal to the structure or relate to modifications internal to it. The existence of the general framework of *attitudes* itself is something we are given with the fact of human society. As a whole, it neither calls for, nor permits, an external 'rational' justification. (91, emphasis added)

I read these as claims that the general system of interpersonal relationships, constituted by a wide variety of attitudes, is beyond rational criticism. That system may be described as a web of attitudes and feelings, as Strawson does; but I think it best to characterize the system in terms of relationships rather than attitudes. That is because speaking in terms of attitudes here is liable to mislead.

In light of Strawson's focus on the reactive attitudes, it is easy to take Strawson to be claiming that what is beyond criticism are the specific *reactive* attitudes Strawson identifies, including resentment. But Strawson avoids making such a claim, speaking instead of "ordinary inter-personal attitudes in general" (82-3) and "the general structure or web of human attitudes and feelings" (91). If Strawson felt that he could claim, straightaway, that the reactive attitudes in particular are impossible to eliminate, or beyond justification, there would be no payoff to connecting those reactive attitudes to the demand for goodwill or to interpersonal relationships in general. I do not think that those connections are irrelevant to Strawson's argument; they seem, instead, quite central to it. As I read

relationships is not subject to revision in light of reasons, because it is part of the framework within which such reason-driven revision makes sense. Rationally undermining the framework which gives sense to rational appraisal is, in some sense, incoherent.

In one (or more) of these senses, it is not possible for us to withdraw from all interpersonal relationships. (Or, in the last sense, it is not possible for us to have reasons for withdrawing from all such relationships.) Because participation in these relationships entails demanding goodwill, it is similarly impossible for us to drop all such demands. And finally, since that universal exemption is the only way that the reactive attitudes like resentment might be eliminated altogether, it is not possible to eliminate those reactive attitudes altogether.¹⁹

Strawson, then, those connections serve to connect the reactive attitudes to the general structure, about which Strawson feels he can make the “it is beyond rational criticism” claim.

¹⁹ Depending on the way we understand Strawson’s rejection of complete withdrawal from interpersonal relationships, this argumentative move will vary in important ways. I want to emphasize, for those concerned about these matters, that *my main argument in this paper will be consistent with any reading of Strawson’s claims on this point*. Some consideration of these two readings of Strawson will help to make that clear.

If Strawson’s rejection is just a claim about what is motivationally feasible for human beings, his argument will run along these lines: if some consideration (e.g., the truth of determinism) undermines the demand for goodwill, it thereby leads us to withdraw from all interpersonal relationships; but, we are motivationally incapable of withdrawing from all interpersonal relationships; so, no consideration can motivate us to eliminate the demand for goodwill. This version of Strawson’s argument leaves it open that some consideration *ought* to undermine the demand for goodwill, and we *ought* to withdraw from interpersonal relationships, although we are motivationally incapable of doing so.

Alternatively, I have said, Strawson might be making a claim about the role that these phenomena play in structuring human life: they are part of “the general framework of human life” (83), which is required for evaluative claims to make sense. A rational evaluation of the entire framework simply does not make sense. If some reason leads us to think that all demands for goodwill are unjustified, and so to think that all relationships are unjustified, we are engaged in just that senseless rational evaluation of the entire framework.

Some of Strawson’s discussion is strongly suggestive of this second interpretation. He writes of “the general structure or web of human attitudes and feelings,” saying that “questions of justification are internal to the structure or relate to modifications internal to it. The existence of the general framework of attitudes itself is something we are given with the fact of human society. *As a whole, it neither calls for, nor permits, an external ‘rational’ justification*” (91, my emphasis).

In addition, on this second reading the possibility left open by the first sort of argument—that we ought to drop the demand for goodwill altogether, but we are motivationally incapable of doing so—is closed off. There is no way to assert that we ought to drop the demand altogether, because that would entail that the general framework of human life is somehow rationally undermined, which is incoherent.

The latter claim—that interpersonal relationships and the demand for goodwill are part of the general framework within which talk of justification makes sense, so that it is not coherent to ask if the framework itself is justified—may seem obscure to some philosophers. (For one, it may seem that we can at least ask if the general framework is itself internally coherent or not.)

This argument relies on there being very tight connections between interpersonal relationships, the demands we make within such relationships, and our susceptibility to reactive attitudes like resentment: participation in interpersonal relationships requires demanding goodwill, and demanding goodwill requires that we be susceptible to reactive attitudes like resentment when the demand is violated. Those connections enable Strawson to extend his claim about the status of interpersonal relationships to the reactive attitudes.

In what follows, I want to focus on the connections between these three phenomena. If Strawson is correct that they are this closely linked, the Gandhi-King ideal is not, after all, so ideal. Eliminating altogether our susceptibility to resentment, because it seems dangerous or morally problematic in some way—say, in response to the considerations raised by Gandhi and King—would require an impossible and unattractive exit from interpersonal relationships.

Additionally, if Strawson is correct that these three phenomena all stand or fall together, eliminating the susceptibility to resentment in any individual situation will carry the high cost of exiting an interpersonal relationship. If I am no longer susceptible to resentment in my interactions with someone, I am no longer demanding goodwill of her, and I thereby adopt the objective attitude towards her. Doing this in a particular situation is not impossible in the way that adopting the objective attitude universally is impossible—we can and do adopt the objective attitude in a variety of situations, for instance with ordinary, healthy people who are so

Exactly how to understand this portion of Strawson's argument has been a major focus of philosophical discussion of "Freedom and Resentment." But it is not my focus. Strawson's argument has two main parts: a claim that we cannot exit interpersonal relationships altogether, and a claim that interpersonal relationships, the demand for goodwill, and the reactive attitudes are so closely tied together that they stand or fall as one. I am focused on the second claim; in particular, I will attempt to separate the antagonistic reactive attitudes from the demand and the relationship. Then we can challenge the antagonistic reactive attitudes, without challenging the demand or the interpersonal relationships. We need not address Strawson's argument about the status of interpersonal relationships—or even settle its specifics—to be able to mount that challenge.

obnoxious that it is too much of a strain to maintain engaged interpersonal relationships with them²⁰— but doing so comes at the cost of a retreat to objectivity.

In what follows, I will focus on resentment’s connection to demands and relationships. I will argue that our susceptibility to specifically antagonistic attitudes, like resentment, is not as closely tied to the demand for goodwill or to our participation in interpersonal relationships as Strawson suggests.²¹ We can rely on other, non-antagonistic attitudes to secure the demand for goodwill and participation in involved interpersonal relationships, while eliminating susceptibility to antagonistic reactive attitudes.

II

Antagonistic response to wrongdoing *does* do some work keeping the demand for goodwill in place and maintaining interpersonal relationships.

When Strawson introduces the phenomenon which he will go on to label the demand for goodwill, he describes it as a *desire*, pointing out “*how much we actually mind, how much it matters to us*, whether the actions of other people...reflect attitudes towards us of goodwill,

²⁰ See pp. 79-80.

²¹ It may be helpful to compare my reading of Strawson with the one R. Jay Wallace develops in his influential book *Responsibility and the Moral Sentiments*. Wallace accepts that the reactive attitudes and a certain sort of demand are closely connected elements of what he calls “holding someone to an expectation” see pp. 21-5. Wallace thus accepts the connection that I will argue against, between making demands of someone and reacting in an antagonistic manner. (I will, however, accept a more limited version of this connection: see section III below, especially pp. 33-5.) Wallace certainly does think that the reactions in question are antagonistic: p. 93 makes it clear that his project of normatively assessing holding people to expectations is predicated on the antagonistic character of such reactions. Wallace also rejects connecting the reactive attitudes to interpersonal relationships (see pp. 25-33). He supports this rejection by pointing out—quite rightly—that simply defining the reactive attitudes as those attitudes we have within interpersonal relationships will be quite unsatisfying. Curiously, Wallace does not, so far as I can see, consider tying the reactive attitudes to interpersonal relationships by way of an expectation or demand. That is, he does not seem to consider the claim—which I take to be much more plausible, and quite central to “Freedom and Resentment”—that it is the *demand* which is essential to participation in interpersonal relationships; and so he does not consider the more indirect way of connecting the reactive attitudes to interpersonal relationships, by way of their connection to that demand, that I attribute to Strawson.

affection, or esteem on the one hand or contempt, indifference, or malevolence on the other” (76, my emphasis).

Antagonistic response to the absence of goodwill does express that desire. But of course there is much more to demanding something than simply desiring it. To see what more is required, consider some scenarios in which one might have desires about someone’s attitudes but fail to demand anything of her. These are scenarios in which one takes the objective attitude towards someone—and so fails to engage in a demanding interpersonal relationship with her—even though one is concerned about her attitudes.

A detailed example will be useful. Consider a racist. Suppose that this person shows me ill will. Say she unjustifiably blames my ethnic group for unemployment and other social ills. Suppose that she was raised in an extremely intolerant society—say, in an economically depressed coal-mining town in eastern Kentucky. In virtue of the social and economic pressures this person experienced growing up, she might be unacceptably intolerant towards my ethnic group: she might unjustifiably blame us for unemployment and other social ills.

Now suppose that I respond to the racist with pity and charity. *You poor thing, I say, it’s terrible, the pressures you were under growing up. Look, I don’t take offense, because I know how your background led you to where you are now. But you’re living a very impoverished life. You’re missing out on valuable friendships; you’re stuck in an unrealistic and angry way of viewing the world. You’ll be much better off if I help you to improve your outlook on life.*

Even if this benevolent concern for the wrongdoer’s well-being is sincere, it might strike us as a condescending and belittling way of responding to a racist. I do not demand that she bear me goodwill; I am detached from the quality of her will, and so I take a detached, objective view of her.

I am detached even though I am concerned with the racist's ill will as a problem for *her*. Even though I am concerned about her well-being, her ill will does not matter *to my well-being*. This is a belittling, condescending attitude to take towards someone, precisely because it is purely benevolent. (This combination of benevolence and belittling may explain why it is attractive to take this sort of stance when comforting a decent person who has been insulted or mistreated by a racist. One is still being *nice* to the racist, even while one is dismissing her and her attitudes as unimportant.)

This is a belittling way of viewing the racist because it is a detached, disengaged—in Strawson's terminology, *objective*—way of relating to her. I fall short of demanding goodwill of her. In this case, I think, that is explained by the fact that my concern about the racist's attitudes involves no *vulnerability* to the racist's ill will. Her ill will cannot leave me worse off. I view it only as a problem for her own well-being. I may, as Strawson puts it, see the racist as an "object of social policy" to be "cured" (79); but her will does not matter to me, in the sense that her ill will is no threat to me.

An absence of vulnerability seems to explain some cases of disengaged concern about another's attitudes.²² Being vulnerable to someone's attitudes contributes to participating in an engaged, demanding interpersonal relationship with her.

It may, of course, be quite traumatic to be the object of racist ill will. Then if I do find it traumatic to be the object of racist attitudes in this way, the specific type of disengagement and condescension just described is no longer available. Does that mean that when I am vulnerable to

²² This rules out a Socratic view of wrongdoing, on which the victim of wrongdoing is not seriously badly off because her soul remains just. If vulnerability is important to engaged interpersonal relationships, and we cannot entirely give up such relationships, this Socratic view is problematic: to the extent that it does not involve vulnerability to the attitudes of others, it grounds an impossibly disengaged, objective way of relating to each other.

the racist's ill will, I am therefore making a demand within an engaged interpersonal relationship?

No. Vulnerability is necessary for engagement with the racist, but it is not sufficient. I might be vulnerable to her and still take a detached stance towards her. *How awful*, I might say. *This person is really quite painful to be around. Her mind has been systematically perverted by the pressures she experienced growing up: her view of the world is so warped that she has lost touch with reality. She needs medical treatment—and in the meantime, I'm going to avoid her as much as I can.*

This is disengaged, but in a different way. I am vulnerable to the racist. Her ill will towards me can damage my well-being: the ill will she bears me threatens to harm my emotional composure and sense of self-worth. But I do not take the racist seriously as a person. I think she is unhealthy, warped, to the point that she is unable to recognize moral truths. I do not view her as a healthy, full-fledged moral agent.²³

Because I view the racist as disabled in this way, I do not demand goodwill of her. Because I take her agency to be impaired, I exempt her from the demand for goodwill.²⁴ She is merely a threatening or dangerous part of my environment—“something certainly to be taken account, perhaps precautionary account, of” (79), as Strawson puts it, but not an agent of whom I may make demands.

²³ Whether or not this disengagement is *condescending* will depend on whether or not this is an accurate description of the racist. If it is accurate, it seems entirely appropriate. If it is inaccurate—if the racist is a healthy, full-fledged agent, but I take her to be impaired—my stance seems condescending. I think my description of the influence the racist's upbringing has on her attitudes leaves it open whether she is, in fact, impaired.

²⁴ Thanks to Andrew Khoury for pointing out to me that this is a natural way of reacting to the racist, and one that amounts to exempting the racist from the demand. Interestingly, all three ways of adopting the objective attitude that I describe involve inhibiting the demand (if I am not vulnerable to the racist's ill will I do not demand goodwill of her, either), and so they all fall into Strawson's second category of resentment-inhibitors; but it is most natural to label this specific way of adopting the objective attitude—viewing the wrongdoer as an unhealthy or impaired agent—an “exemption”. To my ear, it is much less natural to say that when I am invulnerable to the racist I am *exempting* her from the demand.

This indicates a second way of falling short of demanding goodwill of someone. Even though one cares about someone's attitudes and is vulnerable to those attitudes, one may fail to view that person as a healthy, full-fledged agent.²⁵

And yet even when satisfying these requirements another form of disengagement is possible. Suppose that I am vulnerable to the racist's ill will. Suppose further that I do take the racist to be a full-fledged, healthy agent. I do view her as an agent whose ill will may injure me. Even so, my response to her ill will might be quite *manipulative*. Perhaps I notice that the racist has a special aversion to direct confrontation. I might then be intentionally confrontational with her, with an eye towards shaping and controlling her behavior toward me.

When I treat the racist in this way, I do view her as a healthy, full-fledged agent. Of course, I do take advantage of a peculiar feature of her psychology: a disposition to avoid angry confrontation. But that disposition does not make the racist less than a healthy, full-fledged agent. Peculiarities like strong aversions are well within the healthy range for human agents. I leave specification of this range open, but I think that any plausible account of healthy human agency will leave room for strong aversions of the sort ascribed to the racist in this scenario.

Although I view the racist as a full-fledged, healthy agent, to whom I am vulnerable, I am once again relating to her in a condescending, disengaged way. Although I recognize her agency, I treat it in an objective fashion. It is condescending of me to attempt to shape the racist's behavior by manipulating her, rather than engaging with her as a person, demanding that she act

²⁵ In the example I give, I take the racist to fall short of being a healthy, full-fledged moral agent in a particular way: by being incapable of recognizing moral truths. There are, of course, many other ways of failing to be a healthy, full-fledged moral agent. Exactly what is required for such agency is a serious and difficult substantive question; I intend my argument in the text to be compatible with many ways of answering it (—although I might have to adopt a different illustrative example, if the “able to recognize moral truths” requirement were to be rejected).

well, and then allowing her to respond. I treat her as something to be “managed or handled...or trained” (79), or “something to be understood and controlled in the most desirable fashion” (82).

This indicates a third requirement on demanding goodwill and participating in an interpersonal relationship with someone. In addition to *viewing* someone as a healthy, full-fledged agent, one must *treat* her as such an agent. That requires avoiding manipulating her.

I began this section with the observation that when the demand for goodwill is violated, angry, antagonistic reactions to that violation, like resentment, seem to do important work keeping that demand in place and maintaining an engaged interpersonal relationship with the wrongdoer. The three conditions I have described enable us to see clearly the work that such reactions perform. Angry resentment expresses vulnerability; it involves viewing the wrongdoer as a healthy, full-fledged agent; and it treats her as such an agent.

At this point one might worry that my characterization of demands made within interpersonal relationships is incomplete. One might think that something more is required by participation in an engaged, demanding interpersonal relationship. I cannot rule that out decisively. If there are further conditions on participation in interpersonal relationships, I would like to hear about them. Such conditions would add to our picture of the work done by angry, antagonistic reactions to ill will.

Now I can state more precisely my central question: is an antagonistic reaction like anger *required* to do this work in the aftermath of a violation of the demand? If there are nonantagonistic reactions capable of doing the very same work as antagonistic reactions,

switching to such reactions would enable us to preserve interpersonal relationships while eliminating antagonistic reactions.

I think we can conceive of some such nonantagonistic alternatives. Consider yet another way of responding to the racist. In light of the racist's upbringing, I might modify my reactive attitudes. Without taking that upbringing to have impaired the racist's agency, my resentment of her might diminish or dissipate altogether. *It's sad how poorly you're treating me*, I might say. *I expect better of you, and you've fallen short. I'm not angry with you, but I am very, very disappointed.* While still demanding that the racist be tolerant and goodwilled, my resentment might be replaced by sadness or disappointment at the racist's ill will.

I want to argue that this shift would not involve giving up the demand for goodwill or exiting an interpersonal relationship. To begin, notice that it is not anger *simpliciter* that keeps the demand in place. One can be benevolently angry at the plight of an agent to whom one is not vulnerable, or angry at the dangerous behavior of a seriously impaired agent. It is *resentment* that expresses vulnerability, and it is *resentment* that involves viewing and treating its object as a healthy, full-fledged agent. It is *resentment*, not anger in general, that satisfies the necessary conditions on participation in engaged, demanding interpersonal relationships.

Resentment is a subclass—what Justin D'Arms and Daniel Jacobson call a “sharpening”—of the more basic emotion of anger.²⁶

The claim that resentment is an emotion raises broader questions, about the nature of emotions, which are D'Arms and Jacobson's primary interest; to avoid being drawn into those questions, I set to one side the claim that resentment is an emotion. Instead I want to draw on

²⁶ See D'Arms and Jacobson, “The significance of recalcitrant emotion (or, anti-quasijudgmentalism).”

their useful framework for understanding the relationship between broader phenomena with affective and cognitive elements, like anger, and narrower phenomena, like resentment.

D'Arms and Jacobson illustrate the sharpening phenomenon that they have in mind with “tenure-denial anger” or “tenure rage,” which is the anger one has that one has been denied tenure. This is a *cognitive* sharpening: the subclass of anger is constructed by specifying a belief that all members of the subclass happen to share.²⁷ There are other ways of creating sharpenings of emotions, D'Arms and Jacobson indicate, for instance by the emotions' causes or by their motivations.²⁸

D'Arms and Jacobson propose that resentment is a cognitive sharpening of anger, distinguished by the belief that one has been wronged.²⁹ Since I take seriously Strawson's focus on the demand for goodwill and involved interpersonal relationships, I would prefer to distinguish resentment not by the belief that one has been wronged, but rather by the beliefs and causal connections that I have just identified as essential to participation in such relationships. This can, I suggest, be captured by a simple cognitive sharpening of anger: resentment is a subclass of anger, distinguished by the belief that one has been left worse off by the ill will of a full-fledged, healthy agent.

This sharpening of anger satisfies the three conditions set out above. Suppose I have such a reaction to the racist. This sharpening of anger does reflect my vulnerability to her bigoted attitudes: this sharpening is, after all, anger at *having been made worse off* by such attitudes. This sharpening also involves *viewing* the racist as a healthy, full-fledged agent: it is characterized by the belief that she is such an agent. And this sharpening of anger *treats* the wrongdoer as a

²⁷ See p. 137.

²⁸ See p. 138, n. 22. D'Arms and Jacobson suggest that spite and vengefulness are motivational sharpenings, though they do not pursue that suggestion.

²⁹ See p. 143.

healthy, full-fledged agent. There is nothing manipulative in believing that someone's attitudes have left one worse off, and being angry about that. (Of course, such an emotional reaction may be *used* to manipulate someone: if I know that the racist is so averse to hostility that she will do almost anything to avoid expressed anger, I may cultivate my resentful reaction to her behavior, in order to manipulate her into avoiding such behavior in the future. But then it is my intention to cultivate my resentment with this aim in mind that is manipulative, not my resentment itself. Depending on the racist's aversions, I might manipulate her by cultivating some other attitude of mine, or by making use of some other feature of the world that she is especially susceptible to.)

I am not fully committed to the details of this analysis of resentment. Perhaps resentment is, in part, a causal sharpening of anger: perhaps resentment is anger which has a distinctive causal history. But it is not necessary for my purposes to get these details of the sharpening that characterizes resentment exactly correct. What I want to point out is that *whatever sharpening distinguishes resentment from anger in general, we can apply the very same sharpening to sadness instead of anger.*

Sadness, like anger, is an affective response to something that one takes to be bad or unfortunate in some way.³⁰ We can describe a subclass of sadness that is set apart from sadness in general by the belief that a full-fledged, healthy agent has shown one ill will, which left one worse off. Or, we can describe a sharpening that is distinguished by its causal origin in the attitudes of a full-fledged agent which leave one worse off. Or—to take D'Arms and Jacobson's

³⁰ Sharpening happiness or some other emotion in the same way is ruled out here. An interpersonal masochist who is happy to be the object of ill will would be vulnerable to that ill will; she might very well view and treat her tormentors as healthy, full-fledged agents; but she would not thereby demand goodwill of them. Her happiness, even though it might be sharpened in the appropriate way, would not express any desire that her tormentors show her goodwill instead.

own account of resentment—we can describe a sharpening of sadness that is distinguished by the belief that one has been wronged.

This resentment-like sharpening of sadness might be called a form of disappointment (although I do not claim that it is the same as the natural phenomenon of disappointment). Because, by stipulation, it involves the same elements that ensure that resentment keeps one within an engaged, demanding interpersonal relationship, this sharpening of sadness also keeps one within an engaged, demanding interpersonal relationship. If I respond to the racist's ill will with unsharpened sadness—or, for that matter, unsharpened anger—I am not responding to her as a participant in an engaged interpersonal relationship. I can, however, respond to her ill will with a sharpening of sadness that is related to sadness in general as resentment is related to anger in general. That response satisfies the conditions on participation in an interpersonal relationship, just as well as resentment does. When I am *disappointed* with the racist, in a way that parallels being *resentful* of her, I continue to demand goodwill of her and maintain an engaged interpersonal relationship with her.

Some space does remain to resist this conclusion, but it is tightly circumscribed. Earlier, to characterize the work done by antagonistic reactions, I identified three necessary conditions on participation in an engaged interpersonal relationship: vulnerability; viewing someone as a healthy, full-fledged agent; and treating her as such an agent. At that point I acknowledged that this list might be incomplete: there might be some other requirements on participation in such a relationship, which antagonistic reactions satisfy.

That alone might seem to enable resistance to my conclusion, but it does not. For if the proposed additional condition is satisfied, not by anger *simpliciter*, but by *the sharpening that*

characterizes resentment, then the parallel sharpening of sadness is equally well-equipped to satisfy that condition. What is needed is an additional condition that is satisfied by the mere fact of antagonism, not by the sharpening that characterizes resentment. In addition, this condition must be such that it can only be satisfied by antagonism. I do not see why interpersonal relationships would require this, although I do not have a conclusive argument against the possibility. So I conclude that replacing resentment with the sharpening of sadness I have identified would not involve giving up the demand for goodwill.

I take it to be clear that this possible sharpening of sadness would not be an antagonistic emotion in the way that anger is: it does not involve the antagonistic motivations or affects characteristic of anger. And yet it does keep us in demanding, engaged interpersonal relationships. Hence our participation in such relationships does not stand or fall with our susceptibility to antagonistic reactions: we might shift to this sharpening of sadness, and thereby eliminate susceptibility to antagonistic reactions, without dropping the demand for goodwill or exiting interpersonal relationships.

III

I have described a shift in our reactions to wrongdoing, from antagonism to a sharpened form of sadness. I argued that this shift would not require dropping the demand for goodwill or withdrawing from interpersonal relationships. If that is correct, whatever considerations Strawson draws on to block exiting from interpersonal relationships altogether, or dropping the demand altogether, fail to block dropping resentment altogether.

I am uncertain whether Strawson would resist my argument. He does pay extensive, recurring attention to the antagonistic reactive attitude of resentment. And yet near the beginning

of his discussion of resentment, Strawson indicates that he is using it as a stand-in for a broader range of attitudes. “Resentment and gratitude are... a usefully opposed pair,” he says; “but, of course, there is a whole continuum of reactive attitude and feeling stretching on both sides of these and—the most comfortable area—in between them” (77). The way Strawson frames his use of resentment here does at least open the way for it to be a stand-in for a wider range of attitudes, including ones that are not antagonistic.³¹

There is, however, a single point at which Strawson seems to indicate that antagonistic reactions are an essential part of the demand for goodwill. I mentioned this discussion at the beginning of this chapter; now it is worth quoting and considering at length:³²

[T]hese attitudes of disapprobation and indignation [the third-person analogues of resentment] are precisely the correlates of the moral demand in the case where the demand is felt to be disregarded. The making of the demand *is* the proneness to such attitudes. ... The partial withdrawal of goodwill which *these* attitudes entail, the modification *they* entail of the general demand that another should, if possible, be spared suffering, is, rather, the consequence of *continuing* to view him as a member of the moral community; only as one who has offended against its demands. So the preparedness to acquiesce in the infliction of suffering on the offender which is an essential part of punishment is all of a piece with this whole range of attitudes of which I have been speaking. (90, Strawson’s emphasis)

Strawson first identifies the demand³³ with susceptibility to particular attitudes (here, the third-personal, vicarious analogues to resentment, namely indignation and disapprobation). He then

³¹ Moreover, some remarks Strawson makes at the end of the essay suggest that he is open to being corrected about the specific forms the reactive attitudes can take. He writes of “... the great variety of forms which these human attitudes may take at different times and in different cultures.” That, he thinks, “makes one rightly chary of claiming as essential features of the concept of morality in general, forms of these attitudes which may have a local and temporary prominence” (92-3).

³² One might think that Strawson’s talk of “our natural human commitment to ordinary interpersonal attitudes” (83; see also 91) indicates that we have such a commitment to the antagonistic attitudes in particular. I think that is a bad misreading of Strawson’s claim, however, which is carefully directed at the *general* “framework,” “structure,” or “web” of attitudes, not at particular attitudes. (See also n. 10 above.)

³³ Here, within the context of Strawson’s discussion of the vicarious, third-person analogues of resentment, the demand in question is “the moral demand,” rather than the demand for goodwill; and it is situated within “the moral community” rather than interpersonal relationships. Strawson’s remarks on p. 84 indicate that this demand made by the moral community is the same as the demand for goodwill made within an interpersonal relationship. And on p.

claims that those attitudes entail a *partial withdrawal of goodwill*, and a *greater willingness to acquiesce in the infliction of suffering*. Violation of the demand for goodwill entails a withdrawal of goodwill and a willingness to see someone suffer. Making the demand entails susceptibility to antagonistic attitudes.

I have argued that we could respond to violation of the demand with a form of sadness that is not connected with antagonism. So Strawson's claims here are incorrect.

But I do not wish to reject this passage in its entirety. There is an important claim being made in it that, contrary to initial appearances, is compatible with my argument. Strawson *identifies* a demand with the susceptibility to certain attitudes. I think he is correct in this. Demands are partly constituted by the attitudes they give rise to; it is impossible to separate a demand from its characteristic attitudinal expressions.³⁴ I do not dispute this connection.

Identifying demands with the susceptibility to particular attitudes seems to undermine my argument. I claimed that we can eliminate our susceptibility to antagonistic attitudes, without eliminating the demand for goodwill. But now I am agreeing that the demand for goodwill, like any demand, is partly constituted by susceptibility to particular attitudes. Then eliminating susceptibility to those attitudes must involve eliminating the demand.

My conclusions survive this concern if they are formulated more carefully. In a loose sense, which I adopted in previous sections, we may speak of "the demand which is essential to interpersonal relationships." I argued that this demand is compatible with responding to violations of the demand with either (sharpenings of) anger or (sharpenings of) sadness. In a strict sense, though, we should individuate demands by their characteristic expressions. Then I

86 Strawson indicates that "the moral community" is itself a generalization of demanding interpersonal relationships.

³⁴ This has been recognized by others working in Strawson's shadow, most prominently R. Jay Wallace. See *Responsibility and the Moral Sentiments* pp. 12 and 20-25.

should distinguish between “the demand which issues in (sharpenings of) anger” and “the demand which issues in (sharpenings of) sadness.” Speaking carefully, my argument in the previous section established that either the anger-issuing demand or the sadness-issuing demand suffices to keep us in the engaged interpersonal relationships which are so important to Strawson.

What we cannot imagine eliminating altogether, then, is not susceptibility to *resentment*, or *antagonism*; it is susceptibility to some attitudes which keep in place an interpersonally involved, vulnerable demand for goodwill. That demand may take different forms, involving different reactions. There is a real sense in which the demand issuing in (sharpenings of) anger is distinct from the demand issuing in (sharpenings of) sadness. But any such demand—antagonism-generating or not— keeps in place involved interpersonal relationships.

IV

At the beginning of “Freedom and Resentment,” Strawson indicates that he agrees with his “optimist” character’s claim that “the facts as we know them supply an adequate basis for the concepts and practices which the [incompatibilist] pessimist feels to be imperilled by the possibility of determinism’s truth” (73). Strawson’s argument gives an adequate basis to those concepts and practices which cannot be suspended without exiting an interpersonal relationship. I have argued that making some involved, vulnerable demand for interpersonal regard is secure in this way; but our having specifically antagonistic reactions to violations of the demand is not similarly secure. I think that clearly acknowledging this limit to Strawson’s argument has intriguing consequences.

Within our moral practices, and compatible with Strawson's claims that we cannot imagine eliminating those practices altogether, the possibility of deep and far-reaching revision remains. Consider a universal version of the shift to disappointment. Suppose that, inspired perhaps by Gandhi and King, we universally respond to wrongdoing and ill will with disappointed sadness rather than resentment or other antagonistic attitudes. In responding this way we do not stop demanding goodwill of each other; people's attitudes still matter to each other; we do not withdraw from genuine, demanding interpersonal relationships with each other. But the texture of those relationships is dramatically changed. They are far less antagonistic.

I have not argued that this would be a better form for human life to take. I have not argued that it is, psychologically speaking, possible for this change to come about. (I consider some arguments along those lines in the next chapter.) But I have established that Strawson's argument leaves room for such a change.

As Gary Watson has pointed out,³⁵ Gandhi and King seem to realize this ideal, without giving up on demanding goodwill: they attempt to "*stand up* for themselves and others;" they "*confront* their oppressors," while "*urging* and even *demanding* consideration for themselves and others." But "they manage, or come much closer than others to managing, to do such things without vindictiveness or malice."

Some might doubt that Gandhi and King really managed to eliminate antagonism altogether. Perhaps in the end that is not, psychologically speaking, a possibility for human beings.³⁶ It might still stand as an ideal for human relationships, one that we endeavor to realize

³⁵ At the end of his discussion of Strawson's argument in "Responsibility and the Limits of Evil," p. 140.

³⁶ Even if Strawson's argument involves the claim that *exiting interpersonal relationships* is a motivational impossibility for human beings, and even if my criticism of his argument is successful, the possibility remains that it would be motivationally impossible to *eliminate antagonistic reactions to wrongdoing*. That motivational impossibility would have to be grounded in some fact about us other than our attachment to interpersonal

as much as possible, and in light of which we appraise—or disavow—our actual responses to wrongdoing.

On this ideal of human relationships, it would be better to eliminate antagonistic reactions if we could; and even if we cannot eliminate them entirely, we should condemn and disavow them wherever they do appear. I have not defended Gandhi's and King's claims about the moral urgency of this ideal. I have established that eliminating antagonism as they demanded is compatible with remaining in engaged interpersonal relationships.

That makes arguing for the Gandhi-King ideal easier, in two ways. First, because Strawson's claim (that certain features of human life are impossible to eliminate) does not extend to antagonistic reactions, we cannot immediately rule out a universal elimination of antagonism in the way that we might immediately rule out a universal exit from interpersonal relationships. Second, my argument separates dropping antagonistic reactions from exiting interpersonal relationships. Giving up the antagonistic reactions in any individual case does not, then, require dropping the demand in that case and exiting that interpersonal relationship. The cost of dropping the antagonistic reactions to any extent is much lower than it would be if Strawson's argument did extend to antagonistic reactions.

This second point makes a significant difference to appraisal of the Gandhi-King ideal. If Strawson's argument did extend to antagonistic reactions, eliminating antagonism in any case would always involve exiting an interpersonal relationship. At the very least, that would provide us with grounds for doubting that eliminating antagonism is as commendable as Gandhi and King seem to think it is.

relationships: perhaps, for instance, some contingent fact about our neurophysiology which means that judgments about wrongdoing will always generate antagonistic attitudes.

Even though Strawson's argument is limited in the way I have described, it should inform and circumscribe the Gandhi-King ideal. The modification of our reactive attitudes that King and Gandhi recommend might be momentous and far-reaching; but it cannot involve abandoning some of the most basic aspects of human life. It cannot, for instance, involve abandoning our vulnerability to each other, as we might if we were to respond to wrongdoing with pure benevolence.

The facts as we know them may provide an adequate basis for our demanding interpersonal regard; but the facts as we know them do not thereby provide a basis for angry, hostile, antagonistic responses to wrongdoing. Our current—reactively antagonistic—form of life is not so secure.

Strawson's account fails to provide a defense of our antagonistic, resentful responses to wrongdoing. In the next chapter I turn to a different strategy for justifying antagonistic, angry resentment. This strategy involves appealing to our relationship to the victim of wrongdoing, rather than our engaged relationships with the wrongdoer.

CHAPTER 3: ANGER AND THE VICTIM

I. Essential vs. Psychological Connections

I now want to consider some alternative ways of defending angry resentment. These approaches do not appeal to our participation in engaged interpersonal relationships, as in chapter 2, but they do attempt to connect angry, antagonistic responses to other important features of our lives.

Before I begin, though, it is important to clearly distinguish two different sorts of connections that one might appeal to in defense of antagonistic response. I will use the appeal to engaged interpersonal relationships, which I have already discussed at length in chapter 2, to make the distinction clear.

In chapter 2, I discussed an attempt to connect angry resentment to an immensely valuable—indeed ineliminable—part of human life: our participation in engaged, demanding interpersonal relationships. The argument I considered aimed to show that experiencing some degree of angry resentment was *essential* to demanding goodwill of someone within such a relationship, and seeing that demand violated.

There is a different, though superficially similar, way in which these phenomena might be connected. One might argue that, as a matter of contingent human psychology (perhaps having to do with our biological makeup, or our cultural background), demanding goodwill of someone and seeing the demand violated is inseparable from experiencing angry resentment. I did not discuss this connection extensively in chapter 2. Here is why.

If the essential connection discussed in chapter 2 had obtained, we would have had a very strong defense of resentment. There I emphasized two consequences of this connection. First,

eliminating resentment altogether would entail exiting interpersonal relationships altogether—something that might not be possible, and in any event would be extraordinarily costly. Second, in any individual case in which resentment was eliminated (without excusing the wrongdoing—without showing there to have been no wrong in the first place), doing so would have to come at the expense of exiting the demanding interpersonal relationship. There would be no way to keep the demanding relationship in place in the aftermath of wrongdoing without angry resentment.

It can seem that little would be different if the connection were a matter of contingent human psychology. Suppose that there is such a connection between being in engaged, demanding interpersonal relationships and responding to wrongdoing antagonistically. Suppose further—plausibly, I think—that this sort of connection within human psychology is difficult, if not impossible, for individuals to alter or break. Then it might be true that the only route practically open to us for eliminating resentment would involve exiting interpersonal relationships altogether; and it might be true that in many individual cases in which the demand for goodwill is violated, the only practical route to eliminating antagonistic resentment would involve exiting that particular relationship.

In spite of these similarities, the essential-connection defense of antagonism is much more powerful.

One way to draw out the difference is to point out some possibilities that are left open by the contingent-psychological-connection claim. One: we could imagine some other rational agents, built somewhat like us and engaged in demanding, vulnerable interpersonal relationships as we are, but without the susceptibility to antagonistic response. These creatures would enjoy everything valuable about engaged social relationships, without the (unfortunate and problematic, if critics like Gandhi and King are correct) human quirk of susceptibility to

antagonism. Two: we might be able to exert a great deal of effort, perhaps over many generations, and alter human psychology enough to eliminate the susceptibility to antagonistic response. That alteration need not cost us our engaged, demanding relationships; it need not involve alienation or adoption of the objective attitude as Strawson warns.

If, on the other hand, antagonistic response were essential to engaged relationships, these would not be real possibilities: any beings participating in engaged, demanding interpersonal relationships would ipso facto be liable to antagonistic response.

There is another way of looking at this that may help to make the significance of an essential connection clear. If antagonism and engaged interpersonal relationships are essentially connected, criticizing antagonistic response would amount to criticizing an essential element of vitally important social engagement. If, by contrast, antagonism and interpersonal relationships are only contingently psychologically connected, we can separate the criticized antagonistic response from the engaged interpersonal relationships that seem so important.

Remember the characterization of antagonism I set out in chapter 1. Antagonistic response to wrongdoing involves taking the wrongdoer's suffering to be intrinsically valuable. It involves viewing someone's suffering as good independently of any end or goal to which that suffering might contribute. As I emphasized, this seems deeply morally problematic. If this apparently problematic antagonism turns out to be an essential part of our engaged relationships, there is something apparently morally problematic with one of the most deep-rooted, valuable, and important parts of human life. We might then view the moral concerns of Gandhi and King as answered; or we might view this as revealing an inevitable moral problem with social life. In either case, moral concerns about antagonism would turn out to be directed squarely at our engaged relationships.

By contrast, if the connection between antagonism and engaged relationships turned out to be a contingent psychological fact about humans, we could separate our moral condemnation of antagonism from our appreciation of the importance of engaged interpersonal relationships. We just have the bad luck that, given our contingent psychological makeup, those important engaged relationships are accompanied by bad antagonistic responses.

This would allow for King and Gandhi to condemn antagonistic response without thereby condemning other valuable aspects of our interpersonal lives. But if there is an essential connection here, those who condemn antagonistic response must be prepared to condemn our engaged interpersonal relationships as well. That makes the Gandhi-King condemnation of antagonism much less attractive.

In what follows I will consider further attempts to argue that antagonistic response is essential to some deeply important parts of human life. I will argue that these attempts fail. But that discussion does identify some *genuine psychological* connections. It is true that antagonism is closely connected, in our actual psychology, to some important aspects of our lives. In sections IV and V below, I will turn to consider the normative significance of these psychological connections.

My discussion in this chapter focuses on emotional, and in particular angry, responses to wrongdoing. One might doubt if there really is any difference here between essential connections and the connections that are fixed by human psychology. Emotions just are human psychological phenomena, after all; if anger is always connected to some other phenomena in human psychology, doesn't that mean that the two are essentially connected?

The answer is no. We can distinguish essential connections from connections that are merely necessary given contingent human psychology. Suppose, for instance, that human

psychology was such that we inevitably cried whenever we were sad. Even if there were such a connection—even if crying were inseparably connected, in our psychology, with sadness—crying would not be essentially connected to sadness.

II. Essential Connections: Attachment to the Victim and Anger

Begin with a phenomenon that is, plausibly, just as deep and important a part of human life as our engaged, demanding relationships with wrongdoers: our having some attachment to—some concern, care, or at minimum respect for—the victim of wrongdoing.

Consider an example introduced in chapter 1. Suppose that a driver deliberately, viciously runs down my small child on the sidewalk. It seems appropriate for me to respond angrily. That seems appropriate because of the very strong attachment I have (or ought to have) to my child. It seems plausible that weaker forms of attachment to a victim require some correspondingly weaker forms of anger.³⁷

Is an angry reaction an essential part of being attached to the victim of wrongdoing? That is, is there something about being attached to someone—loving or caring for her—that entails that one will experience anger at her being wronged? Let me explore some ways of making out such a connection.

Jeffrie Murphy writes, of response to wrongdoing done to oneself, that “proper self-respect is essentially tied to the passion of resentment, and...a person who does not resent moral injuries done to him...is almost necessarily a person lacking in self-respect” (Hampton and

³⁷ I should note that I cast a wide net here. Although I am primarily interested in antagonistic response to wrongdoing—response which involves taking the wrongdoer’s suffering to be intrinsically valuable—this section will discuss whether *any* angry responses to wrongdoing are required by attachment to the victim. I suspect that anger is always antagonistic, but I will not argue for that claim; instead, I will argue that *no* form of anger is required by our attachment to the victim of wrongdoing. If that is right, then no antagonistic anger is required either.

Murphy, 16). The self-respect of which Murphy speaks extends beyond the first-personal case, though; it just involves viewing oneself as important enough that one ought not be treated in certain ways: “not to have...resentment when our rights are violated,” he writes, “is to convey—emotionally—either that we do not think we have rights or that we do not take our rights very seriously” (17). This is “a failure to care about the moral value incarnate in my own person” (18).

Murphy appeals to respect for the victim of wrongdoing, which involves thinking that she has rights, taking those rights seriously, and caring about “the moral value incarnate in” the victim. And he suggests a way of connecting these to the emotion of resentment: such a reaction is an emotional expression of these respectful attitudes.

It does seem right that we should respect ourselves, and others, as persons with rights and moral value. Murphy does not try to spell out in detail what having rights, or being moral value incarnate, comes to; but for present purposes only two specific claims are important. First, if we think that someone has rights or moral value, when we see those rights violated or that value disrespected, we should view that violation or disrespect as seriously bad. Second, creatures capable of affect—like us³⁸—will have some affective response expressing the fact that we take the treatment to be seriously bad.

Even all this does not entail that the affective expression of viewing something as seriously bad will take an antagonistic form like angry resentment. As I pointed out in the

³⁸ That we are affective creatures may itself be a contingent fact about our psychology. So even here there may be a break in the essential connection between attachment to the victim and antagonistic, angry response to wrongdoing: angry response is only essential to attachment to the victim in creatures capable of affect. I do not want to place much weight on that disconnect, though, because our being affective creatures is itself another deeply valuable feature of our lives.

previous chapter, sadness is another affective response that can express the fact that one takes something to be seriously bad.³⁹

Attachment to the victim of wrongdoing, viewing her being wronged as seriously bad, and expressing that through some affective response: these stand or fall together. But having an *angry* affective response is not similarly connected to these phenomena.

That is not to say that *my* failure to have an angry response when my child is wronged wouldn't reveal that *I* lack an adequate attachment to my child. Given my psychological makeup, the only explanation for my failure to be angry in such a situation would be that I do not think my child has rights, or do not take them very seriously. But that is a contingent psychological connection, not an essential one. I explore the normative significance of these psychological connections beginning in section IV below.

George Sher provides a more detailed explanation of the connection to antagonism that does not allow for sadness as an acceptable alternative to angry antagonism. If we are concerned with the victim of wrongdoing and with the moral requirements violated by wrongdoing, Sher points out in his book *In Praise of Blame*, we will desire that the wrongdoing not have occurred. But the wrongdoing has occurred; there is nothing that we can do to bring it about that it did not occur. We have desires that cannot be satisfied. Sher thinks that this generates “motivational

³⁹ Appeals to the expressive function of another sort of response to wrongdoing, criminal punishment, face the same problem. In “The Expressive Function of Punishment,” Joel Feinberg argues that punishment is a way for a community to express “judgments of disapproval and reprobation.” Feinberg is not willing to say that the hard treatment is the *only* way to express this condemnation:

One can imagine an elaborate public ritual, exploiting the most trustworthy devices of religion and mystery, music and drama, to express in the most solemn way the community's condemnation of a criminal for his dastardly deed. Such a ritual might condemn so very emphatically that there could be no doubt of its genuineness, thus rendering symbolically superfluous any further hard physical treatment. ... Perhaps this is only idle fantasy; or perhaps there is more to it. The question is surely open. (115-6)

Feinberg concludes that this raises a specific problem about the justification of punishment: “the problem of justifying punishment, when it takes this form, may really be that of justifying our particular symbols of infamy” (116).

energy,” which tends to issue in negative feelings like anger.⁴⁰ Interestingly, sadness does not seem like a similar outlet for motivational energy.

But that does not mean that anger is required here. It is not even psychologically required. We do have another outlet for this frustrated motivational energy: it can be redirected toward a “related but more achievable goal” (Sher, p. 104)—for instance, attempts to make the victim whole or attempts to protect others from similar wrongs. (I might, for instance, pour the motivational energy produced by my desire that this driver not have run down my son into an attempt to have a barrier installed to protect pedestrians from cars.) Anger is one outlet for frustrated motivational energy; sadness is not an alternative outlet; but there are other nonantagonistic outlets for this energy.

Consider one final way of trying to argue that anger is essentially connected to attachment to the victim of wrongdoing. Focus on some specific motivations that are especially important in the wake of wrongdoing. If one has adequate respect and care for the victim of wrongdoing, one will act to protect him. I will suppose that this encompasses at least immediate defense of the victim, and the prevention of further wrongdoing directed at him.

If we are adequately attached to the victim, then, we will be motivated to act to protect him. In some circumstances—with some wrongdoers or potential wrongdoers who will not be stopped or deterred in any other way—that may require acting to inflict suffering on the wrongdoer.⁴¹

⁴⁰ See pp. 104-5. Notice that the situation is much different if we are able to change whether the wrongdoing occurred. I will pursue that silly-sounding suggestion in chapter 5.

⁴¹ Inflicting suffering on the wrongdoer is not quite required by proper attachment to the victim, then—the wrongdoer might stop of her own accord, and there might be no threat of further wrongdoing—but a preparedness to inflict suffering on the wrongdoing, and an ability to do so, are. If we are attached to the wrongdoer, and some

We must, then, be ready and able to inflict suffering on the wrongdoer. Is anger essential here? The answer, again, is no.

For all I have said so far, one only needs to regard the wrongdoer's suffering as instrumentally valuable to act in this way. To have a way of referring to the broader category of suffering-inflicting behavior, regardless of whether it is antagonistic or not, I will label such behavior *aggressive*.⁴² Attachment to the victim entails being ready to act aggressively to protect her, if that is necessary. A readiness to act aggressively is an essential part of adequate attachment to the victim. But anger is not an essential part of being ready to act aggressively.

Anger is one familiar, natural, and easy way of being motivated to act aggressively. But it is clearly not essential to such motivation. It is clearly possible to be motivated to inflict suffering on someone just by the recognition that doing so will contribute to some other end. That does not require anger; nor does it require viewing the suffering one inflicts as intrinsically valuable. I inflict suffering on a student when I refer a plagiarized paper to the Dean of Students for disciplinary action; but I need not be motivated by anger to bring myself to do so. I may not be angry at all; I may instead be motivated simply by, e.g., a desire to deter cheating.

I have identified several ways of connecting attachment to the victim of wrongdoing and antagonistic anger, and argued that none of these are essential connections. Being adequately

further conditions outside our control obtain, we must motivate ourselves to act to inflict suffering on the wrongdoer.

⁴² *Any* behavior that involves intentionally bringing about suffering is, on my use of the term, aggressive behavior. This is an extremely broad, and perhaps unnatural, use of the term. But it has some precedent in the psychological literature. There researchers have identified what they call "proactive aggression," which is aggressive behavior undertaken to accomplish some goal or end. Proactive aggression is not associated with subjective reports of anger or with the physiological forms of arousal that are associated with anger. In fact, this form of aggression is associated with a *decline* in arousal. In one experiment, lower measures of physiological arousal were correlated with an increased likelihood that a child would engage in proactive aggression (see "Anger and the Reactive-Proactive Aggression Distinction in Childhood and Adolescence," Hubbard et al.). (Research on proactive aggression has focused narrowly on aggression in children, although the psychologists discussing these phenomena are often willing to apply these distinctions more generally.)

attached to the victim of wrongdoing requires that one be able to motivate oneself to act aggressively. But antagonistic anger is not required to be so motivated. Being adequately attached to the victim of wrongdoing also requires that one have a (frustrated) desire that the victim not have been wronged; that will generate motivational energy, which may very well need some outlet; but it need not issue in anger. Being adequately attached to the victim of wrongdoing also requires that one view the wrongdoing as seriously bad; that will be expressed with some sort of affect, but it need not be expressed through anger in particular.

All these ways of trying to argue that anger is an essential part of being attached to the victim of wrongdoing fail. In the next section I turn to a different, indirect way of appealing to essential connections to defend anger.

III. Essential Connections: Sadness and Attachment to the Victim

At several crucial points in the past two chapters, I have drawn on an alternative to anger: a form of sadness that, I argued, can fill several of the roles that appear to require angry resentment. That alternative was central to my argument in chapter 2; I have also drawn upon it in this chapter.

But sadness seems to be in tension with some of the important motivational traits that I have just discussed. Sadness seems *demotivational*. Being sad seems to be in tension with being motivated to act aggressively at all.

It is important to be clear about the potential problem here. The mere fact that sadness does not motivate us to act aggressively is not a problem. As I have pointed out, we do have other, nonantagonistic ways of so motivating ourselves. If sadness does not motivate aggression,

our saddened appreciation of the victim's plight can prompt us to adopt the end of protecting her, and thereby motivate us to act aggressively.

But sadness doesn't seem to be just neutral with respect to motivating aggression. It seems to be connected with a *depression* of motivation, and in particular a depression in our motivation to act aggressively or assertively to confront problems and pursue our ends. This is particularly clear in the case of serious forms of sadness like grief and depression. Deep grief or depression is often characterized by a withdrawal from the world, or otherwise experiencing some decrease in motivation.⁴³

It might be essential to sadness that it get in the way of aggressive motivations. If so, the moralized sadness that I described in chapter 1 is essentially opposed to motivations that are required by adequate attachment to the victim of wrongdoing. That opposition may not mean that we are unable to act as required; it is possible to muster motivational resources to overcome these sorts of obstacles. But it does mean that substituting sadness for anger in the way discussed in chapter 1—and so avoiding antagonism—has the cost of making it more difficult to be motivated in the way required by adequate attachment to the victim of wrongdoing.

This would at least complicate our appraisal of reactive anger and reactive sadness: anger provides us with a way of preserving relationships with the wrongdoer and adequately responding to the victim's value, at the price of being antagonistic; sadness provides us with a

⁴³ I am extremely skeptical about untutored philosophers' appeals to psychiatric diagnostic tools to support their claims. Nonetheless, I will point out that this broad connection does seem to be reflected in the DSM-IV. Among the criteria for the presence of a "Major Depressive Episode" (i.e., nine criteria are listed, and five must be present for the diagnosis) are "markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day;" "fatigue or loss of energy nearly every day;" and "diminished ability to think or concentrate, or indecisiveness, nearly every day." Among the criteria for a diagnosis of "Chronic Depressive Disorder" are "low energy or fatigue" and "poor concentration or difficulty making decisions."

But, again, these are diagnostically useful criteria. They may be extreme forms of a motivational inhibition that is essential to sadness; or they may be reflections of some contingent connections in human psychology.

way of preserving relationships with the wrongdoer without being antagonistic, at the price of presenting some opposition to adequately valuing the victim.

Is sadness essentially in tension with being motivated to resist wrongdoing? It takes some work to see exactly where the tension might be located.

Sadness first of all seems to involve a general suppression of motivation. Perhaps sadness is in tension with the motivation to act aggressively because it essentially involves the suppression of *all* motivations. But this is not essential to sadness. When asked, people do indicate that there are some behaviors to which sadness inclines them: to spend time alone, to listen to music, and to sit and think, for instance.⁴⁴ These activities may not involve high levels of arousal, but they are behaviors which sadness can motivate. Sadness *can* create rather than inhibit motivations. Doing so is not in tension with what is essential to sadness: we do not think that someone whose sadness motivates her to act in these ways is sad in an unusual, perplexing, or paradoxical way.

But these behaviors do all involve a sort of disengagement or withdrawal. They do not involve attempting to engage with whatever state of affairs made one sad in the first place. Perhaps sadness is opposed to aggressive motivation because it essentially involves either a lack of motivation, or a motivation to withdraw and disengage—either of which are in tension with acting aggressively to alter or address the state of affairs that makes one sad.

This is not quite right, either. Sadness does sometimes generate motivations to engage with the external world, in ways designed to address the state of affairs that makes one sad.

⁴⁴ These examples are the top 3 answers given by college students when asked to select, from a list, those activities “that they would be most likely to perform if they felt depressed or sad” (Cunningham, p. 316; see also p. 320, for the table of activities).

Sadness does sometimes motivate one to seek help addressing what makes one sad. Grief, for instance, often motivates comfort- and reassurance-seeking behaviors.⁴⁵

Of course, seeking comfort, or help changing some bad situation, is not the same as acting aggressively—inflicting suffering on someone—in order to change it. Sadness might be compatible with some motivations to engage with the world in order to address what makes one sad, but in tension with motivations to do so by inflicting suffering. That is, what is essential to sadness might be in conflict with motivations to act aggressively, in particular.

I don't think that is correct either, but a confounding factor here makes it a difficult thesis to evaluate. Anger, I have pointed out, does tend to generate motivations to inflict suffering on others. And in many cases we might look at, to try to determine whether sadness makes aggressive motivation more difficult, the natural alternative or contrast to being sad is being angry. So being motivated to inflict suffering will seem more difficult when we contemplate changing from anger to sadness; but it will be difficult to determine whether that is because something essential to sadness is in tension with aggression, or due to the absence of aggressive-action-promoting anger.

To control for this, consider a pair of cases, one of which involves sadness and the other of which involves no affect. Considering these two cases will make clear that what is essential to sadness is not in tension with motivations to inflict suffering at all.

⁴⁵ See “‘Sadness’—Is There Such a Thing?” (Barr-Zisowitz 611): “A third issue with which North American psychologists have grappled, in the effort to determine a function for sadness, is the seeming contradiction between the sad person’s tendency to turn inward in order to find solace or solve problems, and the tendency to turn outward, cueing others that he or she needs help.” (And notice that even the turning-inwards behavior is described as having problem-solving potential.)

Grief in particular often involves help-seeking behavior, as Barr-Zisowitz notes (although grief is generally experienced in response to a loss which cannot be reversed: the person grieving is seeking help in coping with the loss, not in acting to change it). Barr-Zisowitz also argues that, in the 17th and 18th century in England and North America, attitudes regarding help-seeking sadness went from acceptance to distaste. She thinks that “the later effort to maintain good cheer seems to have coincided with the notion that one had to take care of one’s own difficulties rather than lean on external help” (612-613).

Consider a parent's response in two different scenarios:

(A) I am sad that my child's performance on his homework is not as good as it could be.

To make this a natural case of sadness, not anger, given our actual psychology, suppose that I do not blame my child for this: I just think that his performance could be better, and I am sad that things are not going as well as possible for him. With the right motivation, he would work even harder and do even better on his homework. I inflict suffering on him to provide such a motivation: I take away his internet and texting privileges until his performance improves.

(B) In the second case, I make the very same appraisal of my child's performance; but I am not sad or upset that he isn't doing as well as possible. I just dispassionately judge that he could be doing better. I also recognize that, with the right motivation, he would work even harder and do even better on his homework. So I inflict suffering on him to provide such a motivation: I take away his internet and texting privileges until his performance improves.

In both cases I have the same reason to take the aggressive, suffering-inflicting step of taking away my son's texting and internet privileges. The only difference is that I am sad in scenario (A). Does the addition of the sadness make it more difficult to act in that way? It does not seem to. It is no more difficult to act to inflict suffering on my son in scenario (A) than in scenario (B).

I conclude, then, that it is not essential to sadness that it be opposed to motivations to inflict suffering on others: sadness is not opposed to that specific motivation, and it is not opposed to motivation more generally.

That is not to say that there are no connections between sadness and depressed motivation. Sad people often do withdraw from the world and experience a depression of motivation. But that is a contingent fact about human psychology. It is not essential to sadness

that it have such an effect. Sadness does not necessarily throw up obstacles to being motivated to act aggressively.

IV. Psychological Connections: Defending Anger

To this point in the dissertation I have focused on essential connections. In chapter 2 I argued that antagonistic resentment is not essential to our engaged relationship with a wrongdoer. In this chapter, I have argued that anger is not essential to an adequate attachment to the victim of wrongdoing, and I have argued that there is no essential tension between being saddened by wrongdoing and being motivated to respond aggressively to it.

As I indicated at the beginning of this chapter, I focus on these essential connections because if they obtained, they would provide a very strong defense of antagonistic moral response. If such connections did obtain, Gandhi and King's critique of antagonistic reactions would be aimed at an essential aspect of deeply valuable parts of human life. (Or, if there were an essential tension between sadness and aggressive motivation, the alternative to antagonism would itself be problematically in tension with an important part of human life.)

The connections between anger and attachment to the victim that I have identified—that anger is an expression of that attachment; that it motivates aggressive response required by that attachment—are not essential connections. But there may be genuine psychological connections here. Now I want to consider whether psychological connections like these might provide any defense of antagonism. They will not ground the strong defense that an essential connection would; but they may nonetheless provide for *some* defense of antagonistic response.

Consider, for instance, the fact that anger is often an expression of one's care and concern for the victim of wrongdoing. That we express this sort of care in these sorts of circumstances

with anger is, for present purposes,⁴⁶ nothing more than a brute fact of human psychology. Doesn't the fact that this sort of anger is expressing a morally good attachment provide for *some* sort of justification of it? This anger is performing a morally important function, after all. And given other facts about our psychology and physiology, anger is a readily apparent, widely recognized way of expressing this attachment. The speed with which we become angry, and the high level of arousal involved, are also often difficult to either replicate intentionally or fake; so this is a particularly immediate, and particularly sincere, way of expressing attachment to the victim.

To avoid getting caught up in the specifics of human psychology, physiology, and anthropology, which are surely more subtle and complex than this, suppose that the strongest versions of these connections obtain. Suppose that anger is the easiest-to-recognize, most-widely-recognized, quickest, and most-difficult-to-fake way of expressing attachment to the victim of wrongdoing. Also suppose—plausibly, I think—that this expressive anger, directed at a person one holds responsible for harming someone one cares about, is antagonistic: that it involves some disposition to value the wrongdoer's suffering apart from any good ends it might promote. Wouldn't that provide for *some* sort of justification of an angry, antagonistic response to wrongdoing?

⁴⁶ The "for present purposes" qualification is important. This connection likely has an evolutionary explanation, for instance; it is not entirely accidental. But that evolutionary explanation won't provide a justification for the response.

The evolutionary explanation will likely involve emphasizing some of the roles that antagonistic, angry response plays in social interactions—which is just what I am examining directly in the text. Contemporary evolutionary psychology has made progress addressing certain very circumscribed questions about human interaction—primarily, questions about what response strategies generate stable outcomes in iterated prisoner's dilemma interactions, and questions about what response strategies people actually use in such interactions. This research has generated insight into the way certain human behaviors may serve to solve the problems that arise when we interact in certain specific ways. Its significance outside of iterated prisoner's dilemma interactions, and the normative significance of the fact that certain strategies are successful in the long run within such circumstances, are not yet clear.

If we were dealing with a mode of expressing this valuable attachment that were minimally morally problematic, these advantages might help to justify the response. If we were worried about antagonistic anger just because it creates some discomfort for others, or because it seems occasionally disruptive, these advantages might counterbalance those concerns.

But these advantages need to be set alongside the deeply morally problematic features of antagonism. Antagonism involves taking its object's suffering to be intrinsically valuable: valuable in itself, without consideration of any further good. Antagonism essentially involves viewing suffering as valuable for its own sake. That may not involve actually acting to make the object of angry antagonism suffer; but it does involve wishing for such suffering, or being inclined to welcome such suffering without an eye to whether or not it might serve any use or purpose.

This stance towards another's suffering seems—to King and Gandhi, and to many others attracted to the Gandhi-King ideal—deeply morally problematic. I must rest with an appeal to intuition here: it does not seem that this problematic feature of antagonistic anger will be overcome by its advantages in terms of sincerity and the like. Such concerns are, morally speaking, much less compelling than the problems with antagonism.

V. Psychological Connections: Violence Generates Antagonism

Antagonism is also particularly closely connected to being motivated to inflict suffering on others. I have argued that anger is not an essential part of being motivated to act aggressively. And even in our actual psychology, such motivations do not require anger, or any other sort of

antagonism. Humans sometimes do motivate themselves to inflict suffering on others just by recognizing—nonantagonistically—that the suffering would be instrumentally valuable.⁴⁷

But there are some close connections between antagonism and acting aggressively in our contingent psychology. I want to focus on one that is of particular practical significance.

When discussing King's and Gandhi's absolute, categorical rejection of violence, I suggested that they might think that overt violence inevitably leads to antagonism. This is suggested by some of King's remarks in his Nobel Lecture. Even though he is willing to create conflict and tension, King rejects violence. Calling it destructive, he says that it "creates bitterness in the survivors and brutality in the destroyers."

This is a different connection from the ones that I have so far discussed. Instead of worrying that morally problematic attitudes are required to motivate violence, King suggests that such attitudes might be the *result* of violence. The direction of causation is reversed. It is possible to motivate oneself to act aggressively just by recognizing some good end to which the aggression will lead. But, King worries, violent aggression has the potential to *generate* morally problematic attitudes like antagonism.

And yet violence sometimes seems like the only way to respond adequately to wrongdoing. Orwell makes the case in his discussion of Gandhi:

Without a free press and the right of assembly, it is impossible not merely to appeal to outside opinion, but to bring a mass movement into being, or even to make your intentions known to your adversary. Is there a Gandhi in Russia at this moment? And if there is, what is he accomplishing? (469)

This is a common point about the limits of nonviolence. King and Gandhi both strive to argue that, in the end, violence is self-defeating. If they are right, then violence is just as limited. I am dubious that this is true; in any event, I will suppose that there are some circumstances (e.g.,

⁴⁷ I argued that viewing suffering as instrumentally valuable is not antagonistic in chapter 1.

resistance to Hitler or Stalin) in which (a) it is morally important to at least *try* to resist oppression, and (b) the only strategy that has any hope of succeeding involves violence.⁴⁸

If there is a causal connection in human psychology between violence and antagonism, we may be faced with a difficult choice. The deeply morally problematic antagonism that will arise in the wake of violent resistance is not inevitable; we could avoid it by avoiding violence. But an adequate response to certain sorts of wrongdoing—totalitarian oppression, for instance— involves violence that will give rise to such antagonism. It was immensely important to act to resist Stalin, and Hitler, even if doing so required antagonism-generating violence.⁴⁹

In the face of this dilemma, it is important to examine the connection between violence and antagonism more closely. This is an empirical issue: to what extent does violence really tend to generate antagonism? It does not seem to be true, for instance, that spanking a child tends to create brutality or antagonism in the parent. But resistance to Hitler and Stalin involves much more serious and intense violence. It is more plausible that those who engage in such violence will tend to experience antagonistic attitudes.

I will focus on an extreme, but salient, form of violence. Soldiers who kill in warfare do tend to experience a combination—sometimes, an unstable combination—of antagonistic and nonantagonistic attitudes. It is worth looking carefully at these actual reactions. On a cartoonish

⁴⁸ Interestingly, at the beginning of the portion of his Nobel Lecture cited above, in which King asserts that violence is “both impractical and immoral” (and argues that violence is self-defeating), he explicitly limits himself to “violence *as a way of achieving racial justice*” (emphasis added).

⁴⁹ Orwell reports that Gandhi was unwilling to endorse violent resistance to Hitler, suggesting instead that the Jews should have committed collective suicide, which “would have aroused the world and the people of Germany to the evils of Hitler’s violence” (Fischer 50; Orwell quotes this passage at p. 468 of “Reflections on Gandhi”). But the actual recommendation comes in a slightly different, and more complex, context. Gandhi does in fact recommend collective suicide, but he contrasts it with actual Jewish acquiescence to the Holocaust and a characterization of his view as recommending “passive submission” to Hitler. While Gandhi clearly does not view violent resistance to Hitler as a possibility, his suggestion is meant to emphasize that he recommends an assertive, active refusal to submit.

version of King's concern, warfare is inevitably morally corrosive, leading to racism, hatred and the like directed at the opposing side. That cartoonish picture is not quite correct, but there is an important causal relationship between killing in war and certain specific forms of antagonism.

In *On Killing*, Dave Grossman distinguishes several phases often experienced by soldiers after killing in combat.⁵⁰ Grossman reports that modern training techniques (which involve what is, he argues, a form of operant conditioning) often ensure that in combat a soldier will kill relatively automatically, with little affect or deliberation. Soldiers do, however, characteristically go through a progression of responses after killing, which Grossman divides into the stages of exhilaration, remorse, and rationalization.⁵¹

The final stage, rationalizing the fact that one has killed, can involve cultivating antagonistic attitudes (racist hatred, for instance) that diminish the humanity of the enemy who has been killed. This may fit with the cartoonish version of King's worry, on which killing in war has a corrupting influence that leads a soldier to take a dehumanizing view of the enemy.⁵² But, importantly, rationalization need not take antagonistic forms. Identifying some goals furthered by the killing—immediate goals of self-defense, or long-term goals involving the purpose of the conflict—can also provide rationalizations. These rationalizations work precisely by enabling the soldier to think of the killing as something done for a reason rather than for its

⁵⁰ See Grossman, pp. 234-241. Grossman is a professor of Military Science and former Army Ranger with a Master's degree in counseling psychology.

⁵¹ Grossman compares this progression to psychological models of the characteristic stages of grief, and emphasizes that individual experiences vary widely; individuals may skip or pass extremely quickly through any of these stages. He also argues that disorders like PTSD can often be explained in terms of a failure to complete the rationalization stage.

⁵² This is distinct from (though it surely bleeds into) the racist, dehumanizing attitudes that are often inculcated in soldiers to make them more likely to kill in the first place. Such attitudes are instances of a phenomenon I discussed in section II: they are antagonistic attitudes that make it easier to be motivated to act aggressively. Grossman argues that this is only one of many techniques (not all of which are antagonistic) that can increase the psychological distance between the soldier and enemy, to make it easier to violently kill the enemy. Once again, antagonism may be a useful means to an end, but is not required.

own sake. Social and institutional support has a strong influence on this rationalization process. Thus it is not only possible to go through rationalization without resorting to antagonism; it is possible to design social and institutional structures which encourage that.

The exhilaration phase is more troubling. Grossman uses ‘exhilaration’ to name an experience of satisfaction that follows a successful kill; it is more often, or at least more openly, reported by those who kill at a longer range, especially fighter pilots, and it is distinct from the adrenaline rush of participating in combat generally. It is illustrated by some of the reports that Grossman quotes:

“Once you’ve shot down two or three [planes] the effect is terrific and you’ll go on till you’re killed. It’s love of the sport rather than sense of duty that makes you go on.” [A British fighter pilot, World War II]

“Christ! He’s coming to pieces, there are bits flying off everywhere. Boy! What a sight!” [A World War II fighter pilot]

“I suppose it is brutal...but I had a felling of the most intense satisfaction as the wretched Turk went spinning down.” [A British field marshal, World War I]

“The excitement was just fantastic...the exhilaration, after all the years of training, the tremendous feeling of lift, of excitement, of exhilaration, it was like the first time you go deer hunting.” [An American tank commander in World War II]

“Well-intentioned souls now offer me their sympathy and tell me how horrible it all must have been. The fact is, it was fun. ...granted I may be looking back through rose-colored glasses. *But it was great fun* [Anderson’s emphasis]. ...Only a veteran can know about the thrill of the kill...” [Letter to the editor in *Soldier of Fortune*, R. B. Anderson, American squad leader, Vietnam]

...when he killed enemy soldiers who were “coming over the wire” in a fire-fight, he felt what he called “satisfaction, a satisfaction of anger.” [An American veteran of Vietnam]⁵³

This exhilaration or satisfaction may have nonantagonistic elements. Removing a threat brings relief; completing a challenging task brings satisfaction. But others tasks in combat involve these

⁵³ See *On Killing* pp. 236-8 for these quotes.

elements and do not bring the same reported response. The act of killing itself, apart from any instrumental value it might have, is sometimes experienced as satisfying. That appears to be a psychological fact about human response to killing.

This exhilaration might itself be useful in motivating those who experience it: witness the fighter pilot's report that "the effect is terrific, and you'll go on till you're killed." But more often—reassuringly—it tends to be immediate, short-lived, and followed by remorse and a need to rationalize. It is a limited and contained form of antagonism. In most people it is a transient reaction to what one has done, not an enduring disposition to value and bring about more death. That is reassuring: this antagonism need not be socially destructive or dangerous. But it is still seriously antagonistic, in a way that does seem morally problematic—as indicated by soldiers' guilt and need to rationalize what they have done. It is still pleasure or satisfaction at having ended a life.

This form of antagonism may, in fact, have the most serious consequences for the soldier who experiences it. Psychological research into the effects of killing is limited, but there is mounting evidence that having killed contributes to psychological distress over and above the contribution of having experienced life-threatening or otherwise traumatic combat situations. There appear to be associations between having killed in combat and experiencing PTSD symptoms.⁵⁴ In addition, one study found that veterans reporting killing experiences "were twice as likely to report suicidal ideation as those who did not kill, even after accounting for PTSD, depression, and substance use disorder diagnoses."⁵⁵ Another found that "Vietnam veterans who

⁵⁴ See Maguen et al., "Killing in combat, mental health symptoms, and suicidal ideation in Iraq war veterans." Similar associations have been reported in police officers who kill in the line of duty; see Komarovskaya et al.

⁵⁵ Maguen et al., "Killing in Combat May be Independently Associated with Suicidal Ideation," p. 5.

killed and experienced light combat had more PTSD symptoms than those who did not kill and experienced heavy combat.”⁵⁶

This research does not isolate the impact of experiencing exhilaration at having killed. Grossman does, however, believe that exhilaration is important here. He reports that “the response of veterans to the killing response stages—particularly to the interaction between the exhilaration and the remorse stages—is the most powerful I have ever experienced...when the remorse stage sets in they believe that there must be something “wrong” or “sick” about them to have enjoyed it so intensely” (336).

King warns about creating “brutality” in those who use violence. The cartoonish worry that engaging in violence is inevitably, permanently corrupting is not accurate. But killing in warfare does tend to elicit antagonistic reactions—reactions that appear deeply troubling to those who experience them. If antagonism is morally problematic, those reactions are morally problematic as well.

And yet violence—military violence, in particular—does sometimes seem morally required. We may, then, be morally required to act in ways that will create antagonistic exhilaration, even if the antagonism itself is of no instrumental value, and even though it could be avoided.

VI. More Attempts to Defend Antagonism

⁵⁶ Cited in Litz et al., p. 697.

I want to finish this chapter by considering some additional attempts to defend antagonistic anger. This section will inevitably have a piecemeal quality, as I want to deal with a number of possibilities none of which require sustained discussion.

First consider a broader psychological fact about humans. It is likely that it will not be psychologically possible to eliminate antagonistic response to wrongdoing altogether. Some such reactions are likely to be outside our ability to control or eliminate.

Suppose—though I will not argue for this empirical claim—that it is not psychologically open to us to eliminate antagonism in favor of the nonantagonistic alternatives that I have described. On top of that, as I have argued in section V, we might be morally required to act in ways that generate yet more antagonism. What, then, is the practical significance of all this exploration of inaccessible alternatives?

In a different context, while describing an ideal theory of punishment, Herbert Morris acknowledges that:

...the theory is, of course, not intended as a description of any actual practice of legal punishment or even as realistically workable in a society such as ours. ...At this point it may be thought, 'fair enough, but then what is the point of the whole exercise?' ("A Paternalistic Theory of Punishment," 269)

Morris answers that his theory "has value, for it provides an important perspective upon actual practices; it throws into relief our society's failure to realize the conditions" that would provide for a just practice of punishment. Part of Morris's point is that this perspective can help to guide social reform. But the perspective itself—appreciating what is and is not of value in our society, regardless of our ability to act in light of that knowledge—seems valuable. It seems similarly valuable to recognize the ways in which our natural human constitution connects valuable and morally problematic elements, even if that is not something that we can change.

Next consider the self-reactive attitudes: the reactions one has to oneself, when one has done wrong. Merely being disappointed with oneself seems problematically self-centered or self-indulgent. If I have betrayed my friend and I am simply disappointed with myself for not doing better, my friend could quite plausibly be upset with that fact. Shouldn't *she* have a more central place in my reaction to what I have done?

I think this is an instance of the importance of being concerned with the victim of wrongdoing: being disappointed with oneself seems self-centered, because it is focused on the wrongdoer (in this case, oneself), and so does not seem to express adequate care and concern for the person who was wronged. But my earlier discussion applies here as well. As I argued in section II, sadness is not *required* to express this attachment. And, as I indicated in section IV, while antagonistic anger might have some advantages as a way of expressing this attachment, those advantages do not seem to be adequate to overcome what seems so problematic about antagonism: that it involves taking its object's suffering (in the self-reactive case, one's own suffering) to be intrinsically, noninstrumentally valuable.

I think that the motivational concerns previously discussed arise in the self-reactive case as well, although the relevant motivations are a bit different here. When responding to one's own wrongdoing it is important to be motivated to act in a variety of ways, which may include exercising increased self-restraint; making amends; and working to reform oneself. (These motivations seem to be connected to the importance of caring for the victim of wrongdoing; they also seem to be connected to the importance of respecting and caring about the moral law or requirement that one has violated.) It may be that anger directed at oneself sometimes makes it easier to be motivated in these ways. But, once again, it does not seem that angry antagonism is

required: there are other ways of generating these motivations. (In addition, being saddened or disappointed in oneself is not in conflict with being motivated in these ways.)

I am tempted to add a bit more here. Being disappointed with oneself can be very motivating indeed. In the moral case, perhaps, this form of disappointment is an instance of what is often labeled *shame*. Perhaps disappointed shame can be contrasted with angry *guilt* at one's moral failings. Both of these emotions seem powerfully motivational.⁵⁷ I cannot explore these connections in depth here, though, so I will set them to one side and rest with the claim that what has been said about anger and motivation in the third-personal case also applies in the first-personal case.

Finally, think about the sadness that I have described. In the previous chapter it was disappointed sadness at the wrongdoer's having violated some demand. But here I have described it as sadness at the victim's having been wronged. There seem to be two emotion-tokens here: sadness generated by, and directed at, the wrongdoer; and sadness generated by, and directed at, the victim's being wronged.

Resentment or indignation (to use Strawson's name for the third-personal analogue of resentment) seems, by contrast, capable of unifying our reactions to the wrongdoer and the victim. My angry resentment or indignation at wrongdoing seems capable of expressing *both* my demand of the wrongdoer *and* my care for the victim.

⁵⁷ I make these connections tentatively because I cannot do justice here to the rich philosophical, historical, and anthropological discussions of guilt, shame, and their relationship. I will, however, note that shame is often discussed in the context of the debate over alternative sanctions and restorative justice. In that literature, the alternatives to traditional punishment that are proposed are sometimes said to involve forms of shame. (See especially the work of John Braithwaite, e.g. *Crime, Shame and Reintegration* and "Repentance Rituals and Restorative Justice.") If my tentative claim—that guilt involves anger where shame involves sadness or disappointment—is correct, then my claims about saddened disappointment and angry reactions to wrongdoing ought to be explored in the context of those alternative sanctions. Again, I cannot do that here. And much care is required to avoid flattening important distinctions here: resentment, anger, and guilt are not simply identical; nor are disappointment, sadness, and shame, though they are closely linked.

If we look more carefully at the way that resentful anger is directed at both the wrongdoer and the victim, however, I think it will become apparent that the parallel form of sadness I have described can work in the same way.

Resentful anger is directed at the wrongdoer in the sense that it is generated by her failure to live up to some demand I make of her. It is directed at the wrongdoer because the demand is directed at her. And resentful anger is directed at the victim in the sense that the wrongdoer falls short of this demand by mistreating the victim. The *demand* does the unifying: the demand is *addressed to the wrongdoer*, and it is issued *on behalf of the victim*. When the wrongdoer fails to satisfy the demand, that failure is a fact about her and a fact about how the victim has been treated.

The form of sadness that I have described involves the very same demand, in the same way. My disappointment focuses on the wrongdoer, in that I am disappointed by her failure to do better; and it focuses on the victim, in that I am disappointed by the fact that she has not been treated better.

VII. Conclusion

In this chapter I have pursued a variety of attempts to connect antagonistic, angry responses to wrongdoing with our valuable attachments to the victim of wrongdoing. I have been flexible about exactly what is involved in being attached to the victim, and how we might make a connection between that attachment and antagonistic response. That is because I wanted to see if *any* such connection can be made.

My main conclusion so far is this: I can find no essential connection between attachment to the victim of wrongdoing and angry, antagonistic response to that wrongdoing. In addition, as

I argued in chapter 2, there is no essential connection between our engaged, demanding interpersonal relationships and antagonistic response to wrongdoing. I can find no reason why some beings who are rational, emotional, social, and vulnerable like us could not have all these valuable features of human life, without any antagonistic response to wrongdoing whatsoever. Neither our engaged, demanding interpersonal relationships, nor our care, concern, and respect for each other when we are wronged, requires antagonistic response. There do seem to be *contingent psychological* reasons for thinking that it may be impossible *for us human beings* to realize this state of affairs; but the deeply valuable features of our lives do not themselves rule this out.

This is an argument by elimination. If I have overlooked a way in which our engaged relationships or our attachment to the victim of wrongdoing does require angry, antagonistic response, I would like to hear about it. It is *strange* to conclude that such reactions—which seem so deeply embedded in, and essential to, our social lives—are in reality just a contingent accompaniment to what is really of value. But so far as I can see, there is nothing about our valuable attachments that requires angry antagonism.

When we begin with the intuitive familiarity and appeal of antagonistic reactions, this evokes perplexity. But when we begin from the King-Gandhi ideal and its suspicions about antagonism, this can be an inspiring prospect, one that promises to free us from our attachment to morally troubling antagonism without asking us to renounce our human attachments. Pointing out that Gandhi rejected “close friendships” and “exclusive loves,” Orwell worried about the “other-worldly, anti-humanist tendency” of Gandhi’s doctrines (466). It may be true that, given our psychological makeup, attempting to live the Gandhi-King ideal to the extent that Gandhi did

requires suppressing or eliminating our attachments to others. But that does not mean, as Orwell thought, that the Gandhi-King ideal is, essentially, a rejection of the interpersonal relationships and attachments that make human lives valuable.

The way that antagonism is connected, in our psychology, to interpersonal relationships and attachments may, however, mean that pursuing Gandhi's path is too costly for us. Those contingent psychological connections may mean that the cost of a serious attempt to eliminate antagonism would be unacceptably high.

In fact, as I argued in section V above, the situation may be worse than that. There appear to be causal psychological connections between violence and antagonistic attitudes. Killing in warfare appears to generate an antagonistic enjoyment of the killing itself. And yet it also seems that we are sometimes morally required to respond to wrongdoing with violence. If all that is correct, then we are sometimes morally required to respond to wrongdoing in ways that will bring into existence antagonistic attitudes or emotions that we would not otherwise experience.

Perhaps it is not yet time to give up on justifying antagonism. Even if antagonistic anger is not a required part of our interpersonal relationships and attachments, some form of antagonism might be morally important in some other way. In some circumstances, an antagonistic response to a wrongdoer might be the only way to accomplish some morally important end. In the next two chapters I explore this approach to justifying antagonism.

CHAPTER 4: RESTORATION OF RELATIONSHIPS

I turn next to *restoration*. It is clear that wrongdoing often involves the destruction of something of value. Murder ends a life; violent assault can destroy or permanently damage parts of a victim's body; a victim of theft loses property; and so on. It is equally clear that, to the extent doing so is possible, it is good to restore whatever valuable thing or state of affairs has been lost. It is good to restore a theft victim's property, for instance. In cases where complete restoration is not possible—for instance, where a victim's eye has been destroyed in a violent assault—we aim to do what seems like the next best thing: we aim to provide *compensation*, so that the victim's overall level of well-being is restored to what it was before the assault.

I will take these general claims about the importance of restorative response (including the claim that compensation is a form of restoration) for granted. I want to examine whether restoration requires antagonistic response to wrongdoing.

I

I will begin by exploring briefly some appeals to restoration that do not involve antagonistic response. These appeals to restoration may very well provide strong support for responding to wrongdoing in certain ways; they may even support responding to wrongdoing by inflicting some suffering on the wrongdoer. But they do not support responding with antagonism; they do not support taking the wrongdoer's suffering to be intrinsically, non-instrumentally valuable.

Consider first some very straightforward attempts to restore the victim to her pre-wrongdoing level of well-being. This is often accomplished in a way that makes the wrongdoer

worse off. We might, for instance, take stolen property from the wrongdoer and return it to the victim of wrongdoing. In other cases, where restoring exactly what has been destroyed is impossible or extremely difficult, we might take resources from the wrongdoer in order to provide the victim with compensation. It is, obviously, likely that the wrongdoer will find either of these losses of resources to be unpleasant.⁵⁸

But this way of treating a wrongdoer is not antagonistic. The suffering we impose on the wrongdoer is not intrinsically valuable; it is, instead, instrumentally valuable. It is valuable as a way of restoring the victim's well-being. Any account on which the wrongdoer's suffering helps to bring about restoration will be like this: even if such an account provides good reasons for making the wrongdoer suffer, it will be nonantagonistic, because it will make that suffering out to be instrumentally, not intrinsically, valuable. It looks like restorative accounts are badly positioned to justify antagonistic response.

This is not the end of the road for a restorative justification for antagonistic response. It does, however, emphasize that we will have to pay careful attention to what accomplishes the restoration. On the account just suggested, taking some resources or stolen property from the wrongdoer—making the wrongdoer worse off in some way—restores the victim to her pre-wrongdoing level of well-being. That makes the wrongdoer's suffering merely instrumentally valuable.

Consider a different restorative account: one on which certain responses to wrongdoing provide the victim with *satisfaction*. A victim of wrongdoing often has a variety of desires about what will happen to the person who has wronged her. A victim might desire to express

⁵⁸ I am casting a wide net here; the responses I sketch are often examined in discussions of torts or reparations. And the responses I discuss later are often associated with criminal punishment. I cast my net this widely because my focus is on antagonistic response in general; I want to know if any of these forms of restoration can justify any form of antagonistic response.

antagonistic attitudes like hatred and malice at the wrongdoer. She might be aggravated and discontented until she does so; satisfying those desires might extinguish them and restore her to a calmer state of mind.

Suppose that satisfying the victim's desires will restore her peaceful state of mind. In this case—given the specific desires that the victim has—doing so requires that she express antagonistic attitudes at the wrongdoer. The desire in question is a desire for antagonistic response. Only antagonism will satisfy the desire in question; only antagonism will bring about a restoration of the victim's peaceful state of mind.

Consideration of the victim's desires might provide some support for such a response. But it is doubtful that this will be sufficient to render seriously morally problematic responses morally acceptable. A victim of wrongdoing might, for instance, desire that the wrongdoer's neighborhood be razed to the ground. The fact that doing so would restore the victim's peaceful state of mind does not mean that we are justified in doing something so seriously morally problematic. As I emphasized in chapter 1, antagonistic response also seems deeply morally problematic; so it seems similarly problematic to justify such response by saying that it will restore the victim's peaceful state of mind.

Of course, if I were successful elsewhere in this dissertation—if I could identify some other reason for thinking that antagonistic response to wrongdoing is morally acceptable—things would be different. Then the fact that this antagonistic response would also provide the victim with satisfaction would be an additional consideration counting in favor of the antagonistic

response. But this consideration itself does not seem to be sufficient to overcome the serious moral qualms about antagonistic response.⁵⁹

II

I think that antagonistic response to wrongdoing can be restorative in a different way. To describe this, I need to draw materials from some other accounts of restoration in other contexts. In that vein, I turn next to an extremely influential appeal to restoration: the account of criminal punishment that Herbert Morris sets out in his influential paper “Persons and Punishment.” Morris describes a society in which each citizen takes on a burden of self-restraint in order to obtain the benefits that come when others exercise similar self-restraint. In such a society, Morris writes, “if a person fails to exercise self-restraint even though he might have and gives in to such inclinations, he renounces a burden which others have voluntarily assumed and thus gains an advantage which others, who have restrained themselves, do not possess” (477). Criminal wrongdoing in such a society gives the criminal an unfair advantage.

⁵⁹ There is another restorative strategy that might be more successful, which I cannot explore in depth here. Once again I am making reference in passing to the “restorative justice” movement promoted by John Braithwaite and others. This movement—which I mentioned above (chapter 3, n. 57) for its discussion of shame—proposes replacing traditional forms of criminal punishment with alternatives involving what Braithwaite calls “reintegrative shaming” and “repentance rituals” (see Braithwaite, “Repentance Rituals and Restorative Justice”). These interactions are restorative: they are intended to restore relationships, and restore the wrongdoer into society.

The shaming and rituals may very well be unpleasant for the wrongdoer. But once again much of the wrongdoer’s suffering in such interactions is likely to be merely instrumentally good. It is a means to bringing about reintegration and the restoration of damaged community relationships. So this theory does not seem to propose antagonistic response to wrongdoing.

But it is not entirely implausible that a certain form of suffering is itself good in these types of programs. It might be *essential* to the good of recognizing that one has done wrong that one be pained at having done wrong. (This is suggested by Morris in “A Paternalistic Theory of Punishment;” see also his “Guilt and Suffering.”) The suffering might, plausibly, be *an essential part of* the recognition *I have done wrong*. In that case the suffering is not a means to a good end, or a side-effect of something good; the suffering is itself good. Bringing this about seems, then, to be antagonistic. It is also quite minimal: the suffering in question does not reach beyond whatever suffering inevitably accompanies recognizing that one has done wrong. I cannot explore these issues further here, but I hope to in future work.

To restore the equilibrium of benefits and burdens, we must take away the criminal's unfair advantage. Morris thinks we do that through punishment. He does not say much about the way punishment restores this equilibrium, beyond remarking in passing that it involves "what is generally regarded as a deprivation of some thing of value" (479-80). That does not deprive the criminal of the unfair benefit she has already enjoyed, the benefit of casting off self-restraint at the moment of her crime—nothing could deprive her of that benefit, because she has already had full enjoyment of it—but it seems plausible that depriving her of something else of value will reduce her net benefit to the level that the rest of us enjoy.

Jean Hampton has forcefully criticized the picture of a social equilibrium of benefits and burdens upon which this entire approach is built. She points out that it depends on characterizing lawbreaking as the taking of an unfair advantage. But, she argues, it is strange to view lawbreaking as being unfairly beneficial to the lawbreaker. A rapist or murderer may well benefit from not exercising self-control, because doing so means not having to control one's criminal impulses. But the benefit of relaxing one's self-control is not an *unfair* benefit, in the sense of being one that we would all like to keep, but that we have given up in order to live together. It is grotesque to think that a murderer has unfairly relaxed her self-control when the rest of us are continuously exerting ourselves to restrain from murdering each other. The vast majority of people don't need to exercise self-control to keep from murdering each other in the first place.⁶⁰

If Hampton is correct, it is not true that lawbreakers enjoy an unfair advantage over the law-abiding in the way Morris sketches. That knocks the foundation out from under Morris's account.

⁶⁰ See *Forgiveness and Mercy*, pp. 115-116.

Despite these problems, Morris's account introduces an important wrinkle to ordinary restoration strategies. He shifts our attention away from the victim of wrongdoing, and what might be done to restore her well-being, and towards a social *balance* or *equilibrium* that is disrupted by wrongdoing.⁶¹

The valuable state of affairs that Morris identifies is a state of social *balance* or *equilibrium*; that equilibrium is disrupted by wrongdoing, and restored, he suggests, by criminal response. In the remainder of this chapter, I will describe a balance or equilibrium that obtains within special interpersonal relationships, rather than within a legal or communal setting. Disruption and restoration of this balance is a familiar dynamic within interpersonal relationships.

In the ordinary interpersonal cases that I describe first, the disruptions in question need not involve moral wrongdoing. Moreover, the responses that bring about restoration need not be antagonistic. What I describe is a nonmoral and nonantagonistic phenomenon.

But, I will go on to argue, analogous disruptions may be caused by moral wrongdoing. And in the moral case, analogously restorative responses *are* antagonistic. In addition, excitingly, it turns out that antagonistic response is the *only* way to restore the interpersonal equilibrium. In previous chapters, I identified some other important roles that antagonistic response does play; but, I pointed out, antagonistic response is not required to play these roles. That is not the case this time. The moralized sadness that I have pointed to as an alternative in other chapters will not accomplish this restoration; nor will any other nonantagonistic response. If we want to accomplish this particular sort of restoration, antagonism is required.

⁶¹ Of course that is not to say that restoring the victim's well-being is unimportant. But Morris does identify a different variety of restoration that is also morally significant.

But avoiding this one problem does not mean this account is problem-free. Whether we are morally justified in undertaking this restorative antagonistic response to wrongdoing is a further question, one that I will not be able to address until the restoration has been described in sufficient detail.

III

I begin by setting aside moral wrongdoing and response altogether, to focus on restoration within special interpersonal relationships, like friendships. Very roughly, I will argue that, within such relationships, an equality of the attitudes that the participants in the relationship have towards each other—equality of regard and well-wishing—is important, such that the participants in the relationship have reason to try to restore this equality when it is lost. A friendship that involves unequal regard is “one-sided” and unhealthy; the friends involved in such a relationship have reason to try to restore equal attitudes to the relationship, and so to restore the relationship to health.

To see this, consider a specific friendship. Suppose that John and Olivia are close friends. Each enjoys spending time with the other and, although they do not have any explicit standing commitments, they do meet up for dinner a few times a month. At dinner they discuss politics and commiserate about their jobs. They regularly e-mail each other news articles and other items of interest, thus keeping in touch casually when they are unable to get together for dinner, and providing conversation materials for future dinners.

Consider a situation that might develop out of this one. Suppose that Olivia’s involvement in the relationship changes abruptly and for no good reason. Perhaps she simply meets some new people and drops John (along with his e-mailed articles) like yesterday’s news.

But even after the shift in Olivia's intentions and affections has become quite clear, John's attitudes towards Olivia persist. John continues to look forward (now, perhaps, somewhat wistfully) to spending time with Olivia, just as he did before her change. He continues to e-mail her, and he continues to keep open space in his schedule for dinners that never get scheduled.

This relationship seems unhealthy. It seems that something ought to be done. I think that a significant part of the problem with this relationship is the asymmetry in John's and Olivia's attitudes. Olivia is less engaged in the relationship than John. That is problematic.

To resolve this problem, equality of attitudes must be restored. That might be accomplished through Olivia's resuming her former level of involvement. It might also be accomplished through John's withdrawing his involvement in the relationship, to match Olivia's. In the latter case, Olivia and John would become more distant friends; but equality of attitudes would be restored, and their relationship would no longer seem imbalanced.

Before I discuss these equality-restoring reactions in greater detail, consider an alternative explanation for the intuition that Olivia's withdrawal creates a problem with the relationship. Surely there are *minimum* requirements on the attitudes that individuals have within an interpersonal relationship. A friend, for instance, ought to maintain some minimum level of positive intentions and desires towards her friends: she must wish her friend well, intend to help her friend out in some circumstances, and the like. Olivia's attitudes violate these standards. Perhaps that is the only fact we need to point to, to explain our intuitive sense that there is a problem with their relationship.

Olivia's change in attitudes does violate the minimum standards of her relationship with John, and that is a problem; but inequality is a problem as well. To see that, consider a variation of the case that involves no violation of the minimum standard. Suppose that Olivia grows apart

from John: she finds that John is never willing to go much beyond discussing current events, and becomes dissatisfied with the relatively superficial level at which their friendship remains. For this reason, Olivia's involvement in the relationship gradually diminishes. Suppose that her attitudes diminish to the same extent as in the original case. Suppose further that, in this situation, John's involvement with Olivia remains the same. He continues to look forward (now, we might want to say, somewhat foolishly) to spending time with her; he continues to e-mail and to plan.

In this version of events, it is much less natural to say that Olivia violates any specific minimum standard—even though her attitudes towards John ultimately diminish to the same extent. It seems, rather, that the relationship and its minimum standards are changing. They are diminishing, in a way that John fails to notice or fails to acknowledge.

Both the “Olivia unceremoniously dumps John” scenario and the “John doesn't acknowledge that the relationship diminishes” scenario involve a relationship that, at the end, seems problematically out of balance. The best way to explain the problem shared in both of these scenarios is to focus on John's attitudes in comparison to Olivia's. In each scenario, John's attitudes are far stronger than Olivia's; in each scenario, that inequality in attitudes is a problem.

The fact that equality of attitudes has been disrupted does have clear significance for the way the participants in the relationship ought to react. A relationship like a friendship ought to involve equal attitudes. The language of health captures this well: a relationship that involves asymmetric, unequal attitudes is *unhealthy*. The participants in a relationship have reason to act

to keep their relationship healthy, and that they have reason to act to restore the relationship to health when it is unhealthy.⁶²

As I pointed out above, restoring equality of attitudes does not require the person whose attitudes have diminished to restore her attitudes. The other friend can instead withdraw attitudes to match the initial withdrawal. Whether one of these or some combination of the two will seem appropriate depends in complex ways on additional features of the situation. When John has failed to recognize a diminishment of the relationship, for instance, it seems appropriate for him to withdraw his attitudes to match Olivia's withdrawal.

I will not investigate those further considerations. I do, however, want to clearly distinguish how the problem of inequality should be resolved from what the problem consists in. The best or appropriate way to resolve the problem will depend on additional considerations, but the resolution of this problem is always *restoring equality* of attitudes.

Let me reiterate two caveats about this dynamic within special interpersonal relationships. First, the initial disruption need not be wrong. When Olivia's involvement with John gradually diminishes, her change disrupts the equality of their relationship; but she has not wronged John. Second, the restorative response need not be antagonistic. When John responds to

⁶² I want to emphasize that the equality in question is equality of *current* attitudes: the attitudes Olivia and John have *right now* must be equal. A friendship's history is deeply important in many ways; it might be tempting to say that the sum total of the attitudes two friends have borne each other over the history of the relationship must be equal. I think that is a mistake. Consider again the Olivia-John relationship *where John fails to notice that the relationship is legitimately diminishing*. Suppose that the relationship returns to (current) equality because John withdraws his attitudes to match Olivia's. If it were equality of attitudes over the history of the relationship that mattered, this relationship would still seem unbalanced and unhealthy, because for a period of time John's attitudes were unequal with Olivia's. It would seem important that John compensate for that period of inequality—perhaps by bearing Olivia *even lower* attitudes for an equivalent period of time. John would need to show *even less* care for Olivia than she now shows for him, in order to really restore the relationship to health. But—especially considering that John was the one who did not recognize that the relationship was legitimately diminishing—that does not seem right.

I do not mean to assert that historic inequality in a relationship is never problematic. But the way it is problematic is not captured by bluntly requiring that a relationship involve equal attitudes over the course of its entire history. Relationships can grow and diminish asymmetrically, with one participant growing attached before the other; however we account for the importance of relationship histories, this sort of asymmetry should not be labeled problematic.

Olivia's gradual diminishment with a matching withdrawal—when he acknowledges that they have grown apart as friends—he is not being antagonistic towards Olivia.

IV

Thus far I have relied on a rough, intuitive notion of equality, and an unsupported—though I think intuitively appealing—claim that special relationships like friendships must involve equal attitudes to be healthy. Before moving on to consider whether this dynamic applies to moral wrongdoing and response, I must identify more specifically the sort of equality that is issue.

It is not in fact easy to say what equality of attitudes really requires. It is certainly not a matter of equality at the level of specific attitudes, like individual intentions. Consider John and Olivia again. Perhaps Olivia intends to teach John how to bake. In that case it is unlikely that John also intends to teach Olivia how to bake. John may admire Olivia's pastry-baking talents, because they far surpass his own; Olivia will likely recognize this and does not have the same admiration for John's pastry-baking skills. These inequalities on the level of specific attitudes do not make their relationship unequal in an unhealthy way.

If we leave specific, concrete equality behind, however, comparisons seem very difficult to make. It is hard to see how to measure John's admiration for Olivia's talents against her intention to teach him, to decide if those attitudes are really equal or not.

To make sense of equality of attitudes without demanding implausibly exact equality, we need to distinguish two sorts of attitudes involved in a friendship. Being friends with someone involves two distinct elements. First, one attaches a special value to one's friend. And, second,

one reacts to that value by wishing that person well and undertaking to do good things for her: by having goodwill for her, in a broad sense of the term.

John values Olivia: he takes pleasure in their time together, he is interested in her opinions, her support helps him to deal with his job, and so on. And he wishes her well. He is supportive of her travails at her job, he makes an effort to make time to spend with her, and he actively searches for news articles he thinks she will enjoy. Supposing that the relationship is healthily equal, Olivia values John similarly, and wishes John well to a similar extent.

These two elements are doubtless interrelated in complex ways. But we need to distinguish them to be able to compare the attitudes of different participants in a relationship. It does not make sense to compare John's admiration for Olivia—an aspect of his valuing Olivia—to her intentions to help him. But within each of these two categories, comparisons of attitudes do make sense. We can compare the value that John attaches to Olivia and the value Olivia attaches to John; and we can compare the well-wishing and good intentions that each has for the other. We can speak of (at least rough) equality between John and Olivia, within each of these categories.⁶³

I want to emphasize that both 'goodwill' and 'valuing' are being used in a very broad sense here. Goodwill for someone may, for instance, involve an intention to help her; a desire that she do well; an inclination to be pleased at things going well for her; a willingness to take on burdens in order to benefit her; and so on.⁶⁴ In spite of these variations, within the broad category of goodwill we are able to make comparisons. Of course, these judgments will be rough, and

⁶³ This sketch of the structure of a friendship is derived, with simplifications, from Aristotle's discussion of friendships in books VIII and IX of the *Nicomachean Ethics*. I diverge significantly from Aristotle in my understanding of the valuing element. See below, especially n. 69.

⁶⁴ It should be clear that this goodwill is not equivalent to the Kantian notion of having a good will. It also seems more narrow than Strawson's notion of goodwill in "Freedom and Resentment." Strawson's notion of goodwill is centered on attitudes that satisfy the demands made within interpersonal relationships; those demands seem to encompass both what I am labeling "goodwill" and what I am labeling "valuing."

there is substantial room for variation in goodwill within a healthy relationship. But, most importantly, we are quite capable of identifying problematic levels of inequality. If, for instance, Olivia intends to go to great lengths to teach John pastry baking, but John's only goodwilled attitude towards Olivia (beyond the minimal level of goodwill he would have towards a total stranger) is a disposition to be pleased when things go well for her—if he is unwilling to go to any trouble to help her—their relationship is imbalanced.

My use of the term 'valuing' also requires some elaboration. 'Valuing' calls to mind a response to someone's antecedently or independently valuable features.⁶⁵ But it is odd to say that John values Olivia simply for her skills; we want to say that some friends, at least, are valued *for themselves*. Harry Frankfurt has set out some of the dimensions of this sort of valuing in his discussions of what he calls *caring*. When a person cares about something, Frankfurt says, he is "...invested in it. He identifies himself with what he cares about in the sense that he makes himself vulnerable to losses and susceptible to benefits depending upon whether what he cares about is diminished or enhanced" ("The importance of what we care about," p. 83).

One can identify in this way with something, Frankfurt thinks, even if that thing is not itself independently valuable.⁶⁶ Caring need not be a response to something's being independently valuable; in fact, caring *makes* the thing cared about valuable to the person who cares about it. Caring *generates* value that is specific to the person who cares.⁶⁷ When John comes to care about Olivia, then, that makes Olivia *valuable for him* in a way she was not

⁶⁵ On Aristotle's account of friendship, the valuing element is precisely this: we respond to a friend's usefulness, pleasantness, or to her virtuousness.

⁶⁶ Although there may be *some* limits to the things one can identify with in this way. Frankfurt gives the example of caring about not stepping on cracks in the sidewalk: that is not worth caring about.

⁶⁷ See "The importance of what we care about," p. 90, on the subject-relativity of this valuing (contrasting that with the objectivity of the values of rationality and morality), and p. 92 on the creation of value: "if there is something that a person does care about, then it follows that it is important to him. This is not because caring somehow involves an infallible judgment concerning the importance of its object. Rather, it is because caring about something *makes* that thing important to the person who cares about it."

before.⁶⁸ As a matter of the history of their relationship, this caring may have its origins in recognition of antecedently valuable features of Olivia's, for instance her talents or the fact that John finds her pleasant to spend time with; but when John comes to care about Olivia, she comes to have a value for him that she did not before.⁶⁹

There are, then, two sorts of equality that are required for a relationship like a friendship to be healthy: equality of valuing, and equality of goodwill. The participants in a relationship of this sort have reason to try to restore equality when it is lost. That may involve either restoring the valuing or goodwilled attitudes that have gone missing, or a matching withdrawal of attitudes (or some combination of the two). Which of those options will be appropriate in a given situation depends on other features of the situation. But either will restore the relationship to equality.

Before investigating whether this model can be applied to moral response, I want to consider an apparent problem for my claim about the reason-giving significance of unequal attitudes in interpersonal relationships.

I claim that participants in certain sorts of relationships have reason to try to make their relationship equal when it is unequal. Although I have restricted most of my discussion to the

⁶⁸ One might wonder why I am not simply adopting Frankfurt's terminology and saying that it is part of a friendship that John cares about Olivia. Why use the term "valuing"? I do so because Frankfurt also claims that the phenomenon he labels caring has a motivational element: if we care about something, we direct and shape our lives by reference to it. Caring involves what I am labeling goodwill. Frankfurtian caring combines the motivational and valuing elements that I need to distinguish, in order to make sense of talk of equality.

⁶⁹ This understanding of valuing allows us to understand how certain relationships involve equal valuing where Aristotle's view, on which one responds to independently valuable features of one's friend, has trouble. Consider an odd-couple friendship: a friendship between one steady, reliable person and one irresponsible, unreliable person. Of course, sometimes (as in Neil Simon's 1965 play "The Odd Couple"), the irresponsible, unreliable person is a valuable counterbalance to the reliable friend's uptightness. But that is not always the case; sometimes an irresponsible friend is not in fact very valuable. But if we allow that caring can *create* value—can *render* someone valuable to the person who cares—then we can make sense of an irresponsible, unhelpful friend and a responsible, useful friend being equally valuable to one another.

case of friendships, equality of attitudes seems to be similarly important to romantic relationships. If that is right, then participants in an unequal romantic relationship have reason to try to restore their relationship to equality when it is unequal.

In some circumstances, this generates very unappealing results. Suppose, for instance, that John develops extremely strong romantic feelings for Olivia. Then, it seems, Olivia has a reason to increase her own attitudes towards John, in order to make their relationship healthy.

But that is a stalker's dream come true. Just by having affection for someone, one can give her a reason to try to reciprocate the affection. Even if there are other reasons (prudential ones, for instance) not to get involved in such a relationship, this seems to give stalkers a normative, reason-giving power they do not have. Call this the "leveling-up" objection⁷⁰ to the claim that participants have reason to make special relationships equal.

It is useful to work through the reasoning about a stalker in a bit more detail. First: an equal romantic relationship is more healthy than an unequal one. That claim's application to the stalker case *is* plausible. The relationship between a stalker and the object of his attentions would be healthier if they both valued each other equally. Other things being equal, if the two were indifferent towards each other, or if they were in an ordinary, mutually affectionate romantic relationship, their relationship would be healthier than it is right now.

Next: since the relationship would be healthier if it were equal, both the stalker and the object of his attentions have reason to do what they can to get to a situation of equal valuing. I think this is clearly correct in many of its applications: the stalker has reason to do what he can to control and eliminate his attentions. The object of the stalker's attentions has reason to try to dampen her stalker's affections as well.

⁷⁰ Analogous to the "leveling-down" objection to the claim that equality is a political value.

This line of reasoning only generates unappetizing conclusions when the target of a persistent stalker applies it to herself. Apparently, if the stalker will not adjust his attitudes, the stalking victim has reason to act unilaterally to do what she can to bring about a “leveling up” in her attitudes towards her stalker.

We want to reject this conclusion. It seems offensive to give a stalker this kind of normative power. But I don’t think that is the right way to respond to this situation. We should admit that the stalker is utilizing the dynamics of interpersonal relationships to his advantage. By introducing an unhealthy inequality into an interpersonal relationship *and refusing to address that inequality himself*, the stalker does bring it about that his victim has some reason to increase her affection for him, as the only available way of getting out of the unhealthy relationship.

The stalker’s victim does not, I want to emphasize, have an all-things-considered reason to increase her valuing and goodwill towards the stalker. There are strong countervailing moral and prudential reasons for her to avoid doing so. But the stalker does manipulate the reasons that apply to his victim. That is part of what is wrong with stalking.

V

To summarize the interpersonal case: certain types of interpersonal relationships must involve equality of valuing and equality of goodwill to be healthy. Participants in a relationship that is unhealthy in this way have reason to try to restore the relationship to equality: they have reason to alter their attitudes, or to take actions to bring about alterations in their attitudes, so that they are equal once again.

I suggest that wrongdoing disrupts the equality of an interpersonal relationship in this way. And, I want to argue, an antagonistic response to wrongdoing is often required to restore the relationship to equality.

How does wrongdoing disrupt the equality of an interpersonal relationship? And how does an antagonistic response restore equality? Begin with the equilibrium that is to be disrupted. Suppose that there is some basic moral relationship: one that obtains between all people, just in virtue of their being persons. If there is such a relationship, it seems, that relationship should involve equality of valuing and equality of goodwill.

Wrongdoers often fail to value their victims adequately: they fail to have the minimum amount of care and concern for their victims that anyone deserves just in virtue of being a person. Wrongdoers also often fail to bear their victims adequate goodwill: they fail to have the minimum morally required intentions to help or (at least) take care to avoid inflicting harm. Wrongdoing that involves these attitudes disrupts the equality of the basic moral relationship.

If the wrongdoer will not restore her valuing and goodwilled attitudes towards her victim, the victim can restore equality to the relationship by withdrawing goodwill and valuing attitudes in a way that matches the attitudes of the wrongdoer.

Suppose, for instance, that the wrongdoer acted recklessly: that she did not care adequately about her victim's well-being, and she was insufficiently motivated to avoid harming her victim. If the wrongdoer will not restore her attitudes, the victim can restore their relationship to equality by withdrawing some concern for the wrongdoer's well-being, and her motivations to avoid harming the wrongdoer, to a similar extent.

In the case of recklessness, this matching withdrawal may not seem antagonistic. But suppose instead that the wrongdoer acts maliciously, in a way that is antagonistic. Suppose that

the wrongdoer views the victim's suffering as intrinsically valuable, and is motivated to bring that suffering about for its own sake. If the wrongdoer will not withdraw these malicious attitudes, the victim can restore their relationship to equality by adopting similar malicious attitudes. The wrongdoer's valuing attitudes toward the victim are extremely negative because they involve valuing her suffering for itself. If the victim is to take a similar devaluing attitude toward the wrongdoer, she must value the wrongdoer's suffering for itself as well. And the wrongdoer's motivations toward the victim are similarly negative because they involve being motivated to bring about the victim's suffering for its own sake. If the victim is to have a similar attitude toward the wrongdoer, she must also be motivated to bring about the wrongdoer's suffering for its own sake. If a victim who has been the object of antagonistic attitudes is to restore her relationship with the wrongdoer to health, she must bear the wrongdoer similarly antagonistic attitudes.

If all this is correct, antagonistic response to wrongdoing is a way to restore equality to a relationship in the wake of antagonistic wrongdoing. It is important for me to acknowledge that, on this account, antagonistic response is not quite *required* to restore equality; the wrongdoer might instead reverse her antagonistic attitudes towards her victim. But when a victim is facing a recalcitrant, antagonistic wrongdoer, she can use antagonistic response to restore their relationship unilaterally. On this account, only antagonistic response can *empower* a victim of wrongdoing.

If, that is, the basic moral relationship resembles the interpersonal relationships like friendships with which I began. This entire account rests on the claim that there is a moral relationship which is healthy when it involves equal regard.

But that is false. The basic moral relationship is not like special interpersonal relationships in this respect. That is apparent if we think more carefully about the sort of moral response under consideration.

A wrongdoer values her victim less than is morally required, or she does not have the minimum morally required goodwill for her victim, or both. To match those attitudes, and so restore the moral relationship to equality, the victim of wrongdoing must bear the wrongdoer attitudes that also fall below the level that is morally required. But that is *ipso facto* morally unacceptable. It cannot be morally acceptable to bear someone attitudes that fall below the morally required minimum.

Compare a situation in which one person shows another impermissibly low levels of goodwill or regard, and another situation in which two people show each other equally impermissibly low levels of those attitudes. It is, I think, readily apparent that the first situation, involving one-sided impermissible attitudes, is preferable to the second, involving impermissible attitudes on both sides. It would be better if at least one of the participants in the relationship showed the other goodwill and regard.

I pointed out above that there are two sorts of norms that apply to the attitudes of participants in a friendship: ones that set minimum requirements on the attitudes of each friend, and ones that require equality of attitudes between friends. Because equality is independently important in special relationships like friendships, when a friendship is disrupted by a decline in one participant's attitude, the other participant can restore the relationship to health by withdrawing her own attitudes to match. The friendship changes; it is less close than it was. But this new, more-distant friendship can still be a healthy relationship. In the moral relationship, on

the other hand, the minimum requirement on participants' attitudes of overwhelming importance; attitudes that drop below that minimum moral standard are, by definition, morally unacceptable.

Consider a way of trying to resist these conclusions. One might argue that wrongdoers deserve the disregard or ill will that they receive, in a way that their victims do not. Then, it seems, the victim's disregard can restore equality to the relationship without being morally unacceptable. Consider, for instance, a wrongdoer who lies to someone. In doing so, the wrongdoer withdraws goodwill for his victim, just to the extent of withdrawing the intention to tell her the truth. Suppose that the victim responds by withdrawing her intention to tell the truth to the wrongdoer. This withdrawal may be justified: "he doesn't deserve truthfulness from me," the victim might rightfully assert. And this withdrawal matches the wrongdoer's violation, thus restoring the relationship to equality.

Or so it seems. If the victim's willingness to lie is in fact morally acceptable, it does not restore equality to the relationship. The claim that the victim's withdrawal of her truth-telling intention restores the relationship to equality is quite sketchily grounded in the fact that her willingness to lie and his willingness to lie are similar in content. But that similar content is not enough to guarantee equality in attitudes after the withdrawal.

Well, how do we determine whether two people's attitudes are equal or not? In this case, there is a sense in which the absolute level of good or ill will that each person wishes the other is the same: each person is willing to lie to the other. But *ex hypothesi* those attitudes differ in their relationship to what is morally required. The wrongdoer's attitudes fall below those requirements; the victim's attitudes do not.

We are considering the basic moral relationship; when attitudes within that relationship are compared, what matters is their strength relative to what is morally required, not their specific content. If specific content mattered, then someone who wrongfully withdrew his intention to tell the truth and a decent person who withdrew her truth-telling intention for good reasons (for instance, in order to contain information about a surprise party) would both have the same effect on their respective moral relationships. But they do not.

If the victim's willingness to lie is morally acceptable, then, within the basic moral relationship, her (acceptable) attitudes towards the wrongdoer are not equal to his (unacceptable) attitudes towards her. If the victim responds in a morally acceptable way, she does not restore equality to the moral relationship.

VI

In this chapter I pursued several ways of appealing to restoration to justify antagonistic response to wrongdoing. At the beginning, I pointed out that it was not clear how specifically antagonistic response would serve to restore the specific items of value that are lost when wrongdoing occurs. Returning stolen property might be unpleasant for the wrongdoer; but that is not antagonistic in the way that hatred, anger, and malice are antagonistic. It does not involve treating the wrongdoer's suffering as good in itself.

But, I pointed out, an equilibrium of attitudes might be disrupted by antagonistic wrongdoing and restored by antagonistic response. In that case we do not look for antagonistic response to restore a specific item of value; instead, antagonistic response helps to restore equilibrium or balance to the situation. The only way for the victim to accomplish this is through

further antagonistic response. Only similarly antagonistic attitudes will balance the wrongdoer's antagonistic attitudes.

But restoring balance in this way must be morally problematic. To balance the wrongdoer's morally problematic attitudes, the victim would need to have similarly morally problematic attitudes.

In fact, I argued, when we examine the moral case it appears that equality of attitudes is not independently valuable, as it is in many special interpersonal relationships like friendships. Other interpersonal relationships can diminish but remain healthy, so long as they are still equal; diminishment of the basic moral relationship—consisting of the basic attitudes we are morally required to bear towards each other—is problematic no matter how equal that relationship is. Moral relations are in this way significantly different from other interpersonal relationships. These differences undermine the attempt to justify antagonistic response by appeal to its ability to restore an interpersonal equilibrium.

CHAPTER 5: NULLIFICATION, NOT RESTORATION

I turn finally to the classic retributive notion of *nullification*. Many retributivists suggest that some apparently antagonistic response to wrongdoing—often criminal punishment in particular—is required to nullify or cancel out the wrongdoing to which it is a response.

It is important to sharply distinguish *nullifying* responses to wrongdoing from the *restorative* responses to wrongdoing discussed in the previous chapter. Nullifying responses are retroactive in a robust sense: they involve altering the past in some way, by eliminating or altering some morally significant aspect of a past wrongful action. Restorative responses are not retroactive in this way; they involve bringing back—restoring—some good state of affairs that was destroyed or altered by the wrongdoing. The two are often run together in the literature, but they are distinct.

The difference between restoration and nullification is familiar in other contexts. Consider, for instance, the difference between annulling a marriage and ending it through divorce. A marriage annulment brings it about that (from the law's perspective) the parties involved were never married.⁷¹ By contrast, a divorce ends a marriage without wiping it from legal history; instead, it restores the parties' unmarried status, from the date of the divorce onwards.

The difference between justifying punishment as a restorative response to wrongdoing and justifying it as a nullifying response might be captured by saying that restorative justifications are forward-looking, because they are concerned with bringing about a good state

⁷¹ This is a drastic oversimplification of actual legal practice, which varies from state to state, and which I discuss in greater detail in section II below. I do, however, think that this captures the common understanding of the significance of annulment.

of affairs in the future, while nullification justifications are backward-looking, because they are concerned with altering a problematic event which occurred in the past. But restorative accounts are still backward-looking in an important sense: they look back to the wrongdoing to understand what problem there is to be addressed going forward. They are sensitive to—they focus on restoring—the specific good state of affairs that the wrongdoing destroyed. (Or, equivalently, they are sensitive to, and focus on eliminating, the bad state of affairs that has been created by wrongdoing.) This is a natural⁷² and important way of being backward-looking. (By contrast, deterrent-utilitarian justifications of punishment may pay much less attention to the nature of the wrongdoing itself; they are concerned with using that wrongdoing as an occasion for deterrence.⁷³) There are many senses in which a justification of punishment may be backward-looking. Since nullification theories aim to change the past, they are backward-looking in a very serious way; but I do not wish to deny the (purported) honorific *backward-looking* to other views.

It would be a powerful recommendation for antagonistic response that it is capable of retroactively eliminating or altering wrongdoing, so that the event turns out to have been less bad than it otherwise would have been. But annulment is often the target of philosophical scorn, with the suggestion that we can actually change the past being described as “metaphorical,”⁷⁴ “obscure,”⁷⁵ and “incoherent.”⁷⁶

⁷² To choose an example nearly at random, in “Justice in Reparations,” Chris Kutz describes restitutions (in the form of the return of state-confiscated property to residents of post-Communist Czechoslovakia) as backward-looking, apparently because such restitutions aim to restore property to its prior owners. He does so in passing and contrasts that with “forward-looking” aims also active at the time, like quickly bringing about the privatization of state-owned property.

⁷³ And yet, as Rawls points out in “Two concepts of rules,” even within a deterrent-utilitarian framework specific instances of criminal punishment may be backward-looking in a robust sense.

⁷⁴ Jean Hampton (*Forgiveness and Mercy*, p. 114) and David Dolinko (“Some Thoughts About Retributivism,” p. 549) both describe nullification this way. Hampton does, however, go on to make the most serious attempt yet to

I will argue that we can make sense of nullifying wrongdoing, without running into incoherence or taking refuge in obscurity and metaphor. I begin by looking more closely at the worries philosophers have had about annulment. Then I draw on examples of annulment outside the wrongdoing-response context, including marriage annulment, to distinguish several sensible ways of talking about annulment and changing the past. This enables me to more clearly specify the strongly retroactive form of annulment that is my focus. Next, drawing on the work of Herbert Fingarette, Jean Hampton, and others, I describe several clear, nonmetaphorical ways in which wrongdoing could be annulled by punishment. I argue that these attempts fail to justify punishment—but for moral reasons, not because of obscurity or incoherence in the annulment proposed. Finally, building on a proposal of Jean Hampton’s, I argue that we do in fact have a substantial moral reason to bring about a certain form of annulment.

I

None of the annulment proposals I will describe propose eliminating an event from the past altogether. But just altering an event in the past—even if that does not involve eliminating the event altogether—seems similarly impossible. This basic problem facing nullification is a metaphysical, rather than a moral, problem: it is a problem with the possibility of successfully doing what nullification theories propose, not a problem with the moral significance of doing so.

It does not help matters that Hegel, of all philosophers, is regarded as the originator and most prominent proponent of a nullification theory of punishment. At the beginning of his discussion of annulment theories of punishment, for instance, Ted Honderich quotes Hegel

make sense of nullification, and Dolinko engages at length with Hampton’s attempt. I discuss both in greater detail below.

⁷⁵ See *Punishment: The Supposed Justifications Revisited*, p. 42.

⁷⁶ See “Morality and the Retributive Emotions,” p. 5.

discussing annulment in the *Philosophy of Right* and remarks, “All this is, of course, obscure. It is by Hegel” (42).

Well, yes, I also find Hegel obscure. Although I am confident that careful examination of Hegel’s work will uncover an interesting, powerful account of punishment (one that may or may not actually involve the specific, strongly retroactive nullification that is my focus), I will not attempt such an examination here. Instead I will mine some contemporary discussions of punishment for material that will help me to develop some clear, concrete proposals about the ways that we annul some aspect of wrongdoing.

I describe this as “mining” the work of other philosophers because most philosophers who discuss nullification—perhaps driven by concerns about the impossibility of genuine, retroactive nullification—slide easily from talk of nullification to talk of restoration. Hence there are not many straightforward, focused accounts of nullification to engage with.⁷⁷

John Cottingham, for instance, suggests that annulment is better understood as analogous to civil restitution: he thinks we should understand talk of nullifying a crime on the model of paying damages for breaking a window. On this model, punishment is a way of paying back the victims of crime.⁷⁸ But that does not really nullify the crime at all; the crime itself is not retroactively altered in any way. Instead, punishment-as-compensation is supposed to *restore* the victim to the level of well-being she enjoyed before the wrongdoing. (Cottingham is—in my view rightly—dubious that this will justify criminal punishment.)

⁷⁷ This has the unfortunate, potentially frustrating consequence that the criticisms I make below may not have force against the actual views of the philosophers whose work I pillage. I am open to the objection that I am constructing straw men here; I only want to insist that these are theoretically interesting straw men, ones that may serve to inspire the construction of similar, but less straw-filled, accounts.

⁷⁸ See “Varieties of Retribution,” p. 244.

Other philosophers just treat annulment and restoration as the same thing. Honderich, for instance, runs annulment and restoration together in a single sentence, saying that “A punishment is an annulment, a cancellation or a return to a previous state of affairs;”⁷⁹ David Wood discusses a number of accounts on which “...punishment annuls, cancels out, rectifies, or repairs the wrong.”⁸⁰ And even in her serious, focused discussion of nullification (which I discuss in greater detail below), Jean Hampton sometimes talks about punishment serving to “reaffirm a victim’s equal worth in the face of a challenge to it”—suggesting that punishment simply restores the state of affairs in which the victim’s worth is adequately established or affirmed.⁸¹

A. M. Quinton has restorative accounts in mind when he discusses annulment:

The doctrine of "annulment", however carefully wrapped up in obscure phraseology, is clearly utilitarian in principle. For it holds that the function of punishment is to bring about a state of affairs in which it is as if the wrongful act had never happened. This is to justify punishment by its effects, by the desirable future consequences which it brings about. (135)

Quinton points out that purported nullification theories that are concerned with making things as they had been before the punishment—with restoration—are thereby focused on the positive future effects of punishment. I will set aside the tendentious claim that such theories are utilitarian,⁸² but that jolting declaration about purportedly retributive theories—theories that are

⁷⁹ See p. 41. This is just before Honderich quotes Hegel. Honderich appears to find it similarly obscure how punishment could restore things to the way they were before, say, a murder occurred. He does, however, go on to discuss some ways of filling in the restorative suggestion in greater detail, without attempting to do the same with the suggestion that we annul wrongdoing retroactively, as we annul a marriage.

⁸⁰ “Retributive and Corrective Justice, Criminal and Private Law,” p. 547. Wood discusses at length a number of variations on what he labels the ‘annulment’ theory; all involve repairing some problematic state of affairs connected to wrongdoing.

⁸¹ *Forgiveness and Mercy* p. 126. In later work—mainly “Correcting Harms Versus Righting Wrongs” and “Punishment, Feminism, and Political Identity”—Hampton appears to largely retreat from her claims about nullification and settle on this *reassertion* as the central justifying function of punishment.

⁸² Considering the current philosophical antipathy towards utilitarian theories of punishment, accepting this label would really be very tendentious. After all, restorative theories do not focus on the good of deterrence which appears to provide a utilitarian motivation for punishing the innocent (it also seems doubtful that the goods that restorative theories do focus on would be promoted by punishment of the innocent). Hence restorative theories do not seem to be susceptible to a standard objection to utilitarian justifications of punishment. The label also just seems inapt.

often presented as *antithetical* to utilitarian justifications of punishment—is useful. It reminds us of the strong sense in which annulment theories, but not restorative theories, are backwards-looking: on annulment theories, punishment is justified not by the desirable future effects it brings about, but by what it does to change a past wrongdoing.

As I have said, I suspect that this slide from annulment to restoration is driven by the perceived difficulty of providing some specific, concrete explanation of the way that punishment alters the past. If punishment cannot retroactively make it so that the wrong did not occur, perhaps it can instead make it so that things are—now, and going forward—as they were before the wrongdoing occurred. Altering the future is not as difficult as altering the past.

I should, however, note that some of these philosophers worry that a restorative justification of punishment will be similarly difficult to make out. Quinton, for instance, writes that “theft and fraud can be compensated, but not murder, wounding, alienation of affection or the destruction of property or reputation... things cannot always be made what they were” (135); and Honderich points out that “my death or imprisonment, after I have killed a man, does not make things what they were before” (42). No punishment for murder restores the state of affairs that obtained before the murder; no punishment will resurrect a murder victim.

In application to some crimes, then, restorative theories face problems similar to those facing annulment theories. But at least the basic restorative strategy of bringing about good

After all, a utilitarian thinking about punishment ought to look to all the consequences that punishment might have, beyond just the wrongdoing-reversing effects of punishment: she ought to, for instance, incorporate the deterrent effects of punishment into her moral calculus. A theory that restricts its attention to the restorative consequences of punishment, and refuses to take deterrent or other consequences into account, likely has nonutilitarian reasons for doing so.

future effects is a nonmysterious one. It is, by contrast, difficult to see how the basic nullification strategy of altering the past is possible at all.

II

This may seem to overstate the case. As I have pointed out, nullification does make sense in some circumstances. When a marriage is annulled, for instance, it is retroactively voided; we are not troubled by that legal process. Why be so troubled by talk of changing the past?

The short answer is that a marriage and a wrongful action are, of course, quite different. But that oversimplifies matters; even marriage annulment can operate through a variety of different mechanisms. It will be useful to work through these in some detail. Doing so will enable me to more clearly specify what would be involved in altering past wrongdoing.

Begin by considering an incestuous marriage. Suppose that, unbeknownst to everyone involved, two people who get married are siblings. Suppose further that (as is actually the case) marriages between siblings are legally prohibited. The marriage will be annulled if the (putative) spouses' blood relationship is discovered. This annulment is simply a *recognition* that the people weren't legitimately married all along. Someone with legal authority may have performed a ceremony; the state (and everyone else) may have acted as if these two people were married; but they did not satisfy one of the conditions on being married. The purported spouses were mistaken, as was the legal system and its representatives.

Call this annulment as recognition: to annul something in this sense is to recognize that it never existed in the first place. There is more variation in state marriage law than I can do justice to here, but this is a common way of treating incestuous and bigamous marriages. Such a marriage was never valid in the first place; annulment is just a recognition of that fact. Often

state laws are explicit about the retroactive dating of this recognition; California law, for instance, describes these marriages as “void from the beginning.”⁸³ Notably, it is not up to the parties involved whether to be married or not. Even if they want to be married, they are not.⁸⁴

There is an attenuated sense in which recognition annulment is retroactive: our understanding of the events that occurred in the past is altered. We now recognize that when the state accepted a joint tax return from two siblings who entered into a void marriage, it made a mistake. (That is not to say that the correct course of action upon recognizing this mistake would be to act as if two unmarried people fraudulently filed a joint tax return. Other considerations are surely relevant in the aftermath of this sort of mistake. Annulment law often treats children born or adopted by a couple within a void marriage as being legitimately the children of both parents, likely out of concern for the well-being of the children in the aftermath of this sort of mistake.⁸⁵) But this alteration in our understanding of the past clearly does not involve *altering* the past. It involves recognizing what has been true all along.

Bigamous and incestuous marriages are often grouped together in state laws as “void” marriages. They are separated from another group of marriages, often labeled “voidable” marriages.⁸⁶ These include marriages entered into under duress or because of fraud, and marriages in which one or both parties are underage, have a mental illness, or are incapable of consummating the marriage.

⁸³ Cal. Fam. Code §§2200, 2201.

⁸⁴ Although, of course, they may succeed in concealing their biological relationship, and so deceive the state into treating them as if they are married.

⁸⁵ See N.Y. Dom. Rel. Code §24-1, 750 IL Comp. Stat. §5/212, Colo. Rev. Stat. §14-10-111. California law does not make this presumption. (The distinction between legitimate and illegitimate children is nowadays quite minimal, but it can have some significance for immigration purposes.)

⁸⁶ Compare NY Dom. Rel. Code §§5 (“Incestuous and void marriages”) and 6 (“Void marriages”) to §7 (“Voidable marriages”), and California Family Code §§2200 and 2201 to § 2210.

Annuling these marriages is not a matter of recognizing that they never existed in the first place. Instead, they involve conditions which give someone (often only one party—e.g., the one who was subject to duress, or underage) an opportunity to retract his or her consent to the marriage. That is reflected in the limits on this sort of annulment: in many states, if someone enters into a marriage because of fraud, and then *after discovering the fraud* continues to cohabit as spouse, she cannot seek an annulment. The continued cohabitation as a spouse indicates consent to be married in spite of the fraud.⁸⁷

There is, again, tremendous variability in state laws here; but this second class of marriage annulment is clearly different from the first. Although these do involve recognition of some fact (duress, for example) that has obtained all along, that does *not* mean that we discover that the marriage never existed. Instead, we discover some fact that gives some parties the power to have the marriage annulled.

Interestingly, many state laws do not retroactively date this second sort of annulment in the same way as recognition annulment. In New York, for instance, this sort of marriage is null “from the time its nullity is declared by the court.”⁸⁸ The refusal to retroactively date this second class of annulment is not universal, however; Colorado law, for instance, declares categorically of both sorts of annulment that “Marriages declared invalid...shall be so declared as of the date of the marriage.”⁸⁹ In Illinois, making annulment retroactive is the default “unless the court finds, after a consideration of all relevant circumstances, including the effect of a retroactive

⁸⁷ See Cal. Family Code §2210. Similar conditions are attached to annulment of underage marriages and marriages obtained by force.

⁸⁸ NY Dom. Rel. Code §7.

⁸⁹ Colo. Rev. Stat. §14-10-111, ¶5. Along with considerations like force, fraud, and duress, Colorado allows for marriage annulment if “one or both parties entered into the marriage as a jest or dare.”

judgment on third parties, that the interests of justice would be served by making the judgment not retroactive.”⁹⁰

There is some intuitive pressure towards saying that this sort of annulment is not retroactive. The impact of this sort of annulment is that, going forward, the persons in question will be treated by the law and legal institutions as if they never were married. The sense in which we alter the past is extremely limited: we alter the past as we will think of it, from this point forward.

Call this sort of annulment as-if annulment: we proceed *as if* something has not happened.⁹¹ In spite of the intuitive pressure towards thinking of as-if annulment as merely a change in our behavior going forward, I want to insist that it can sensibly have a strongly retroactive element. This retroactive element is, however, very circumscribed: it is retroactive alteration of the past *within a conventional context*. Within such a context, and for the purposes pursued within such contexts, the only past that matters is the past as it is recognized by the convention or conventional institutions. Within the law, and for the purposes pursued by the law, the only marriages that have taken place are the ones the law recognizes to have taken place. Organized games are another such context. In football, after the end of a play a defensive player can be declared to have committed a foul. The game then proceeds *as if* the ball had not been moved during that play; the motion of the ball on that play did not happen, for the purposes of the game. Other aspects of the play are not annulled: the time elapsed is not put back on the game clock, and the foul itself has an impact, in the form of a 5-yard benefit to the opposing team. This situation has further parallels to as-if marriage annulments: the opposing team can decline the penalty, and decline to have the play annulled (if for instance they scored during the

⁹⁰ 750 IL Comp. Stat. §5/304 (titled “Retroactivity”).

⁹¹ Thanks to Gavin Lawrence for pressing me to think more carefully about this kind of annulment.

play or wish to retain some other benefit obtained during the play). The foul provides the opposing team with an opportunity to have aspects of the play annulled. These are genuine changes to the past, as it matters within the conventional context.

But of course this sort of annulment is limited. Its reach is limited to features of events insofar as they have significance within conventional contexts. We do not alter the past, apart from altering its conventional significance.⁹²

I am pursuing the proposal that some responses to wrongdoing, like punishment, are justified because they annul morally problematic features of wrongdoing. I do not want to limit this proposal to recognition annulment, or as-if annulment. We might sometimes be able to recognize some sense in which wrongdoing is already null,⁹³ and some wrongdoing might be significant primarily within a conventional context, in a way that opens the door for as-if annulment. But I do not think that annulment is limited to these possibilities. It is possible to describe annulment that, unlike recognition annulment, involves genuinely altering the past—and, unlike as-if annulment, extends beyond conventional contexts.

The best way to establish this is to describe a response that annuls wrongdoing while being neither recognition annulment nor as-if annulment. But first it will be useful to describe an

⁹² Fingarette (“Punishment and Suffering,” 507) and Honderich (*Punishment: The Supposed Justifications Revisited*, 42) make similar points.

⁹³ Hegel’s discussion of punishment sometimes suggests something like this. Hegel asserts that a crime itself is null, without punishment’s having taken place; punishment accomplishes something like revealing this nullity, or making it manifest (which may not come to the same thing). See, e.g., *Elements of the Philosophy of Right* §97 ff. But Hegel also talks, repeatedly, about “canceling” a crime, which is (to my unindoctrinated ear, at least) suggestive of changing what has already happened. What is really required to resolve these issues, of course, is a detailed and systematic understanding of Hegel’s philosophy as it is applied to punishment. I cannot undertake that task here. Allen Wood has attempted to do so; he argues that Hegel’s talk of annulment should be jettisoned or, at best, folded into an argument that the criminal *consents* to be punished for her actions. (This is Wood’s gloss on Hegel’s claim that the criminal’s will is null. See Wood, *Hegel’s Ethical Thought*, Ch. 6, especially pp. 112-114.)

extremely general schema for the alteration of past events outside of conventional contexts.

Doing so will make clear one final pitfall that annulment theories should avoid.

What non-conventional features an event has can depend on the context within which the event takes place. The death of an individual passenger pigeon can be the end of a species or not, depending on whether there are any more passenger pigeons around. Moreover, what non-conventional features an event has can depend on aspects of the context that come after the event itself occurs. Whether a rock's falling is the beginning of an avalanche or not depends on aspects of the context that take place after the fact—on what happens after the rock falls. If all this is true, then it is possible to alter the non-conventional features of an event after it occurs, by altering aspects of its context that occur after the fact.

It is important to be careful here. In this latter case, the feature of the falling that depends on what happens afterwards is *purely a matter of what happens later*. All we change is what happens next. It might be tempting, in light of this, to say that the event itself is unchanged. But that is oversimplifying. There is an important difference between a single rock falling and a single rock falling that is the beginning of a massive rockslide. If I were nearby, that would be one of the features of that event that would concern me the most.

Still, the feature that changes depending on the event's consequences is merely *what its consequences will be*. In some cases, it is misleading to focus on this retroactive change. Suppose that I see a gust of wind knock a precariously balanced rock down the side of a mountain, where it threatens to start a rockslide that could destroy a village below. If I block the rock, saving the village, I will be justifiably proud of my actions. But it would be strange for me to brag about altering the past—about altering that initial gust of wind. It's true that my actions changed that gust in a particular way, but that fact has no moral significance apart from the averted

consequences themselves. Talk of altering the past in this case is a red herring: we directly and immediately alter the event's consequences, and those consequences are what really matter.

Thinking through the impact that has on the past event is an irrelevant distraction.

But altering a past event by altering what consequences it goes on to have is not always a red herring of this sort. The significance of an event's having different consequences might go beyond those consequences themselves. To illustrate this possibility, suppose that I recklessly dropped the rock that is about to start a deadly avalanche. Suppose that I then block the rockfall, stopping it from setting off an avalanche. What I change about my earlier rock-dropping is purely a matter of affecting what consequences it has. But that does not mean that we should ignore this fact and focus on the consequences that I do or do not prevent by blocking the rocks. It is important that by changing what consequences my earlier rock-dropping has, I change whether or not I am responsible for an avalanche and the deaths of some hikers. Whether or not my reckless dropping of a rock has certain consequences is significant in ways that go beyond the consequences themselves; so there is good reason to pay attention to the fact that, in affecting those consequences, I also retroactively alter the dropping itself.

I have identified two sorts of annulment that I will not pursue further: recognition annulment, which involves discovering that something never occurred in the first place, and as-if annulment, which involves proceeding as if something has not occurred. Either of these sorts of annulment may feature in interesting, substantive accounts of criminal punishment or other response to wrongdoing; but I am after something else. I am after annulment that goes beyond uncovering what really occurred, or proceeding as if something did not occur, to really *alter* some aspect—some independently significant aspect—of past wrongdoing. I now turn to describing some forms of annulment that meet all these criteria.

III

Consider first a claim made by Herbert Fingarette about the law in his article “Punishment and Suffering.” He thinks that it is essential to the law that it is an exercise of *power* over the wills of its subjects. When the law prohibits some action ϕ , it does not merely *request* that its subjects not ϕ . It *requires* that they do not ϕ . Issuing this requirement essentially involves exercising power over the wills of those who are subject to it.

One might think that the law has a norm-giving power: a power to issue requirements (e.g., “do not ϕ ”) that its subjects are required to obey. This power—the authority of these norms—is not threatened or undermined by the fact that some citizens break the law, any more than the authority of mathematical norms is threatened by my making arithmetical mistakes.⁹⁴

Fingarette thinks the law must do more than this. It must effectively *exercise* will-restricting power over its subjects: the law must actually restrict its subjects’ wills. It ordinarily does so by giving its subjects reasons: the law’s subjects regard its prohibitions as reasons not to act in the prohibited ways. (Those reasons may not be the subjects’ actual reasons for not acting in those ways, Fingarette acknowledges—e.g., I have my own moral reasons for not murdering anyone—but, he thinks, the law’s subjects view its prohibitions as sufficient reasons for not acting in those ways, even if such a reason is a redundant backstop that plays no role in their actual deliberations.)

⁹⁴ Fingarette acknowledges this possibility only in passing, noting in a scare-quote infested passage that “one might hold that the statute, viewed as an authoritative ‘norm’ within the law as a system of ‘norms’, has validity. In some ‘normative’ sense of the term—a sense we may here leave vague—it could be held that we are ‘required’ to conform to the statute.” If that is all there is to the law’s power, Fingarette says, “one who is concerned with doctrine will say there *is* such a requirement; but one who is concerned with power, and with questions of a ‘practical’ kind, will say the ‘requirement’ is not ‘serious’, not enforced, is in practice *not* a requirement at all” (507).

Lawbreaking is a gap in the law's exercise of power over its subjects: the lawbreaker's will is not effectively restricted by the law's prohibition. Suppose, for instance, that I am subject to the law, that I know about the law's prohibition against ϕ -ing, and yet I ϕ . Then the law did not exercise its ordinary power over me. The law did not play its ordinary role restricting my will.

Unless, that is, the law subsequently and as a result of the lawbreaking exercises control over my will. Suppose that, as a result of my ϕ -ing, the law (that is, some people acting on behalf of the law, or as a part of it as an institution) incarcerates me for a period of time. That is a restriction on my will. I am restricted in what I am able to will; I cannot decide to go where I want, and so on.⁹⁵ My ϕ -ing is not, then, a gap in the law's exercise of power over my will. In ϕ -ing I did not escape from the law or throw off its shackles; rather, in ϕ -ing I was subject to its power.

Absent a punitive will-restricting response, lawbreaking would be a gap in the law's exercise of power over its subjects. Punishment—a restriction of the lawbreaker's will, imposed by the law—annuls this gap: it makes it the case that there never was any such gap in the first place. It ensures that the law exercises power even over a lawbreaker. As Fingarette puts it:

To say, then, that the law has power over my will comes down to saying that (a) whatever the law requires that I do, I generally *do* do, of my own will, and if for no other reason than that the law says so (I am subject to law, *and* I respect it), or, (b), if and when I *don't* respect the law, don't do as the law requires, the law inescapably—and as a direct consequence—constrains my will (I am in any case *subject* to law, under its power). (509, Fingarette's emphasis)

⁹⁵ Setting aside Albritton-type concerns about whether the will can be restricted at all. (See "Freedom of Will and Freedom of Action.") What Fingarette seems to have in mind is what Albritton would label a restriction on my freedom of action. In any case, if we say that my will is restricted when I treat legal prohibitions as ruling out certain courses of action for me, it is surely also restricted by incarceration, which is the important point: either way, the law exercises a similarly restrictive power over my will.

Even when someone fails to obey the law, the law's exercise of power over that person's will can be preserved. Doing so requires that the law inflict some constraint on the lawbreaker's will in response to the specific lawbreaking action.

This is genuine nullification, not just restoration. This is not a matter of restoring power over this person, power that was for a while diminished. Rather, what would have been a gap in the law's power over the lawbreaker's will is retroactively eliminated by punishment. To see this, consider what we would say to a gloating wrongdoer, giddy at her freedom from the law, if we were certain that her lawbreaking was going to result in punishment: "You think you have escaped from the law's power, but alas you have not. To be sure, you have exercised some power of your own—to determine how the law will exercise its power over you. But it is not the case that your will went unconstrained by the law. The law's power hasn't manifested quite yet, but it will. See, the constable is coming to take you away as a result of what you did."

This is not a matter of recognition annulment. Punishment does not simply reveal that there never was a gap in the law's exercise of power over the lawbreaker; absent the punishment, the law really failed to exercise power over the wrongdoer.⁹⁶

What about as-if annulment? The law is conventional; we might decide to proceed within that conventional context as if a law had not been broken. Within the legal system we might decide to proceed as if, for instance, I had not taken your wallet by force. This would require either proceeding as if there were no prohibition against taking property by force, or proceeding as if I did not take your wallet in the first place.

⁹⁶ This might be resisted if we have a strongly determinist view, on which future contingents are either true or false—in which case, it has always been true that whatever occurs was going to occur. Then it was always true that this person would be punished, and at best our punishing her reveals that fact and so reveals that there was no gap in the law's power. I set aside such worries here—but I might try to convince someone attracted to a strongly determinist view that it makes a difference that, on Fingarette's view, the punishment does not merely reveal the lawbreaker's subjection to the law. It is also the feature of the world on which that subjection depends.

The former involves annulling a gap in the law's power by, in effect, annulling the law itself, at least in a particular instance; while we may have reason to retroactively withdraw laws in some circumstances, that is clearly not an effect of hard treatment like punishment. (It also seems clear that it would be practically problematic to retroactively withdraw legal prohibitions very often; doing so would undermine the stability of those prohibitions.)

The latter option—proceeding as if, for legal purposes, I did not take your wallet—would involve making it the case that, for legal purposes, the law never was broken in the first place. Once again, this is not something that punishment accomplishes. (And, once again, the law generally foregoes this option, for good reason. A system of laws that often proceeds as if the law was not broken would not be useful for many of the purposes that we have such a system.)

Punishment accomplishes something else. It alters the context in which my taking of your wallet occurs: that theft leads to, and is the reason for, the law's restricting my will. Since my taking of your wallet takes place in this context—since it prompts a (quite direct) exercise of the law's power over my will—it is not a gap in the law's exercise of power over me.

Notably, only will-restricting punishment will accomplish this nullification. If a response doesn't involve restricting the will of the lawbreaker, there is no sense in which the law exercised a will-restricting power over the lawbreaker when she broke the law.

But that does not mean that we have a justification for will-restricting punishment. We need to ask if accomplishing this nullification justifies this response. That is, we need to understand why it is important that the law exercise this gapless power over its subjects' wills.

Fingarette's answer to this question is that, as a conceptual matter, the law *just is* an exercise of power over its subjects. To the extent that there are gaps in the exercise of the law's power, this power-exercising institution is limited.

But that just pushes the question back: why have such an institution? Why not have a different institution, one characterized by something other than the exercise of power over its subjects? Why not, for instance, have the institution which Fingarette sets aside: an institution that has the power to generate authoritative norms which its subjects are required to obey?⁹⁷

Here is a natural way of explaining why it is important that the law exercise a gapless power over its subjects: *it is important that the law's subjects not act in the prohibited ways*. That does seem to be the point of issuing prohibitions, after all: to prevent people from acting in certain ways.

But, again, punishment does not bring it about that—even from the legal perspective—the prohibited action did not occur. We might have a mechanism in the law to accomplish that some of the time; but punishment is designed to address a situation in which, within the legal system, it is acknowledged that the prohibited action did occur.

This emphasizes how generic Fingarette's conception of the law as exercise of power really is. Punishment does not annul a gap in the exercise of the law's power *to prevent wrongdoing*, because punishment does not bring it about that the person did not act in the prohibited way. Punishment only annuls a gap in the law's power *to limit its subjects' wills in some way or other*. Neither those who obey a legal prohibition nor those who violate it have a will that emerges from the law's power altogether: all roads lead to Chelm,⁹⁸ and all courses of action involve a subject's will being hindered by the law.

Once more, this does not ensure that the law's subjects do not act in prohibited ways. This exercise of power seems, then, disconnected from the very point of issuing prohibitions—of having laws! Why care about it, then? Why is it important for the law to exercise power over its

⁹⁷ Cf. Robert Hughes, "Governing the Good."

⁹⁸ See Isaac Bashevis Singer, "When Shlemiel Went to Warsaw."

subjects, when that doesn't serve the purpose of ensuring that people act in accord with the law's commands?

These questions might seem hyperbolic. Punishment—the law's imposing some restriction on a lawbreaker's will—*is* connected to ensuring that the law's subjects do not act in prohibited ways: punishment deters wrongdoing. Punishment—imposing restrictions on people because they violate the law's commands—*does* make it less likely that the law's subjects will act in prohibited ways. Such treatment, imposed on those who do break the law, is likely to *deter* people from violating those commands in the future.

The appeal to deterrence to justify punishment is familiar, and faces familiar problems. More importantly for present purposes, though, this makes the appeal to *annulment* to justify punishment a red herring. If we make this move, punishment is not important because it retroactively alters wrongdoing; it is important because it deters further wrongdoing down the line. We do not need to appeal to annulment to see why deterrence would be a valuable thing to accomplish.

Nor do we need to appeal to annulment to see why punishment deters. Punishment deters because it is a bad outcome of lawbreaking that convinces potential criminals that breaking the law in the future has a good chance of leading to similar bad outcomes. The fact that the lawbreaking is retroactively altered does not contribute to either of these. It does not contribute to the punishment's being a bad outcome. (It might, if the bad outcome that deterred criminals was *being subject to the law's power*. But the bad outcome that deters is generally something like *being locked in jail* or *forking over some money*.) The fact that lawbreaking is retroactively changed does not contribute to convincing potential criminals that future lawbreaking has a good

chance of leading to similarly bad outcomes, either. (It is consistency in the law's responses over time that is convincing in this regard, not its having responded with retroactive annulment.) We can provide just as successful a deterrent justification of punishment while setting aside its retroactive significance. The appeal to annulment adds nothing to simply saying that punishment is important because it deters lawbreaking.⁹⁹

Fingarette thinks it is logically necessary that the law, *qua* power, should punish. The properly qualified way to render this claim is: to the extent that the law does not punish, it does not exercise a will-restricting power over its subjects. I concur; but when we get clear about what that means, gaps in the exercise of that power do not appear to be so important. Even if there are gaps in this power, the law can still issue authoritative, conclusive-reason-giving commands (it retains a normative power, we might say). And annulling gaps in this power does not enhance the law's power to prevent its citizens from willing and acting in prohibited ways, except insofar as that exercise of power is also a deterrent.

The focus on exercising some power seems, in the end, petty and petulant: *I can't prevent you from willing in a morally terrible way, but I sure as hell can interfere with your will somehow!* This is to focus on the exercise of power, without regard to *stopping people from doing wrong*—without regard to what is morally important about exercising power over people in the first place.

⁹⁹ Perhaps Rawls's familiar distinction in "Two Concepts of Rules" between justifying a practice and justifying a particular action falling under the practice will help here. Perhaps the general practice of annulling gaps in the law's power over its subjects is justified by concerns about deterrence, while individual cases of annulment falling within that practice need not deter to be justified as parts of the practice.

I am doubtful that concerns about deterrence would justify us in setting up a system on which "annul gaps in the law's power over its subjects" is (one of) the rules of the criminal justice system. (Why not just have our actual criminal justice system, which doesn't involve an "annul-gaps-in-power" rule?) (At least, our practice doesn't involve such a rule on Rawls's definition of a practice, which requires that the rules that define the practice be public and taken to be definitive of the practice: see p. 24.)

I want to reiterate, here, the ingredients that make this a genuine nullification theory. I first identified a potentially problematic feature of lawbreaking: its being a gap in the law's exercise of power over its subjects. Punishment of a particular sort—the imposition of a restriction on the lawbreaker's will, because of her lawbreaking action—ensures that the law does exercise power over the lawbreaker's will with respect to her lawbreaking. I criticized an attempt to justify punishment by appeal to this annulment on the grounds that retroactively eliminating this gap in the law's exercise of power is not a morally important aim. (The punishment that accomplishes this annulment may also deter further wrongdoing, which is surely morally important; but annulment makes no contribution to that deterrence.) We do have genuine nullification here; but it has little justificatory force, because what we annul is not very significant.

IV

Next I want to take a moment to follow out some variations on Fingarette's suggestion, to see if they will fare any better.

First. The law is, Fingarette thinks, an exercise of power over its subjects' wills; wrongdoing is a gap in the exercise of that power; punishment is a way of annulling such a gap. Perhaps similar suggestions could be made about moral norms. We do, after all, think that moral norms ought to restrain the wills of those subject to them. These norms ought to control or restrict people's wills; wrongdoing is an instance in which the norms do not do so. Could

punishment or some other response manage to annul that gap—to ensure that even a wrongdoer’s will is hindered, and hindered by the moral norms?¹⁰⁰

Robert Nozick makes a suggestion along these lines in *Philosophical Explanations*. He identifies:

...three ways that correct values can have effect in our lives: (a) We can do acts because they are right or good, we can do them as right or good acts. (b) Having acted wrongly, we can repent, and give this repentance effect in our lives, performing repentant actions and so forth. (c) We can have the connection imposed upon us, via punishment.

Nozick expands on the third option, punishment:

This third alternative is worse than the others, but although less desirable it is an alternative of the same sort. It is a way, an inferior one, of falling on the same dimension on which doing something because it is right falls. That dimension is: connecting with correct values. (374-5)

Responding to correct values when one acts is one way a person can be connected with right values. So is failing to respond to correct values, and then being punished because of that failure.

The punishment maintains some connection between oneself and the values.

But this is extraordinarily abstract. A mere connection is surely insufficient. Here is another way of being connected with right values: failing to respond to them when one acts, and then being rewarded for that failure. That would be *some* connection to right values, but surely not the right kind of connection.¹⁰¹

A few lines later, Nozick adds that “through punishment, we give the correct values, *qua correct values*, some significant effect in [the wrongdoer’s] life” (375, emphasis added). The

¹⁰⁰ This suggestion might also be made about the demands I discussed in chapter 2, above. I think that this suggestion has the same problems as the proposal I consider here.

¹⁰¹ Hampton makes a similar suggestion at *Forgiveness and Mercy*, p. 113.

‘qua correct values’ phrase recurs many times in Nozick’s discussion, but he does not provide a direct explanation of what the effect of correct values qua correct values should be.¹⁰²

Rather than attempting to piece together Nozick’s remarks further, let me pursue a natural way of developing these thoughts that draws on Fingarette’s discussion.¹⁰³

We might begin by pointing out that correct values, qua correct values, ought to have the effect of limiting or restricting a person’s will. More specifically, they ought to prevent one from acting in certain ways. The wrongdoer’s wrong action is a gap in this connection between her life and correct values; the values have not restricted her will. We can ensure that those values do have *some* will-restricting impact on the wrongdoer’s life by restricting the wrongdoer’s will because of her flouting of those values. That connects the wrongdoer with correct values, qua sources of restrictions on her will.

¹⁰² Nozick does demand that the correct values on a wrongdoer have “as significant an effect on [the wrongdoer’s] life as the magnitude of his flouting these correct values” (377). He uses this to suggest that capital punishment may be the only way to match the magnitude of the flouting found in extreme crimes. But this focus on magnitude of impact itself needs explanation, and it still does not explain why the impact needs to have the specific character of punishment.

¹⁰³ There are some suggestive remarks in Nozick’s discussion of “linkage with correct values” that might help. At two points (pp. 379-80, and again at p. 385), Nozick ties the linkage he is discussing to a communicative, moral reform-focused aspect of punishment that he discusses earlier in *Philosophical Explanations*. Nozick thinks that the very same punishment which links the wrongdoer with correct values is also an attempt at communicating to the wrongdoer the values which she has violated and convincing her of the wrongness of her action. And he almost seems to say that what effects the linkage between wrongdoer and correct values is *the fact that the wrongdoer recognizes that others believe that what she did is wrong*—the fact that the wrongdoer understands the message that communicative, moral-reform punishers are attempting to convey (whether or not she believes it). Then the linkage is effected by something very different from the fact that the punishment is a restriction on the wrongdoer’s will: it is effected by the wrongdoer’s understanding of an attempted communication about right values.

If that is what Nozick has in mind, it is very obscure: how does *recognizing a communication* enact a link to the values that are the content of the communication? I recognize that someone else thinks that these values obtain. How does that amount to those values playing a role qua values in my life?

Perhaps the linkage with correct values is (as Nozick suggests on p. 384) that I *regret* what I have done, *because* of those correct values. (Perhaps I must also know that those values are the cause of my regret—perhaps that accounts for the importance of recognizing the intent of those who punish me.) But this does not seem like the right sort of connection to correct values. If I recognize the correct values, and wish (in my degeneracy) that I could have violated them even more, then I might regret my violation. That regret is caused by the correct values (and I know it is caused by them); but it is not the correct way to be linked to those values.

Nozick also concludes his discussion of retributive punishment by adding “a second line of thought, one about which I am unsure.” He suggests that, in addition to connecting a wrongdoer to correct values, punishment “is responsive (negatively) to the wrong qua wrong” (386). But this connection to wrong qua wrong is even less explained, and even more obscure, than the connection with correct values qua correct values.

This is annulment in the same sense as the Fingarette-inspired account developed above.¹⁰⁴ We ensure that correct values do have a will-restricting impact on the wrongdoer's life with regard to her wrongdoing. That is to annul a gap in her connection to correct values.

But this account faces the same question that I posed to Fingarette's account. Why is it important that correct values play some will-restricting role in one's life, if that does not involve restricting one from violating or disrespecting those values? The mere fact that there is some restriction on one's will, apart from *restricting one from violating what is valuable*, does not seem very important. In fact, apart from promoting this sort of important aim, restricting a person's will seems, in general, morally problematic.

Next I want to consider a suggestion about the law that, while similar to Fingarette's, seems to avoid the problem I have emphasized so far. In "The Right to Threaten and the Right to Punish," Warren Quinn characterizes the criminal law as a set of threats: prohibitions backed by the threat of unpleasant consequences. Then when the law prohibits ϕ , what it is really doing is threatening: if you do ϕ , then you will experience some specific unpleasant consequences.

If we don't inflict suffering on those who ϕ —if we don't carry through with the threat—it seems that it is undermined in some way. Quinn claims, for instance, that a failure to carry through on a threat when we could do so reduces a threat to a mere bluff. (See p. 339.) One might go further than this, and add that when we do inflict suffering on someone who ϕ s, we reinforce the threat in some way. Carrying through with a threat is at least evidence that the threat was serious. If these claims are correct (I examine them more closely below), then we have

¹⁰⁴ Nozick says of his account "Punishment does not wipe out the wrong, the past is not changed, but the disconnection with value is repaired (though in a second best way); nonlinkage is eradicated" (379). It seems to me that the past is changed, though in a limited way: that nonlinkage is in fact eradicated from the past.

an annulment justification for punishment: punishment is a way of retroactively eliminating what would otherwise be a gap in a system of threats. It is a way of changing something that would undermine a threat into something that (perhaps) reinforces the threat.

This seems well-positioned to avoid the problem afflicting the power-focused accounts discussed thus far. On those accounts, punishment annuls gaps in the exercise of a power; but exercising that power turned out to be morally insignificant. On the current account, however, punishment annuls a gap in a system of threats. That system of threats does seem morally important. Or, to put it a different way: on the previous account, what punishment accomplished was disconnected from the point of issuing prohibitions: preventing people from acting in certain ways. If the law is a system of threats, however, gaps in those threats seem to undermine its ability to stop people from acting in certain ways. Then punishment can, by closing gaps in those threats, help to ensure that people do not act in certain prohibited ways.

This approach depends on the claim that we are justified in issuing these sorts of legal threats. Quinn focuses on arguing that we have a right to do so. I won't engage with that question, though, because first we need to look more closely at the claim that punishment annuls a gap in this system of threats.

This requires that lawbreaking amount to a gap in a system of threats. But upon closer inspection, that does not seem to be the case. A threat is not undermined by the fact that someone acts in the way that would trigger the threat. If I say to my son, "if you hit your sister again, you'll get a time out," we do not doubt the seriousness of the threat if he hits his sister again. The mere fact that he acts in that way does not mean that what I said was any less of a threat. (The situation is different with my exercise of *power* over my son: the mere fact that he has defied me is itself a gap in my exercise of power over him.) Since there is nothing about my son's action

that undermines my threat, there is no space here for retroactively annulling the significance of that action.

The same seems to be true of an instance of lawbreaking: the lawbreaking itself does not undermine a legal threat at all. What *might* undermine the threat is a failure to carry through with the threatened consequence; but that would not be a retroactive alteration of the initial lawbreaking's significance, because the lawbreaking itself has no significance for the threat one way or another.

V

Like most discussions of nullification, Jean Hampton's discussion of "The Retributive Idea" in chapter 4 of *Forgiveness and Mercy* sometimes wavers between talk of nullification and talk of restoration. Hampton's discussion is nevertheless the most developed contemporary attempt to explain how punishment could nullify wrongdoing.

Although she presents it as a discussion of criminal punishment, like Fingarette's account, I want to emphasize at the start that Hampton's account is much broader. It may, in fact, be better positioned to justify nonlegal responses to wrongdoing than to justify criminal punishment for lawbreaking. I will return to this issue once the normative significance of the annulment Hampton describes has been made clear.

Hampton identifies three interlocking features of wrongdoing. First, wrongdoing is a form of "domination," "defeat," or "mastery" of the victim by the wrongdoer. Second, this domination has a particular expressive significance. This is drawn from Jeffrie Murphy's characterization of wrongdoing earlier in *Forgiveness and Mercy*, where he says that wrongful actions are "ways a wrongdoer has of saying to us, 'I count but you do not,' 'I can use you for

my purposes,’ or ‘I am here up high and you are there down below.’”¹⁰⁵ Third, wrongful mastery is also a form of evidence for the truth of these claims—for the wrongdoer’s superiority to his victim.

Hampton most often talks about “moral value” or “moral worth,” not moral status; she makes clear that she is supposing “a Kantian theory of moral worth, which makes people intrinsically, objectively and equally valuable” (124). This is compatible with holding that, in another sense, a wrongdoer who has committed terrible wrongs is worth much less, morally speaking, than a decent person. By fiat I will use “moral status” in what follows to refer to the first, central sense of moral worth or value.¹⁰⁶ I suppose, with Hampton, that all persons have equal moral status.

Wrongdoing is a form of domination; it is a way for the wrongdoer to assert her superior moral status; and it is a form of evidence for the wrongdoer’s superior moral status. Hampton tends to run these together when discussing the nullifying power of punishment, for instance saying that punishment can “annul the false evidence seemingly provided by the wrongdoing of the relative worth of the victim and the wrongdoer. Or to put it another way, it can annul the message, sent by the crime, that they are not equal in value” (131).¹⁰⁷

But when Hampton attempts to explain *how* this evidence or message can be annulled, she points to the other feature of wrongdoing: its being a domination, defeat, or mastery of the victim. We need to begin there.

¹⁰⁵ See p. 25); Hampton quotes this passage at p. 44) and echoes these phrases when talking about what is expressed by wrongdoing throughout chapters 2 and 4 of *Forgiveness and Mercy*.

¹⁰⁶ I choose “status” in part because it lends itself naturally to thinking about hierarchies of greater or lesser status, which is central in what follows.

¹⁰⁷ Hampton makes similar remarks about annulling a message or assertion at pp. 134 and 143, and about annulling evidence at pp. 128-9, 137, and 143.

Hampton does not provide a detailed account of what she means by “domination,” “defeat,” or “mastery.” She uses those three terms interchangeably, to describe a situation in which one person treats another in a way that “humbles the will” of the other, or makes her experience something against her will (126).¹⁰⁸ For clarity’s sake, I will restrict myself to the term ‘domination’. I take dominating someone, in the way that is central to Hampton’s account, to involve intentionally making someone experience or do something that she wills not to experience or do.¹⁰⁹

Importantly, wrongdoing often involves *one-sided* domination: the wrongdoer inflicts something on the victim, and the victim does not treat the wrongdoer similarly. This one-sidedness is central to wrongdoing’s having the expressive significance that Hampton and Murphy emphasize: the wrongdoer can assert her superiority to her victim by dominating the victim *in a way that the victim does not dominate her*. A wrongdoer who intends to assert her superiority through wrongful domination seems to be leveraging the inequality in her interaction with her victim in order to assert that the two of them are unequal in status or value.

This one-sidedness is also central to understanding the annulment that Hampton describes. She says:

To inflict on a wrongdoer something comparable to what he inflicted on the victim is to master him in the way that he mastered the victim.... Whatever mastery he can claim, she can also claim. (128)

¹⁰⁸ These are phrases that Hampton quotes, approvingly, from Fingarette. For the remainder of her discussion Hampton uses the terms “defeat,” “domination,” and “mastery” without any further elaboration.

¹⁰⁹ The “intentionally” qualification is meant to rule out cases in which I interact with someone in a way that goes against her will, although that is not what I am aiming at. This might be the case when I get in line for a ride at Disneyland, and someone gets in line behind me: my being ahead of this person does seem to go against her will (when people queue up at Disneyland, they do aim to get the earliest position in line, making every effort to do so within boundaries set by social norms). But I do not dominate her, because I am not aiming at doing something to her that goes against her will. I do not *inflict* a longer wait on her. I am just getting in line so I can ride *Dumbo*.

Certain responses to wrongdoing involve the domination of the wrongdoer by the victim. This turns the wrongful domination into one part of a larger interaction involving *equal, reciprocal* domination. Such response retroactively eliminates—it nullifies—the one-sidedness of wrongful domination.

It is not immediately clear that we can understand criminal punishment on this model. Hampton thinks of punishment as domination undertaken on behalf of the victim, by a legal authority acting as a representative of the victim. Whether this is sufficient to count as domination *by the victim* is a question I will take up shortly. In any event, this annulment can surely be accomplished directly, if the victim herself dominates the wrongdoer.

One might want to insist, here, that taken by itself the wrongdoing remains one-sided. Accomplishing this annulment does not mean that when we consider the wrongdoing in isolation, it is now reciprocal. Rather it means that it is no longer appropriate to consider the wrongdoing itself, in isolation. The response places the wrongdoing into the context of a broader interaction. Because of that response, it is incorrect and misleading to examine the wrongdoing on its own.

This might place some limits on an annulling response of the sort Hampton describes. For instance, if the wrongdoing and response are to count as two parts of a single, reciprocal interaction, it might be important that the response not happen too long after the initial wrongdoing. I will not explore such restrictions further here.

As with Fingarette's proposal, the annulment that Hampton describes is not just recognition annulment. Reciprocal domination does not simply reveal that the initial wrongful

domination was not one-sided, where that has been true all along.¹¹⁰ Rather, such domination makes it the case that the initial wrongful domination was one part of a reciprocal interaction.

Nor is this as-if annulment. The nullifying response Hampton describes does not involve merely proceeding as if the wrongful domination was not one-sided; it involves actually making it the case that the wrongful domination was not one-sided.

One might worry that this annulment will turn out to be a red herring. When the one-sidedness of wrongful domination is annulled, what we really accomplished is to alter whether or not the wrongdoing leads to some further, reciprocal domination of the wrongdoer by the victim. If that change is important only because of the significance that the further domination has, then we should set aside talk of the annulment of one-sided domination and just focus on the significance of the further domination itself.

To address this concern, we need to see why it is important to annul the one-sidedness of wrongful domination. We need to understand why we would want to turn one-sided domination into reciprocal domination. If the reason to do so is just that the domination of the wrongdoer by the victim is itself valuable in some way, then we should focus on that and set aside talk of annulling the wrongdoing's one-sidedness.

Hampton argues for the importance of the annulment she describes by appealing to the expressive and evidentiary significance that she attributes to one-sided domination. Although, as I have noted, she often treats the two as interchangeable, it is useful to consider the justificatory potential of each separately.

¹¹⁰ Unless, as I mentioned while discussing Fingarette's proposal, one has an extremely determinist view on which all future contingents are either true or false. See above, n. 96.

Consider first the expressive content of wrongdoing. According to Hampton, wrongdoing is a way of claiming that the wrongdoer is superior to her victim.

Is this dependent on the wrongdoing's being one-sided? It can be. A wrongdoer who intends to express this message through wrongful domination seems to be leveraging the inequality in her interaction with her victim in order to assert that the two of them are unequal in status or value. The inequality in the interaction is undermined if we make the defeat reciprocal and two-sided. A wrongdoer who intends to use an unequal interaction between herself and her victim to express the claim that they in fact have unequal status will be undermined if their interaction turns out to be equal.

That is not to say that punishing wrongdoing makes it impossible for wrongdoers to use their actions to express the "I am here up high and you are here down low"-type message that Hampton (and Murphy) thinks they aim to express. H. P. Grice has shown that our use of behavior to express meaning is extremely flexible: it is possible to express a wide range of meanings with a piece of behavior, regardless of its customary or natural significance. A wrongdoer who knows that her treatment of her victim will be punished, and so will not be one-sided, cannot use one-sidedness to express her message; but it is still open to the wrongdoer to defeat her victim, in a reciprocal context, and use that reciprocated domination to get her message across.¹¹¹ We cannot prevent a wrongdoer from using her actions to make a morally problematic claim.¹¹²

¹¹¹ To use a familiar Gricean analysis, the wrongdoer would need to intend (i) that her audience think that she believes herself to be superior to her victim, (ii) that her audience recognize her intention (i), and (iii) that the recognition in (ii) be part of her audience's reason for thinking that. See, e.g., Grice's "Meaning," in *Studies in the Way of Words*.

¹¹² Not to mention that, setting aside the Gricean possibilities, there is always the option of just verbally expressing the morally problematic claim.

But at the moment punishment is not being advertised as a way of *preventing* certain problematic moral claims from being made. Rather we are focusing on punishment as a way of annulling an actual false moral claim, one that has been made in a particular sort of way. Most wrongdoers do expect to be able to one-sidedly dominate their victims; some of those wrongdoers use this one-sided domination to express their superiority to their victims. We can annul this one-sidedness and so remove the resource that these wrongdoers do in fact use to make their assertions of superiority. Doing so annuls an actual communicative action that many wrongdoers do in fact undertake.¹¹³

So, for many actual wrongful actions, we can annul the claim that the wrongdoer actually does make. Does that justify the reciprocal domination required to annul the claim?

The answer seems pretty clearly to be *no*. This proposal faces several problems. First of all, as David Dolinko has argued,¹¹⁴ false moral messages do not themselves seem seriously morally problematic. Dolinko points out that the very same false moral claim can be made in many other ways. For instance, someone might publish a book asserting that men are superior to women, or asserting that he in particular is superior to his victim. If annulling the one-sidedness of wrongdoing is important because that annuls the false moral claim being made by the wrongdoer, it should seem similarly urgent to annul the same false claim made in other forms.

There is a difference between the case of wrongdoing and the cases that Dolinko describes. We have the opportunity to annul a claim made through wrongdoing because the wrongdoer uses one-sided domination to make her claim, and that one-sided domination can be annulled. But the sorts of moral claims that Dolinko describes are not made by way of one-sided

¹¹³ This might not be the case if most wrongdoers in fact expected their actions to draw a response—if few wrongdoers expected to get away with their wrongdoing. (Cf. “Freedom and Resentment,” p. 80: “The prevalence of this happy state of affairs...”)

¹¹⁴ In “Some Thoughts About Retributivism,” pp. 551-2.

domination. They are made with language, paper, and ink. We cannot annul a claim made in a book—we cannot make it so that the claim was never successfully made in the first place—in the way that we can annul a claim made by way of one-sided domination.

Something structurally similar to annulment is sometimes possible with claims made in books, though. Sometimes we can *prevent* such claims from being made—e.g., prevent the wrongdoer’s book from being published in the first place. Preventing a claim from being made and annulling the claim after the fact accomplish the same thing: both bring it about that the claim does not occur. If annulling a wrongdoer’s false claim is deeply important, it should seem similarly important to prevent false claims from being made in the first place.

But it does not seem like a serious moral accomplishment to prevent a false moral claim from being made. There are other ways of addressing false moral claims. We might, for instance, *argue* that the persons in question are in fact equal; or we might *reject* that claim (in a dramatic fashion that both Hampton and Dolinko discuss) by throwing a ticker-tape parade for those whose inferiority is asserted. Doing so would serve as a dramatic reassertion of people’s status in the face of an assertion that they are of lesser status. These responses fall short of preventing or annulling a false moral claim, but they do seem to undermine it effectively.

For these reasons, the mere fact that a false moral claim has been made does not seem seriously morally problematic. At least, it fact does not seem problematic enough to justify acting to prevent such claims from being made in the first place, or acting to annul them after the fact.

Hampton does confront these problems. When she does, she appeals to yet another feature of wrongdoing: its being *evidence* for the victim's inferiority to the wrongdoer.

Discussing the alternative of throwing a victim of wrongdoing a ticker-tape parade, she says:

Suppose we gave a victim a ticker-tape parade after the crime to express our commitment to his value. Still the fact that he had been mastered by the wrongdoer would stand. He would have lost to her, and no matter how much the community might contend that he was not her inferior, the loss counts as evidence that he is. Hence *the victim wants the evidence nullified*. And punishment is the best way to do that. The wrongdoer can't take her crime to have established or to have revealed her superiority if the victim is able to do to her what she did to him. The punishment is therefore a second act of mastery that negates the evidence of superiority implicit in the wrongdoer's original act. (*Forgiveness and Mercy*, pp. 128-9; Hampton's emphasis)

We might not worry about a wrongdoer's false moral message on its own; but, Hampton appears to think, the existence of evidence for the victim's inferiority to the wrongdoer is more seriously morally problematic. We might dramatically reject the wrongdoer's message by throwing his victim a ticker-tape parade; but, according to Hampton, that will not change the fact that there exists problematic evidence for the victim's inferiority to the wrongdoer.

The natural question to ask at this point is why such evidence is more seriously morally problematic than the mere existence of a false moral message. But I will set that question aside, because it is not clear how any such evidence is provided by wrongdoing in the first place. More specifically, it is not clear why such evidence is provided by one-sided wrongful domination, and retroactively eliminated by the nullifying response that Hampton describes. How is it that one-sided domination is evidence of a victim's inferiority, but reciprocal domination is not?

Hampton treats the evidentiary force of wrongdoing as closely connected to, if not identical with, its expressive content.¹¹⁵ Perhaps the one-sided domination of the victim by the wrongdoer is evidence for the victim's relative inferiority in virtue of the fact that it is an assertion that she is inferior to the wrongdoer. That is to treat the fact that someone has asserted "A is inferior to B" as evidence for A's being inferior to B.

But it does not always seem right to treat such an assertion as evidence for its truth. Why shouldn't we instead treat the assertion as evidence that the wrongdoer is deeply mistaken about her worth relative to the victim, and perhaps more generally unreliable in moral matters?

More generally, it seems that we approach the wrongdoer's treatment of the victim from a position of deep confidence that the wrongdoer is wrong. It is in fact true that the victim and the wrongdoer have equal moral status. But then it is difficult to see why anything the wrongdoer does should look like evidence to the contrary. Against a background conviction that people do have equal moral status, we ought to understand the wrongdoer's actions in ways that do not conflict with that conviction: as evidence that the wrongdoer is mistaken, or vicious, or otherwise unreliable in moral matters.

If this background conviction that people have equal moral status were itself weakened, what the wrongdoer does might seem to be evidence for the victim's having lower status than the wrongdoer. If, for instance, one held that human status is correlated with strength, then one might think that a defeat at the hands of another demonstrates that one is lower in status than that other person. On this view of human nature, an unanswered defeat reveals that one has less strength, and hence lower status, than the person at whose hands one has been dominated.

¹¹⁵ E.g.: punishment "*can annul the false evidence seemingly provided by the wrongdoing of the relative worth of the victim and the wrongdoer*. Or to put it another way, it can annul the message, sent by the crime, that they are not equal in value" (*Forgiveness and Mercy*, p. 131; Hampton's emphasis)

If one accepted this account of human status, one-sided domination would be evidence for a victim's inferiority. Reciprocated domination would, by contrast, be evidence for equality of strength and so for the equal status of the victim and wrongdoer. Then annulling the one-sidedness of the wrongdoing would annul the problematic evidence provided by wrongdoing.

The question is why we should accept this account of human status. We believe that the wrongdoer has done wrong precisely because we believe in an egalitarian theory of human worth, on which all people have equal worth independent of considerations like how strong they are.

In later work, Hampton suggests that, while we believe in an egalitarian view of human worth, we *fear* that some different theory of human worth obtains, such that our value is lower or vulnerable to being lowered in a way that is ruled out by the egalitarian theory of human worth that we reflectively endorse.¹¹⁶ This seems psychologically plausible, especially with members of oppressed groups who have been treated in ways that reinforce that oppression. Hampton is particularly concerned with victims of this sort, including African-Americans injured or murdered by whites in the antebellum South¹¹⁷ and female rape victims.¹¹⁸

It is surely true that some people—especially some victims—have these sorts of fears. On the theory of human worth that these victims fear is true, then, their having been mastered by a wrongdoer is evidence for their having lower value. Reciprocally mastering the wrongdoer would, by the lights of this inegalitarian theory of human worth, annul that evidence. It is

¹¹⁶ E.g.: “Our views about human value are more or less secure, and we are more or less committed to them. Fear that we are worth less than we wish (or perhaps less than others think we are worth) is a common human phenomenon, particularly in societies in which non-Kantian inegalitarian theories of worth have gained currency” (Hampton, “Correcting Harms Versus Righting Wrongs: The Goal of Retribution,” 1678).

¹¹⁷ “Correcting Harms” p. 1675 ff.

¹¹⁸ See Hampton, “Punishment, Feminism, and Political Identity: A Case Study in the Expressive Meaning of the Law.”

possible to establish that even if their fears are true, the victims of wrongdoing are not lower in status than the wrongdoer.

But should we engage with this inegalitarian theory of human worth? When we do so, we do not challenge the incorrect theory of human worth itself; instead we work within it. But the theory is itself morally problematic, in ways that extend beyond a single case of inequality. Anyone who is in fact less strong is, on this theory, morally inferior. As Dolinko points out, it then becomes hard to understand criminal punishment as annulment of this evidence: it is the criminal justice system as a whole, not the victim herself, who is shown to be equal in strength to the wrongdoer. Accepting an incorrect theory of worth might also seem to be dangerously liable to bestow legitimacy on such views of human moral status.

It is difficult to maintain that wrongdoing provides problematic evidence for the victim's inferiority. But it still does seem legitimate for a victim to feel threatened by past wrongdoing¹¹⁹ (itself, apart from any danger of further wrongdoing). And it seems that punishment and other reciprocal responses can address the victim's concerns. If we account for these intuitions in terms of evidence, though, they are in tension with the fact that these fears seem to be grounded in an incorrect theory of human moral status, one that even these victims might reflectively reject. I think there is a better way of understanding the victims' concern, one that does not encounter this tension.

VI

Hampton does describe a clear, specific form of annulment: annulment of the one-sided domination involved in wrongdoing. The next step is to explain why this is important to

¹¹⁹ Cf. Hieronymi's description of wrongdoing as "a past action that persists as a present threat" ("Articulating An Uncompromising Forgiveness," p. 546).

accomplish. Hampton tries to do so by pointing to some further features of one-sided annulment: it can be used by the wrongdoer to assert her superiority to the victim, and it can be a form of evidence for the victim's inferiority to the victim. I have argued that both of these strategies face serious difficulties.

I think it is better to take a more direct tack: one-sided domination is itself morally problematic, such that we have strong reason to eliminate it independent of any further problematic features it may have. This reason is grounded in important facts about the victim's moral status. If this is true, and if—additionally—reciprocal domination is not similarly morally problematic, then we have a reason to annul wrongdoing in the way Hampton describes: doing so makes the wrongdoing less morally problematic.

But why think that? Why think that one-sided domination is morally problematic but reciprocal domination is not? The original wrongdoing still involves domination of the victim. We just add more domination to the mix. If that original domination was morally problematic, why not think that we've just added yet more morally problematic domination?¹²⁰

Here is a suggestion: domination itself is not problematic. Instead, *one-sided, unequal* domination is problematic. This domination—this unequal interaction—is incompatible with the victim's and wrongdoer's being morally equal. If the wrongdoer and the victim have *equal* moral status, their interactions should not involve an *unequal, one-sided* interaction.

If two people dominate each other equally, I suggest, their interaction is not in tension with their having equal value. I am *not* suggesting that when two people dominate each other equally, their interaction is morally unproblematic. Other features of their interaction may very well be seriously morally problematic.

¹²⁰ Cf. my objection to the restorative proposal discussed in chapter 4.

Consider, for instance, a sexist man who violently assaults a woman. Part of the problem with this interaction is that it is unequal, incompatible with the woman's being of equal status with the man; a more serious problem with it is the serious physical injury that the woman suffers. Nobody should intentionally inflict that kind of suffering on a person against her will, whether it is reciprocated or not.

Ensuring that this wrongdoing involves reciprocal domination is an improvement along one dimension: doing so brings the interaction in line with what is required by the equal relative status of the victim. But acting to improve the wrongdoing in this way might be disturbing. For the woman to dominate the man as much as he dominates her, she might need to inflict just as much suffering on him as he inflicted on her. And intentionally inflicting that much suffering seems impermissible no matter what other improvements doing so brings.

I think that is correct: sometimes other considerations outweigh the importance of bringing wrongdoing in line with what is required by the victim's equal status. But not always. Consider another case: suppose a colleague betrays my trust. Suppose she shares an embarrassing secret that I had told her in confidence, as a way of getting a laugh out of some of my peers. I have been made to suffer some embarrassment, and had some fact about me made public, against my will. This is problematic in part because of the unequal, one-sided nature of our interaction. It is also problematic because I should not be made to suffer in those ways at all. In order to reciprocally dominate my colleague, I will have to inflict some suffering on her. I will have to add yet more suffering to our interaction. But in this case the suffering itself is comparatively mild. It is much more plausible that the reason I have to make our interaction equal is stronger than the reason I have to avoid causing more suffering of this mild sort.

Note also that reciprocal domination requires, in the first instance, that I dominate my colleague as much as she dominates me. I must inflict something on her that goes against her will as much as what she did to me went against my will. This does not require literal similarity—it need not involve sharing an embarrassing secret of hers behind her back, for instance. (Depending on how my colleague feels about having her secrets shared, the literal response might very well involve too little domination to render our interaction equal—or too much.)

The principle that equal status rules out specifically unequal domination can, I think, be applied in both the violent-assault and the betrayal-of-confidence cases—but only in the latter case are we justified in acting to address the unequal domination. I think that this principle captures our intuition that unequal domination is problematic in a way that reciprocal domination is not; but this is far from a knockdown argument for that principle. Let me provide two more pieces of indirect support for it.

First recall the intuition that I was struggling to do justice to at the end of the previous section, the intuition that prompts Hampton to describe wrongdoing as evidence for the victim's inferiority. It does seem legitimate for a victim to feel troubled or even threatened by the fact that the wrongdoer's domination of her has gone unanswered. Hampton tries to explain this by describing the wrongful domination as evidence for the victim's inferiority. But it is difficult to explain how wrongful domination is evidence for the victim's inferiority while also doing justice to our conviction that the victim is *not* in fact inferior to the wrongdoer. To make the wrongful domination into such evidence, we have to give up or suspend the egalitarian theory of human worth that grounds our confidence in the victim's equality with the wrongdoer.

The principle I am suggesting holds that the victim's equality with the wrongdoer rules out unequal domination. If that is true, we can explain why a victim might feel troubled by

wrongdoing and vindicated or satisfied by a reciprocal response. The wrongdoing is troubling to the victim because it is incompatible with her having equal status with the wrongdoer; when the wrongdoing is reciprocated the victim is satisfied, because this brings her interaction back into line with what is required by their equal status.

In addition, if it is possible for the state to act to annul the inequality of wrongdoing through punishment, a victim might feel additionally aggrieved or threatened by a decision not to punish. If the state recognizes that the victim's equal status rules out unequal treatment, and it is in a position to annul that unequal treatment through punishment, a decision not to do so may indicate that the state does not consider the victim to be the wrongdoer's equal. (A decision not to punish may instead be a response to some countervailing reason not punish: e.g., that punishing in order to annul would involve inflicting suffering that is morally beyond the pale.)

Second, the principle I am suggesting seems to provide the best explanation for the assertion that Hampton and Murphy focus on, the wrongdoer's assertion of superiority to her victim. The wrongdoer is not correct, of course; but there is something natural about using wrongdoing to make that assertion. We can make sense of that against the background of the principle that unequal domination is ruled out by equality of status. If one wants to make some moral claim with one's actions, it is natural to do so by acting in a way that is permissible only if that moral claim is true. (So, for instance, one natural way of asserting that cursing is acceptable in a UCLA dissertation would be to write "damn" in my dissertation.) Unequal, one-sided domination seems like a natural way for a wrongdoer to assert her superiority over her victim, for a similar reason: this unequal, one-sided domination is permissible only if the dominator is in fact of superior status to her victim. Equal status rules out this sort of domination.

If equal status does not rule out *reciprocal* domination in the same way, we would expect it to seem much more strained to try to make the same assertion through domination that one expected to be reciprocated.¹²¹ And that does seem to be reflected in specific cases. Consider again the sexist man who violently assaults a woman. Suppose he does so in part to assert his moral superiority to her. If the sexist expects this woman to answer his domination in kind—if he expects her not just to fight back, but to dominate him just as he dominates her—it is much less natural to see his action as an assertion of superiority. It is much less natural to see an equal exchange as an assertion of superiority.

These are not conclusive considerations. If we brought back the view of moral status on which it is correlated with strength or the ability to dominate, we could make sense of the wrongdoer's strategy: if one accepted such a theory, it would be natural to try to assert superiority over someone by one-sidedly dominating her—thereby demonstrating one's superior strength. But again this makes the assertion out to be sensible only to the extent that one accepts some incorrect theory of moral status. We can understand that the wrongdoer would try to make an assertion in the way that she does, because she accepts a false theory of moral status. And we can understand that the victim might be troubled by that attempted assertion, if the victim also accepts that false theory of moral status. But to the extent that we are confident in an egalitarian theory of moral status, we should be confident that the wrongdoer and the victim are both mistaken. If we are confident in an egalitarian theory of moral status, it is difficult to see why it would be important to undermine the wrongdoer's assertion; it is already undermined, by the fact that it is grounded in an incorrect theory of moral status.

¹²¹ Again, it would not be impossible to do so; the human ability to convey meaning through action is extremely flexible; but it would be much less natural to do so when one expected reciprocation.

If, instead, we understand the wrongdoer's assertion in the way that I am recommending, the wrongdoer is not leveraging an incorrect theory of moral worth to make her assertion. Her assertion does not rely on an appeal to any particular theory of moral worth; instead, it draws on the fact that if two people have equal moral status, their interaction should not be one-sided in certain ways. We only need to grant that fact—we do not need to posit any misunderstanding about human worth generally—in order to understand why the wrongdoer's assertion is sensible. Similarly we only need to grant this fact to see why a victim would be troubled by the wrongdoer's assertion: even if the victim is confident that people have equal status and that the wrongdoer's assertion is false, the wrongdoer makes the assertion by acting in a way that is in conflict with the victim's equal status. The victim might be troubled by the fact that her interaction with the wrongdoer is not compatible with her equal status, even though she is confident that the wrongdoer's assertion is false.

These are not knockdown considerations; but from this point forward I will suppose that this principle is true—that one-sided domination, but not reciprocal domination, is morally problematic because it is ruled out by the victim's equal status with the wrongdoer.¹²² I make this supposition in order to look more closely at the use of this principle to support a nullifying response to wrongdoing.

¹²² I am also uncertain how far to apply this principle. When the state requires that I pay it some portion of my income, that may go against my will. Does the state dominate me, then? Is this problematic one-sided domination? Or is the state in fact morally superior to me, so that this domination is acceptable? (Alternatively, attempts in the social contract tradition to explain how citizens do, or would, consent to the state's wielding authority might—perhaps—undermine the thought that the state's commands go against my will.) What about circumstances in which a boss commands an employee to do something that the employee does not want to do; is that a form of domination?

I set these questions aside; within the context of individuals interacting with and wronging each other as individuals, outside of specific institutional roles, at any rate, there is something problematic about one-sided domination, which is best explained by the equality of the two individuals qua individuals.

Consider some wrongful one-sided domination of a victim who does in fact have equal status with the wrongdoer. On the basis of the principle I have proposed, we can reason as follows:

1. Equal moral status is incompatible with unequal domination.
2. If a victim has equal moral status with a wrongdoer, and equal moral status is incompatible with unequal domination, then unequal domination is seriously morally problematic.
3. If unequal domination is seriously morally problematic, then there is a strong moral reason to eliminate it in favor of something that is not similarly morally problematic.
4. We can change unequal domination into reciprocal domination after the fact, by reciprocally dominating the wrongdoer.
5. Equal moral status is not incompatible with reciprocal domination.
6. So when wrongdoing involves unequal domination of a victim who does have equal status with the victim, there is a strong moral reason to reciprocally dominate the wrongdoer.

Claims (1) and (5) state the central moral principle I have been discussing. Claims (2) and (3) move from the wrongdoing's being incompatible with equal status to our having a strong reason to eliminate it.

Consider an apparent counterexample to claim (2). Here is something else that might be incompatible with someone's having equal status: *asserting* that one is superior to her. As Dolinko has already pointed out, these mere assertions of superiority do not seem deeply morally problematic.

This involves a different sense of incompatibility. If two people do in fact have equal status, it is *incorrect* to say that one is superior to the other. But equality of status does not rule out making such a claim. If it is against the law to turn right at a red light, doing so is not permitted by the law; asserting that we can turn right on red is not similarly ruled out by the law,

although it is incompatible with the law in another sense. It is the ruling-out incompatibility that I am trying to capture here.

In (4) I point to a way of responding to the strong moral reason described in (3). We can eliminate one-sided domination, by changing it into one-sided domination into reciprocal domination. Drawing on Hampton's discussion of annulment, (4) identifies a way of bringing this about after the fact: we can retroactively annul the one-sidedness of wrongful domination, by dominating the wrongdoer in turn. If, as I am supposing, (5) is true—if reciprocal domination is not similarly morally problematic—then we have a strong moral reason, grounded in the victim's equal moral status, to bring about this retroactive alteration of the wrongdoing.

When I first presented Hampton's description of annulment, I wondered whether it would turn out to be a red herring: whether the annulment would only amount to changing the consequences of the past wrongdoing, in such a way that we could justify changing those consequences without reference to the alteration of the earlier event. If that were the case, then we could justify the response equally well without describing it as annulment, by focusing directly on the altered consequences.

On the account I have developed here, annulment *is* just a matter of altering the consequences of the wrongdoing: the annulment makes it the case that the wrongdoer's domination of the victim elicits the victim's domination of the wrongdoer. Could we justify this annulling response equally well without appealing to the retroactive change that makes to the original wrongdoing? The answer is no: to justify altering the consequences of wrongful domination, we appeal to the fact that *doing so improves the original wrongdoing*. Responding with reciprocal domination is important because then the original wrongdoing is no longer one-

sided, and so is not in tension with the equal moral status of the victim and the wrongdoer. To understand this moral improvement, we must look to the retroactive alteration to the initial wrongdoing. In this case, talk of retroactive annulment is not a red herring.

Another long-delayed issue should be addressed here as well. The annulment that Hampton describes, which I am defending, requires that a victim dominate a wrongdoer. That is what ensures that their interaction does not involve problematically one-sided domination. But can third parties respond on the victim's behalf and create equality in the interaction? Can criminal punishment, in particular, accomplish this sort of nullification? Dolinko has raised doubts about this, asking:

But how, precisely, does the state's 'defeating' the wrongdoer convey...as Hampton puts it, 'the experience of defeat *at the hands of the victim*'? How does it get across that the victim has evened the score and can now claim whatever mastery the wrongdoer can? It would seem, rather, that while the wrongdoer claimed 'superiority' by defeating the victim himself, a whole gang of partisans of the victim has now banded together and defeated the hopelessly outnumbered wrongdoer! Perhaps this conveys the message that society as a whole is the equal (or perhaps the master) of the wrongdoer, but it hardly seems an apt way of expressing the message that the victim, individually, is the wrongdoer's equal. (553)

Dolinko points out that the domination of the wrongdoer by "a whole gang of partisans" who band together and "hopelessly outnumber" her does not seem to convey that the victim is equal to the wrongdoer. That is true, from a perspective that focuses on the victim's and wrongdoer's relative *strength*. But we have set aside a focus on strength. Annulment is important because it ensures that wrongdoing is compatible with the victim's equal standing with the wrongdoer. What that requires is not equality of strength but equality of domination: each must impose equally on the other's will. It seems at least plausible that when a whole gang of partisans gang together and defeat the wrongdoer—suppose that they do so in the victim's name, because of the

wrongful treatment of the victim, and at the victim's behest ¹²³—we can say that the victim has dominated the wrongdoer.

At least the present annulment account provides some reason for the victim to respond to wrongdoing herself, by reciprocally dominating a wrongdoer. By doing so she can bring things into line with what is required by their equal moral status. Again, this will not always be a justifiable way of responding; when we consider responding to a serious, violent assault, for instance, it seems likely that the reciprocal domination will be morally problematic in other ways even though it brings the interaction in line with what is required by the victim's equality.¹²⁴ But this problem does not arise with less violent forms of wrongdoing, that involve less serious forms of suffering. I suspect that many familiar, noncriminal, day-to-day wrongs fall into this latter category. And more generally we have an explanation for the appeal of reciprocal responses to wrongdoing, even when that appeal is outweighed by countervailing considerations: such a response brings the wrongdoer-victim interaction into line with what is required by the victim's equality with the wrongdoer.

¹²³ I am tempted by some further considerations—involving deeper issues in political philosophy with which I cannot adequately engage here—that might help to support applying the present annulment story to criminal punishment. Perhaps we can identify the victim as a member of the social, legal, or political institution, and so identify its actions on her behalf as being, in a direct sense, her actions. There is also, as Dolinko hints, the possibility of taking society as a whole to be the victim of the wrongdoing.

¹²⁴ Suppose that punishment is in fact a way of annulling one-sided domination on a victim's behalf. If that is correct, we might be able to extend the range of wrongdoing which can permissibly be annulled. When I discussed the case of a sexist man who violently assaults a woman, I argued that the woman's reciprocally dominating the man would be impermissible even though it would improve the wrongdoing by bringing it in line with what is required by their equal status. Dominating the wrongdoer as much as he dominated his victim seemed to require imposing impermissible levels of suffering on the wrongdoer.

I suspect, tentatively, that criminal punishment manages to, or has the potential to, involve serious domination without being as morally problematic as individually imposed domination. Incarceration, for instance, is a serious exercise of the state's power to dominate a wrongdoer; but the suffering of being incarcerated does not seem clearly, inevitably morally problematic in the way that being the object of violence is problematic. These are extremely tentative thoughts, however. I am not certain why the domination a state imposes through criminal punishment appears less morally problematic than other forms of domination. And clearly our actual penal system often inflicts morally problematic suffering on criminals. This may be an unrealistically sanitized view of what incarceration could be like.

CONCLUSION

It is time to return to antagonism. In chapter 5, I argued that we are sometimes able to annul a problematic feature of wrongdoing: we are sometimes able to act in a way that improves the past. Doing so requires that the wrongdoer be dominated, which will often involve making her suffer. But does this require antagonism? Does it require taking the wrongdoer's suffering to be valuable in itself?

No. The wrongdoer's suffering is, in this context, instrumentally valuable. It is valuable as a way of annulling a problematic feature of wrongdoing. Now in some cases, we might deploy specific antagonistic attitudes, like resentful anger, to annul the one-sidedness of wrongful domination in the way described in chapter 5. Expressing angry resentment at a wrongdoer might be a way of inflicting something on her against her will (most people want to avoid being the object of such anger, after all), and so a way of dominating her as she dominated her victim. In the case of everyday, interpersonal wrongdoing—betrayals of trust, rudeness, and the like, rather than extreme or criminal wrongdoing like murder or rape—this domination-by-antagonism might be equal to the wrongdoer's domination of the victim. Thus antagonistic response to wrongdoing might sometimes be a way of annulling the one-sidedness of wrongful domination.

But once again antagonism of this sort is not *required* to accomplish annulment; it is just one way of doing so, in some circumstances. We might dominate the wrongdoer in some other way altogether, that involves no anger whatsoever. The Gandhi-King ideal is an expression of a deep moral skepticism about antagonism; it is difficult to see how identifying a way in which antagonistic response might be morally helpful, but not required, will answer these deep-rooted moral concerns.

This dissertation is a story of failed attempts to explain why we need antagonistic response to wrongdoing. In chapters 2 and 3, I argued that no such responses are required by any of our deep moral commitments. In chapter 4, I argued that such response is not required to restore equality of attitudes to interpersonal relationships. Finally, in chapter 5, I argued that an important form of annulment, involving domination of the wrongdoer, is justified; but, as I have pointed out, this annulment does not require antagonism.

I want to emphasize that this failure seems *strange*. I can locate no rational ground for the intuitive thought that antagonistic responses to wrongdoing are a necessary part of our social lives. These responses are not essential to deeply important aspects of our lives; they are not essential to accomplishing important ends. At best, it is an unfortunate accident that antagonism is connected, in our contingent human psychology, to other things of value.

Perhaps I am mistaken, and antagonism is essential in some other way I cannot perceive. But as things stand, it appears that King and Gandhi come very close to the truth of the matter when they reject antagonism. Contingent features of human psychology may mean that we could not put this rejection into action without giving up on important aspects of human life; but I can see no reason not to work towards weakening these connections, with the hope of opening the possibility of eliminating antagonism from our lives.

BIBLIOGRAPHY

- Albritton, Rogers. "Freedom of Will and Freedom of Action." *Proceedings and Addresses of the American Philosophical Association* (1985).
- Aristotle. *Nicomachean Ethics Books VIII and IX*. Ed. Lindsay Judson and J L Ackrill. Trans. Michael Pakaluk. Oxford: Clarendon Press, 1998.
- Barr-Zisowitz, Carol. "'Sadness'—Is There Such a Thing?" *Handbook of Emotions*. Ed. Michael Lewis and Jeannette M Haviland-Jones. 2nd ed. Guilford Press, 2004.
- Braithwaite, John. *Crime, Shame and Reintegration*. Cambridge University Press, 1989.
- . "Repentance Rituals and Restorative Justice." *Journal of Political Philosophy* 8.1 (2000): 115-131.
- California Family Code. Division 6. Nullity, Dissolution, and Legal Separation. 2012. Web. <<http://leginfo.legislature.ca.gov/faces/home.xhtml>>
- Colorado Revised Statutes. Title 14. Domestic Matters. 2012. Web. <<http://www.lexisnexis.com/hottopics/colorado?source=COLO;CODE&tocpath=1CPZ9BIKBUIUS4JDA,2TSK7VSJKCDMNR4Z,3PU9SGWRV5JCLC5XO>>
- Cottingham, John. "Varieties of Retribution." *Philosophical Quarterly* 29.116 (July 1, 1979): 238-246.
- Cunningham, Michael R. "What Do You Do When You're Happy or Blue? Mood, Expectancies, and Behavioral Interest." *Motivation and Emotion* 12.4 (December, 1988): 309-331.
- D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61.1 (2000): 65-90.
- . "The Significance of Recalcitrant Emotion (or, Anti-quasijudgmentalism)." *Royal Institute of Philosophy Supplement* 52 (2003): 127-145.
- Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. American Psychiatric Association, 2000.
- Dolinko, David. "Some Thoughts About Retributivism." *Ethics* 101.3 (May 1, 1991): 537-559.
- Feinberg, Joel. "The Expressive Function of Punishment." *Doing and Deserving*. Princeton, NJ: Princeton University Press, 1970. 95-118.
- Fingarette, H. "Punishment and Suffering." *Proceedings and Addresses of the American Philosophical Association* (1977).

- Fischer, Louis. *Gandhi and Stalin: Two Signs at the World's Crossroads*. London: Harper & Brothers, 1947.
- Frankfurt, Harry. "The Importance of What We Care About." *The Importance of What We Care About*. Cambridge University Press, 1988. 80-94.
- French, Peter. *The Virtues of Vengeance*. Lawrence, Kansas: University Press of Kansas, 2001.
- Grice, Paul. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, 1991.
- Grossman, Dave. *On Killing: The Psychological Cost of Learning to Kill in War and Society*. Revised ed. Back Bay Books, 2009.
- Hampton, Jean. "Correcting Harms Versus Righting Wrongs: The Goal of Retribution." *UCLA Law Review* 39.6 (1992): 1659-1702.
- . "Punishment, Feminism, and Political Identity: A Case Study in the Expressive Meaning of the Law." *Canadian Journal of Law and Jurisprudence* XI.1 (February 20, 1998): 23-45.
- Hampton, Jean, and Jeffrie G Murphy. *Forgiveness and Mercy (Cambridge Studies in Philosophy and Law)*. New York, NY: Cambridge University Press, 1988.
- Hegel, Georg Wilhelm Fredrich, and Allen W Wood. *Hegel: Elements of the Philosophy of Right (Cambridge Texts in the History of Political Thought)*. Trans. H B Nisbet. Cambridge University Press, 1991. 566.
- Hieronymi, Pamela. "Articulating An Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62.3 (June 1, 2001): 529-555.
- Honderich, Ted. *Punishment: The Supposed Justifications Revisited*. Ann Arbor, MI: Pluto Press, 2006.
- Hubbard, Julie A, Lydia J Romano, Meghan D McAuliffe, and Michael T Morrow. "Anger and the Reactive-Proactive Distinction." *International Handbook of Anger*. Ed. Michael Potegal, Gerhard Stemmler, and Charles Spielberger. Springer, 2010. 231-242.
- Hughes, Robert. "Governing the Good." Dissertation. UCLA, 2010.
- Illinois Compiled Statutes. Chapter 750. Rights and Families. §5. Marriage and Dissolution of Marriage Act. 2012. Web.
<<http://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=2086&ChapterID=59>>
- King, Martin Luther. "Letter From Birmingham Jail." *UC Daves Law Review* 26 (1992): 835-851.
- . "Nobel Prize Acceptance Speech." 1964. Web.
<http://www.nobelprize.org/nobel_prizes/peace/laureates/1964/king-acceptance.html>.
- . "Nobel Prize Lecture." 1964. Web.
<http://www.nobelprize.org/nobel_prizes/peace/laureates/1964/king-lecture.html>.

- Komarovskaya, Irina, Shira Maguen, Shannon E McCaslin, Thomas J Metzler, Anita Madan, Adam D Brown, Isaac R Galatzer-Levy, Clare Henn-Haase, and Charles R Marmar. "The Impact of Killing and Injuring Others on Mental Health Symptoms Among Police Officers." *Journal of Psychiatric Research* 45.10 (October 1, 2011): 1332-1336.
- Kutz, Christopher. "Justice in Reparations: The Cost of Memory and the Value of Talk." *Philosophy and Public Affairs* 32.3 (2004): 277-312.
- Litz, Brett T, Nathan Stein, Eileen Delaney, Leslie Lebowitz, William P Nash, Caroline Silva, and Shira Maguen. "Moral Injury and Moral Repair in War Veterans: A Preliminary Model and Intervention Strategy." *Clinical psychology review* 29.8 (December, 2009): 695-706.
- Mackie, J L. "Morality and the Retributive Emotions." *Criminal Justice Ethics* 1.1 (1982): 3-10.
- Maguen, Shira, David D Luxton, Nancy A Skopp, Gregory A Gahm, Mark A Reger, Thomas J Metzler, and Charles R Marmar. "Killing in Combat, Mental Health Symptoms, and Suicidal Ideation in Iraq War Veterans." *Journal of Anxiety Disorders* 25.4 (June 1, 2011): 563-567.
- Maguen, Shira, Thomas J Metzler, Jeane Bosch, Charles R Marmar, Sara J Knight, and Thomas C Neylan. "Killing in Combat May Be Independently Associated with Suicidal Ideation." *Depression and Anxiety* (May 13, 2012).
- Miller, William Ian. *Eye for An Eye*. Cambridge University Press, 2007.
- Moore, Michael. "The Moral Worth of Retribution." *Responsibility, Character, and the Emotions*. Ed. Ferdinand Schoeman. Cambridge University Press, 1988. 179-219.
- Morris, Herbert. "Guilt and Suffering." *Philosophy East and West* 21.4 (1971): 419-434.
- . "A Paternalistic Theory of Punishment." *American Philosophical Quarterly* 18.4 (October 1, 1981): 263-271.
- . "Persons and Punishment." *The Monist* 52.4 (1968): 475-501.
- Murphy, Jeffrie G. *Getting Even: Forgiveness and Its Limits*. Oxford University Press, 2003.
- New York Domestic Relations Code. Articles 2-3. Marriages; Solemnization, Proof and Effect of Marriage. 2012. Web.
<<http://public.leginfo.state.ny.us/LAWSSEAF.cgi?QUERYTYPE=LAWS+&QUERYDATA=@LLDOM>>
- Nozick, Robert. *Philosophical Explanations*. Cambridge, MA: Harvard University Press, 1981.
- Orwell, George. "Reflections on Gandhi." *The Collected Essays, Journalism, and Letters of George Orwell*. Ed. Sonia Orwell and Ian Angus. Vol. IV. New York: Harcourt Brace Jovanovich, 1968. 463-470.

- Pereboom, Derk. *Living Without Free Will*. Cambridge University Press, 2006.
- Quinn, Warren. "The Right to Threaten and the Right to Punish." *Philosophy and Public Affairs* 14.1 (1985): 327-373.
- Quinton, A M. "On Punishment." *Analysis* 14.6 (1954): 133-142.
- Rawls, John. "Two Concepts of Rules." *Philosophical Review* 64.1 (1955): 3-32.
- Sher, George. *In Praise of Blame*. Oxford University Press, USA, 2006.
- Simon, Neil. *The Odd Couple*. 1965.
- Singer, Isaac Bashevis. *Stories for Children*. New York, NY: Farrar, Straus and Giroux, 1985.
- Strawson, Peter. "Freedom and Resentment." *Free Will*. Ed. Gary Watson. 2nd ed. Oxford, UK: Oxford University Press, 2003. Oxford Readings in Philosophy.
- Wallace, R Jay. *Responsibility and the Moral Sentiments*. Harvard University Press, 1998.
- Watson, Gary. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." *Free Will and Reactive Attitudes*. Ed. Michael McKenna and Paul Russell. Ashgate, 2004.
- Wood, Allen W. *Hegel's Ethical Thought*. Cambridge University Press, 1990.
- Wood, D. "Retributive and Corrective Justice, Criminal and Private Law." *Scandinavian Studies in Law* 48 (2005).