

INTENTION, PREVENTION, AND THE JUST WAR

by

BEN I. BRONNER

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Philosophy

Written under the direction of

Alec D. Walen

And approved by

---

---

---

---

New Brunswick, New Jersey

October 2019

## ABSTRACT OF THE DISSERTATION

Intention, Prevention, and the Just War

by BEN I. BRONNER

Dissertation Director:

Alec D. Walen

According to *just war theory*, there are a number of conditions that must be met in order for it to be morally permissible for a state or other political community to resort to war.

Likewise, there are conditions on the permissible use of force by combatants in war. The conditions offered by just war theory are, at least for the most part, intuitive and plausible. Yet there are a number of problems that arise when we try to find precise statements of these conditions; or when we try to say, in a particular case, whether these conditions have been met. The following papers examine some of these problems.

“Just Cause and Justification” discusses the idea of a just cause for war, as well as the relationship between just cause and justification. It begins by presenting an account of just cause and arguing that it has been the historically dominant account, from the Roman foundations of the just war tradition until the present day. It then presents several cases in which it seems that war can be justified despite there being no just cause—or, more accurately, despite the fact that the historically dominant account seems to entail that there is no just cause. Finally, the paper considers several possible ways of responding to this fact. Most of these possibilities are either inadequate or would represent a significant

departure from the mainstream of the just war tradition. The most adequate and least revisionary response is based on the controversial claim that an individual can be liable to harm in virtue of posing a threat of unjustified harm to others, even if they are not *responsible* for posing such a threat.

“The Responsibility Dilemma and the Relevance of Intentions” considers a challenge posed by Seth Lazar for Jeff McMahan’s influential account of permissible killing in war. McMahan rejects the common claim that combatants on both sides of a war may permissibly kill their enemies. For according to McMahan, posing a threat to another is not sufficient for being liable to harm. Rather, liability to harm is grounded in responsibility for a threat of *unjustified* harm. According to Lazar, when it comes to responsibility for threats of unjustified harm, many civilians are *as responsible* as many unjust combatants. So if liability is grounded in responsibility for such threats, as McMahan claims, then either (i) unjust civilians are rarely liable to be killed *but many unjust combatants are not liable to be killed either*, or (ii) nearly all unjust combatants are liable to be killed *but so are many unjust civilians*. Hence Lazar contends that McMahan has inadvertently committed himself to either pacifism or total war—that is, to either rejecting the existence of justified wars or else embracing indiscriminate attacks upon civilian populations. I develop a particular response to Lazar’s argument. This response accepts that many unjust combatants are not liable to be killed, but it denies that the intentional killing of non-liable unjust combatants infringes the moral constraint on intentionally harming the innocent. Absent infringements of that constraint, there can be a lesser-evil justification for wars in which non-liable individuals are killed, and hence pacifism can be avoided.

Finally, “The Modal Fog of War” discusses a problem that arises when we try to determine the harm that is prevented by some act, such as the resort to war or an act performed within war. I argue that it is not clear how to draw the distinction between *preventing* harm and merely *refraining from causing* harm. This presents an obvious problem, as well as a less obvious one. The obvious problem is that if we cannot tell *whether* some act would prevent harm, then we certainly cannot tell *what* harm it would prevent, and this may leave us in the dark about the act’s permissibility. The less obvious problem is that if we cannot tell whether some act would prevent harm, this threatens to leave us in the dark about the harm that would be prevented by the agent’s *other* options. After presenting this latter problem, I consider several possible responses. I endorse one and note its limitations.

## DEDICATION

This dissertation is dedicated to those who did not live to see it:

Jerry Bronner, Tim Hall, and Ian Santino

## Acknowledgments

I owe a great deal to many people. I owe the most to my mother, Peggy Woodward, for her support and enthusiasm for all of my endeavors. She has never hesitated to encourage me to pursue my interests, no matter how impractical or idiosyncratic they have been. I have benefited enormously from the support that she has provided—emotionally, financially, and otherwise. I am also grateful for the efforts of my late father, Jerry Bronner. Along with my mother, he worked, saved, and planned for many years so that I could receive an excellent education without assuming the burden of debt. Thanks as well to my entire family for their love and support—to Elaine, Harry, and Martha, my grandparents; to Robin, Tom, Steve, Rika, Ken, and Natalia, my aunts and uncles; to James, Tom, Brad, Shauna, and Vica, my cousins; and to Sarah and Ray, my sister and brother-in-law. Thanks to Chris Konzelman, who along with my mother has generously paid for my travel and hosted me during vacations spent in Oregon. Special thanks to Eli Shupe for her patience and encouragement as I worked on this and other projects.

While at Oberlin College, I had the privilege of learning from a wonderful group of philosophers: Dorit Ganson, Todd Ganson, Tim Hall, Al MacKay, Peter McInerney, Katherine Thomson-Jones, and Martin Thomson-Jones. In addition to the high quality of the education that they offered, I was fortunate that all were generous when extensions were needed during difficult times. I aim to be as understanding and humane with my students as you were with me.

Special thanks to Todd Ganson, my undergraduate advisor. Much to my surprise and delight, Todd treated me more as a peer than as a student. I learned an enormous

amount from his classes, from our discussions, and from our collaborations—both about the subject matter and about the process of doing philosophy.

Thanks also to all of the professors that I have had the privilege of learning from while at Rutgers: Martha Bolton, Robert Bolton, Ruth Chang, Andy Egan, Frances Egan, Johann Frick, Thony Gillies, Alvin Goldman, Doug Husak, Frances Kamm, Derek Parfit, Jonathan Schaffer, Susanna Schellenberg, Holly Smith, Larry Temkin, and Alec Walen. All were extremely enthusiastic, about both my work and their own, and were always approachable and helpful.

Mercedes Diaz has been an invaluable ally in all things administrative. I would not have been able to navigate the Rutgers bureaucracy without her. Her good humor and seemingly endless efforts on behalf of graduate students are truly remarkable.

I owe an enormous debt of gratitude to the members of my dissertation committee: Doug Husak, Jeff McMahan, Larry Temkin, and Alec Walen. Each has provided me with invaluable assistance, extending well beyond the material in this dissertation.

Doug is a fixture in the department, physically present more than any other faculty member that I know of. His office door is always open, and he has been more than happy to talk whenever I have dropped in, regardless of the topic. On more than one occasion, Doug has encouraged me to stop by his office more often, saying that he would like to talk more with graduate students. I very much appreciate Doug's efforts to make time for his students—indeed, for all students in the program, whether his or not—and his warm and welcoming attitude. I have greatly enjoyed and benefited from our conversations.

Jeff has been incredibly generous with his time, even after leaving Rutgers for his current position at Oxford. He has made time to meet with me on a number of occasions while visiting the New Brunswick area, and has continued to correspond by email despite his extremely busy schedule. Far more than anyone else, his views on the ethics of war have shaped my own. Indeed, much of the discussion in the following papers could fairly be described as footnotes to McMahan. Although I focus on our disagreements, they are minor in comparison to the extent of my agreement with the approach that McMahan has carefully developed over many years.

Larry has been my most enthusiastic supporter at Rutgers, from my very first semester until my last. I have received many kind emails from Larry over the past six years, invariably containing a generous helping of the ALL CAPS emphasis for which he is famous. In addition to bolstering my confidence, Larry has encouraged me to apply for opportunities outside of the Rutgers philosophy department and has supported my applications when I have done so. I am sure that my acceptance to the Penn Bioethics Boot Camp—which was instrumental in my developing a specialization in bioethics—was due in no small part to the recommendation that Larry made on my behalf. I am enormously grateful for all that he has done for me.

Alec kindly agreed to direct my dissertation despite us having had little prior interaction. Naturally, he has provided me with more comments, and has read more drafts of the following papers, than anyone else. He has continuously challenged me to write more clearly and to think more deeply, often by questioning the assumptions that have framed my arguments. At the same time, Alec has been more than willing to take on these assumptions, if only for the sake of discussion, in order to engage with my project

and provide internal critiques. To the extent that the arguments in this dissertation have been honed and the prose polished, it is primarily due to Alec’s efforts.

Thanks also to Holly Smith and Alex Guerrero for serving on the committee for my dissertation proposal. In doing so, they provided helpful comments on “The Responsibility Dilemma” and a predecessor of “The Modal Fog of War,” as well as on the shape of the dissertation more generally.

Many thanks to Jonathan Schaffer, who has been generous with his time in reading drafts of my papers, and in discussing issues they have raised concerning causation and mereology. I also thank Jonathan for an extremely helpful dissertation seminar in spring 2018, as well as for his service as the graduate program’s placement director. His assistance and advice helped enormously while I was on the academic job market this past year.

In addition to those mentioned above, a number of people have provided valuable comments or conversations that have informed the following papers. Daniel Schwartz engaged in a very helpful exchange about the views of the early modern authors discussed in “Just Cause and Justification.” Simon Goldstein and Seth Lazar generously provided comments on “The Responsibility Dilemma.” Sam Carter suggested the title “The Modal Fog of War.” He also provided helpful comments on the paper itself, as did Bob Beddar, Ben Blumson, Eddy Chen, John Davenport, Chris Frugé, Verónica Gómez, Jimmy Goodrich, Christopher Hauser, John Holbo, Caley Howland, Hsueh Qu, Hui-Chieh Loy, Tyler John, Hrishikesh Joshi, Nico Kirk-Giannini, Zach Kofi, Ting An Lin, Lauren Lyons, Michael Pelczar, Abelard Podgorski, Jonathan Schaffer, Eli Shupe, Philip Swenson, Saranindranath Tagore, Isaac Wilhelm, and audience members at the 2017

Felician Ethics Conference and the 2019 Eastern APA meeting. Many thanks to each of these individuals, as well as to anyone who has discussed these issues with me but who I have failed to mention.

# Contents

Abstract	ii
Dedication	v
Acknowledgments	vi
Introduction	1
Just Cause and Justification	10
Just cause: A first pass	10
Unresolved questions	14
Refinements	17
Narrow and wide proportionality	19
Individualism and collectivism	23
Justification without just cause	26
Responses	33
References	45
The Responsibility Dilemma and the Relevance of Intentions	48
Clarifying the dilemma	51
Intentionally harming the innocent	53
Lesser-evil justification	58
Objections	65
Concluding summary	69
References	69
Appendix: Remaining difficulties	71

The Modal Fog of War	75
Conditions on preventing harm	76
Accounting for cases	76
Taking stock	80
Intermediate conclusions	84
Objections	88
Problems of application	92
Operation Protective Edge	92
What counts as preventing harm?	95
Counterfactuals and proportionality	99
Doing and allowing	102
Forgetting about prevention	105
Argument by cases	109
References	114
Appendix: Another answer for Abe and Bea	116

## Introduction

The following papers are about the ethics of war. They are written from the perspective of *just war theory*, the dominant approach to the ethics of war among philosophers. Just war theory is a big tent, and there is much that just war theorists disagree about. Yet they agree that war can be justified, and that war stands in need of justification. This distinguishes just war theory from pacifism, on the one hand, and political realism, on the other.<sup>1</sup> To say that war *can* be justified is to say that it can be morally permissible for a community to resort to war, and for an individual to participate in war. It is also to say that this is more than a theoretical possibility. It is not merely that war could be justified if, miraculously, no innocents were to be harmed; or that war could be justified if it were the only way to avert some terrible evil, but that war is never actually necessary in this way. Rather, just war theorists hold that, in some sense, it is a live possibility that there are justified wars. To say that war stands in *need* of justification is to deny that war exists beyond the bounds of morality. The moral law is not silent in times of war. Nor are the moral principles applicable to war radically different than those applicable during times of peace. Moral restrictions on killing the innocent do not disappear, for example.

Just war theory is conventionally divided into two parts: a set of requirements on the *resort* to war, and a set of requirements on *conduct* in war. The former set is most often presented as consisting of the following principles.

(*Just Cause*) War may be waged only against an adversary that has committed a serious wrong or injustice, or that threatens to commit one.

---

<sup>1</sup> These are simplifications, in part because pacifism and political realism each come in a variety of different forms.

(*Right Intention*) Belligerents must intend to achieve the just cause—intend, that is, to prevent or rectify the relevant injustice. And they must not also intend to achieve other, unjust aims.

(*Legitimate Authority*) Only entities with a certain type of authority—traditionally, states—may initiate war.

(*Probability of Success*) There must be a reasonable probability that war will achieve the just cause.

(*Necessity*) War is permissible only if there is no alternative that would achieve the just cause while causing less harm.

(*Proportionality*) War is permissible only if its bad effects are not excessive compared to its good effects.

The second set of moral requirements, applicable to conduct in war, is typically understood to include the following.

(*Necessity*) A violent act within war is permissible only if there is no alternative that would achieve the same good effects while causing less harm.

(*Proportionality*) A violent act within war is permissible only if its bad effects are not excessive compared to its good effects.

(*Discrimination*) Only those who are liable to attack—traditionally, all and only combatants—may be attacked.

We have, then, two sets of moral requirements, often referred to by the Latin phrases *jus ad bellum* and *jus in bello*, respectively. On the traditional view, every one of the requirements in the first set must be met, in order for it to be permissible to resort to war; and every one of the

requirements in the second set must be met, in order for it to be permissible to harm people within war.

Not all just war theorists endorse every one of the requirements listed above. Some reject Legitimate Authority, for example, claiming instead that anyone can permissibly use force in response to wrongdoing that rises to a certain level, such as violations of basic human rights. Some just war theorists also add requirements to those listed above. These might include a requirement that war be publicly declared, for instance. And for all of the requirements that have been proposed, there are questions about how to formulate, interpret, and apply them. Nonetheless, the list given above is a standard one.

The first of the following papers, “Just Cause and Justification,” focuses on the central requirement of *jus ad bellum*. The requirement of Just Cause is central insofar as a just cause is the basis on which war can be justified.<sup>2</sup> The case in favor of war, if there is one, is provided by the fact that war would prevent or rectify the injustice that provides the just cause for war.<sup>3</sup> Moreover, it is plausible that many of the other requirements of just war theory cannot be met, or at least are highly unlikely to be met, if there is no just cause for war.<sup>4</sup>

Suppose that state X is at war with state Y. Then each of the following claims is plausible and is accepted by many within the just war tradition.

1. X’s war against Y is justified only if X has a just cause for war against Y.

---

<sup>2</sup> I use capital letters to distinguish Just Cause (the moral principle or requirement) from just cause (the wrong or injustice that provides the basis for war).

<sup>3</sup> This claim reflects the traditional view, and it ignores complications that are discussed in “Just Cause and Justification.”

<sup>4</sup> As argued by Jeff McMahan, including in his “Just Cause for War” (*Ethics and International Affairs*, 2005, vol. 19.3) and *Killing in War* (Oxford University Press, 2009).

2. X has a just cause for war against Y only if Y's combatants are liable to be killed by X's combatants.<sup>5</sup>
3. Whether an individual is liable to be killed depends on whether they are responsible for an injustice that can be prevented or rectified by killing them.<sup>6</sup>

I argue that these claims cannot all be true. For claims 1–3 together entail that X's war against Y is justified only if Y's combatants are responsible for an injustice that can be prevented or rectified by killing them. And there are a number of cases in which it seems that war can be justified even though the opposing combatants are *not* responsible for the injustice that can be prevented or rectified by killing them—or are not responsible *enough* to make it plausible that they are liable to be killed. Hence we must reject at least one of claims 1–3. War can be justified *without* a just cause, or there can be a just cause *without* liability to lethal harm, or an individual's liability to lethal harm is *not* determined by their responsibility for injustice.

Claim 3 also leads to problems all by itself, and not just when considered in conjunction with claims 1 and 2. For suppose that claim 3 is true. In other words, suppose that whether an individual is liable to be killed depends on whether they are responsible for an injustice that can be prevented or rectified by killing them.<sup>7</sup> Now suppose that X has a just cause for its war against Y. Perhaps X is defending against wrongful aggression by Y. Or perhaps X is fighting to recover territory stolen by Y, territory that contains resources needed for X's citizens to live a

<sup>5</sup> If an individual is liable to suffer a given harm, then inflicting that harm on them would not infringe their rights, or wrong them. Liability is often understood in terms of the *forfeiture* of rights; the idea being that by acting in a certain way, an individual can lose their right against being harmed.

<sup>6</sup> ...among other factors, such as the degree to which the individual is responsible for the relevant injustice, as well as the nature and magnitude of the injustice.

<sup>7</sup> On this view, for example, if an attacker is liable to be killed, that is *because* the attacker is responsible for posing a threat of significant (and wrongful) harm to their victim. To take another example, if an item of property is valuable enough, then someone who is responsible for its theft might be liable to be killed as a means of recovering it (assuming that there is no less violent way of recovering the item).

decent life. If X has a just cause for war against Y, then it seems there must be *some* of Y's combatants who are responsible for wronging X—at the very least, the ones who occupy the territory stolen from X, or who are participating in Y's wrongful aggression. But there will also be many of Y's combatants who are *not* responsible for any wrong committed or threatened, or at least who are not any *more* responsible than many of Y's civilians. For there are members of Y's military who play very minor roles in the war; or who are ineffective at carrying out their duties; or who fire their guns in the air, only pretending to make an effort at harming X's combatants. Such combatants bear little or no responsibility for any wrong committed or threatened against X's citizens or combatants. At the same time, it seems that many of Y's civilians bear some responsibility, even if only a small amount, for the wrongs committed or threatened by Y. They may purchase war bonds, pay taxes that fund the military, vote for pro-war politicians, educate those who will go on to fight for Y against X, or support Y's war against X in other ways. In short, it seems that many of Y's combatants are *no more* responsible than many of Y's civilians are for the wrongs that Y has committed, or threatens to commit, against X.

If liability to attack is based on an individual's responsibility for a wrong committed or threatened, then it seems that many of Y's civilians are *as liable* to be harmed as many of Y's combatants are. So if all or nearly all of Y's combatants are liable to be attacked, then many of Y's civilians are *also* liable to be attacked. On the other hand, if few or none of Y's civilians are liable to be attacked, then many of Y's combatants are *also* not liable to be attacked. This is what follows, at least, if it is true that an individual's liability to attack is based on their responsibility for a wrong committed or threatened, and if it is also true that many civilians are as responsible as many combatants for wrongs committed or threatened. And this seems to leave us with a

choice between total war and pacifism—between a view on which it is permissible to kill civilians in a relatively indiscriminate manner in war, because many civilians are liable to be killed; and a view on which war cannot be permissibly waged, because many of a state's combatants are not liable to be killed, even if the state itself is fighting for unjust aims.<sup>8</sup>

I discuss this problem in “The Responsibility Dilemma and the Relevance of Intentions.” For the sake of discussion, I assume that many combatants are not liable to be killed—in particular, that even combatants whose state is fighting for unjust aims are often not liable to be killed. But I deny that this assumption leads to pacifism, contrary to the view sketched in the previous paragraph. I develop a view on which war can sometimes be justified as the lesser evil, even when war will inevitably involve the killing of many combatants who are not liable to be killed. This parallels the fact that war can sometimes be justified despite its inevitably involving the killing of *civilians* who are not liable to be killed.<sup>9</sup>

If a given war is justified as the lesser evil, then the war prevents a greater evil. Initiating a war in which a certain number of innocent people will be killed, for instance, may prevent a significantly greater number of innocent people from being killed. This brings up questions about how to identify the evils, and especially the harms, that are prevented by war, or by particular acts within war. There are familiar problems in this regard. It is often extremely difficult to predict the effects of war ahead of time, or of particular acts within war. And even after the relevant acts have been performed, it may be difficult to determine what their effects were. There

---

<sup>8</sup> Seth Lazar presents this dilemma as a problem for McMahan's *Killing in War*. See Lazar's “The Responsibility Dilemma for *Killing in War*” (*Philosophy and Public Affairs*, 2010, vol. 38.2).

<sup>9</sup> In making the argument sketched in this paragraph, I am fleshing out a response briefly suggested by McMahan. See, especially, McMahan's “Who is Morally Liable to be Killed in War” (*Analysis*, 2011, vol. 71.3).

is significant debate, for example, about the number of people who died as a result of the 2001 US invasion of Iraq.

There is also a less familiar problem that arises in this context. Suppose that we wish to determine the harm prevented by some particular act. The act might be the initiation of war by a state or its leaders, or it might be an act performed in war by a combatant. In order to determine the harm that is prevented by the act, we must compare the outcome of the act with the outcome of nothing being done to prevent the relevant harm. For example, suppose that Villain attacks Victim, and that you intervene to stop the attack. Villain in fact breaks one of Victim's arms, though he would have broken *both* of Victim's arms had the attack not been stopped. Given these assumptions, it follows that you prevent one (and only one) of Victim's arms from being broken. The harm that you prevent corresponds to the *difference* between the harm that actually occurs and the harm that would have occurred, had nothing been done to prevent Villain from harming Victim. So in order to know what harm is prevented by an act, we must know what *would have* happened, had nothing been done to prevent the relevant threat from coming to fruition. The problem is that it is not clear *what counts* as doing nothing to prevent some threat from causing harm. There are cases in which it is not clear what course of action would *be* non-intervention; and hence in which it is not clear what would have happened, had nothing been done to avert the threat in question. These cases are presented, and this problem is discussed, in "The Modal Fog of War." I consider several possible responses to the problem, endorse one of them, and note its limitations.

Resolving this problem is crucial if we are to apply Necessity and Proportionality—that is, if we are to determine whether these requirements are satisfied in particular cases. The statements of Necessity and Proportionality each refer to the *good* that is achieved by the act in

question (either the resort to war or an act performed within war). And the good that is achieved by an act often consists in the prevention of some threatened harm. So, identifying the good that is achieved by an act often requires identifying the harm that is prevented. If we cannot identify the harm that is prevented by a given act, then we may be unable to say whether the act is necessary and proportionate. In other words, we may be unable to say whether the act satisfies the requirements of Necessity and Proportionality. This may leave us unable to say whether the act is permissible.

The following papers, then, focus on Just Cause, Discrimination, and Necessity and Proportionality, respectively. Each paper considers one or two of these principles, attempting to find ways of formulating and applying them that do not lead to implausible or unacceptable conclusions. I have little to say about Legitimate Authority, Right Intention, and Probability of Success. Contemporary authors are sometimes skeptical about the moral relevance of Legitimate Authority and Right Intention.<sup>10</sup> And it is often suggested that Probability of Success can be incorporated into Proportionality.<sup>11</sup> My focus does not, however, reflect a belief in the relative unimportance of Legitimate Authority, Right Intention, or Probability of Success. Indeed, in future work I plan to give each of these requirements the same treatment that Just Cause receives in “Just Cause and Justification.” In other words, for each of these requirements I plan to identify the way that it has typically been understood within the just war tradition, and to examine the relationship between the requirement (so understood) and the justification of war. It is only after doing so that I would be in a position to suggest that Legitimate Authority, Right Intention, or

---

<sup>10</sup> E.g., McMahan “Just Cause for War,” p. 5.

<sup>11</sup> E.g., Thomas Hurka, “Proportionality in the Morality of War,” p. 37 (*Philosophy and Public Affairs*, vol. 33.1); Daniel Statman, “On the Success Condition for Legitimate Self-Defense,” p. 664 (*Ethics*, vol. 118.4).

Probability of Success are morally irrelevant or are entailed by other principles of just war theory. In the meantime, I hope that the following papers contribute to our understanding of the principles that they do discuss, whether or not these principles are the only ones worthy of attention.

## Just Cause and Justification

This paper discusses the idea of a just cause for war, as well as the relationship between just cause and justification. Sections 1–3 present an account of just cause and argue that it has been the historically dominant account, from the Roman foundations of the just war tradition until the present day. Objections to this argument are considered in sections 4–5. Section 6 provides several examples in which it seems that war can be justified despite there being no just cause—or, more accurately, despite the fact that the historically dominant account seems to entail that there is no just cause. Section 7 considers several possible ways of responding to this fact. Most of these possibilities are either inadequate or would represent a significant departure from the mainstream of the just war tradition. The most adequate and least revisionary response is based on the controversial claim that an individual can be liable to harm in virtue of posing a threat of unjustified harm to others, even if they are not *responsible* for posing such a threat.

### 1. Just Cause: A First Pass

Before discussing *accounts* that attempt to determine which causes are just, it will be helpful to describe the *concept* of a just cause. In the intellectual tradition leading to contemporary just war theory, a just cause of war is simply a cause of just war—that is, a basis on which war can be justified. We see this, for example, in Vitoria’s discussion of just cause, which is offered in response to the question: “What are the permissible reasons and causes of just war?” ([1557] 2006: 313). Vitoria begins with examples of what “cannot be a cause of just war,” before

proceeding to discuss “the sole and only just cause for waging war,” thus equating a *cause of just war* with a *just cause of war* (313–314).<sup>12</sup>

This does not mean that possessing a just cause is *sufficient* for the justification of war. A just cause corresponds to a goal capable of justifying the resort to war. Just war theorists acknowledge constraints on the pursuit of such a goal. Depending on the theorist, these can include constraints on *who* may wage war, the *motives* they may have, and the *means* they may use. Violating such constraints may entail that a war is unjustified. As traditionally understood, however, the case *in favor* of war is provided by the just cause.<sup>13</sup>

The concept of a just cause, then, is that of a basis on which war can be justified. This is not yet to say what can provide such a basis. In other words, it is not yet to provide an *account* of just cause. We now turn to the dominant account in the just war tradition.

On this account, it is *injustice* on the part of one’s adversary that provides the basis for a just war. This view is present from an early point in the history of just war theory:

The legal foundation of the Roman just war was the analysis of contractual obligation... Breach of contract in private law justified a civil suit by the injured party to recover his *damma* and *iniuriae*, his damages and injuries. Similarly, in relations between states the injured city-state enjoyed rights to seek compensation and redress... Hence every just war had to be occasioned by the prior guilt of the offending party. (Russell 1975: 4–5)

This view of just war as a response to injustice was taken up by Augustine, who writes that “just wars are defined as those which avenge injuries.”<sup>14</sup> Augustine’s definition of a just war is itself

<sup>12</sup> Grotius follows the same pattern, placing his discussion of “just cause for undertaking war” under the heading of “What causes of war may be called justifying” ([1625] 2006: 401). Molina writes that “in order that a war be just and licit ... it must have a sufficient cause” ([1593] 2006: 334). Similarly, Suárez begins his discussion of just cause by claiming that “there can be no just war without an underlying legitimate and necessary cause” ([1621] 2006: 347). These authors do not recognize a distinction, to be considered below, between just wars and justified wars (McMahan 2005b, 2014a).

<sup>13</sup> ...as well as by facts about war’s effectiveness in achieving the just cause.

<sup>14</sup> *Questions on the Heptateuch*, bk. 6, ch. 10, in Reichberg et al. 2006: 82. Russell writes that on the earlier view just war “was analogous to the pursuit of compensatory damages in private law,” whereas

quoted in Gratian's influential *Decretum*,<sup>15</sup> along with passages that more explicitly “emphasized the similarity between a judicial process and the just war” (Russell 1975: 63). Both were seen as “means of correcting an unjust situation: the one an ordinary procedure; the other an extraordinary measure warranted by extreme circumstances” (*ibid.*).

This view of just cause was transmitted, largely through the influence of Augustine and Gratian, to Aquinas and on to the early modern just war theorists. During this period, “from Aquinas to Grotius,” just war was “defined as a *response* to grave wrongdoing” (Reichberg 2008a: 21–22).

On this understanding, just war had the status of a sanction by which the injured party pressed its claim by dint of force. It was an extension of the rule of law (i.e., determinable rules of right conduct) into a realm where the standard procedures of law (i.e., the enforceable decisions of courts of law) no longer applied. (22)

Hence the “fundamental line of demarcation, distinguishing just from unjust causes of war, was the issue of prior wrongdoing” (21).

The view of just war as a response to injustice is likewise endorsed by contemporary just war theorists. Michael Walzer, for example, begins with the assumption that war is justified only in response to the violation of certain rights; namely, the rights that states possess to territorial integrity and political sovereignty (1977: 61–62).<sup>16</sup> In a similar vein, Jeff McMahan writes that there “is just cause for war” only when “one group of people … is morally responsible for action that threatens to wrong or has already wronged other people” (2005b: 8). Walzer focuses on the

---

“Augustine’s just war could be seen as a penal sanction analogous to the awarding of punitive damages in private law” (1975: 19).

<sup>15</sup> Pt. 2, causa 23, qu. 2, canons 1–2, in Reichberg et al. 2006: 113.

<sup>16</sup> Walzer proceeds to suggest several exceptions to this initial assumption, but only “because they uphold the values of individual life and communal liberty of which sovereignty itself is merely an expression” (1977: 108).

rights of political communities, while McMahan's analysis proceeds solely in terms of the rights of individuals. They nonetheless agree that a just war is a defense or vindication of rights.<sup>17</sup>

Two points are worth emphasizing before moving on. First, when authors in the just war tradition speak of injuries, they are typically referring to *wrongs* and not merely to *harms*. This is evident in the passages describing the Roman notion of just cause, which refer not only to "the injured city-state" but also the "prior guilt of the offending party" and the "correcting [of] an injus situation" (Russell 1975: 4–5, 63). The connection between injury and injustice is also suggested by the full remark from Augustine. Immediately after stating that "just wars ... avenge injuries," Augustine provides two examples: "if some nation or state ... has neglected to punish a wrong committed by its citizens, or to return something that was wrongfully taken."<sup>18</sup> Similarly, Vattel states that the rights of a nation "are many in number and various in kind," and that "whatever constitutes an attack upon these rights is an *injury* and a just cause of war" ([1758] 1916: bk. 3, ch. 1, §26).<sup>19</sup> Grotius is even more explicit about the connection between injury and injustice: "We have seen what constitutes a 'right' (*ius*); and from this concept we derive also the definition of a 'wrong' or 'injury' (*iniuria*), guided by the basic belief that this term refers to whatever is done in opposition to right" ([1604] 1950: ch. 2).

---

<sup>17</sup> Walzer almost never mentions just cause explicitly. Nonetheless, his discussion of aggression in chs. 4–6 is primarily a discussion of just cause. Indeed, when introducing the initial assumption that only international aggression can justify war, Walzer quotes Vitoria's claim that the only just cause is a wrong received (1977: 62). And when moving from the topic of *jus ad bellum* to that of *jus in bello*, Walzer asks: "Can the rules, then, be set aside for the sake of a just cause?" (124).

<sup>18</sup> *Qu.* 6.10, in Reichberg et al. 2006: 82.

<sup>19</sup> Vattel acknowledges defense, punishment, and the pursuit of reparations as just causes ([1758] 1916: §35, §41). Here and throughout, I have changed capitalization in quoted material, adding or dropping capitalization at the beginning of sentences as appropriate.

Second, injustice has traditionally been viewed as the *only* basis for a just war and not merely as one basis among others. Representative are the following passages from Aquinas, Vitoria, and Vattel, respectively.<sup>20</sup>

A just cause is required, namely that those who are attacked, should be attacked because they deserve it on account of some fault.

A prince cannot have greater authority over foreigners than he has over his own subjects; but he may not draw the sword against his own subjects unless they have done some wrong; therefore he cannot do so against foreigners except in the same circumstances.

The foundation or the cause of every just war is an injury, either already received or threatened.

Likewise, the range of just causes is explicitly restricted to injustices (including *injuries*) by Molina, Suárez, Grotius, and Wolff, among others.<sup>21</sup>

There are, of course, a number of differences among the accounts of just cause provided by the authors mentioned in this section. There are disagreements about the range of causes that are just, for instance, as well as about the relationship between just cause and the other requirements of just war theory. Nonetheless, there is a basic agreement within the just war tradition that injustice is the only just cause for war.

## 2. Unresolved Questions

What has been said so far does not tell us *which* injustices provide a just cause for war. There are familiar questions that arise in this regard. In the absence of an actual attack, for example, what conditions must be met in order for the *threat* of attack to provide a just cause? In particular, are there conditions on the probability or the imminence of attack, or on the intentions of those who

---

<sup>20</sup> Aquinas, *Summa theologiae* II–II, qu. 40, art. 1, in Reichberg et al. 2006: 177; Vitoria [1557] 2006: 314; Vattel [1758] 1916: bk. 3, ch. 1, §26.

<sup>21</sup> Molina [1593] 2006: 334; Suárez [1621] 2006: 347; Grotius [1625] 2006: 401; Wolff [1749] 1934: §619.

pose the threat? To take another example, consider human rights abuses carried out by the victims' own government. Which rights must be violated, and how widespread must their violation be, before there is a just cause for military intervention by a foreign party? These questions are not answered by saying that injustice is the only just cause for war.

Also unanswered are questions less often seen in contemporary discussions. Recall Augustine's example of a war in which "some nation or state ... has neglected to ... return something that was wrongfully taken."<sup>22</sup> Grotius adds "the obtaining of what is owed to us" beyond stolen property ([1625] 2006: 402). Under what circumstances, if any, can the recovery of property or collection of debts provide a just cause for war? How valuable, for example, must the items in question be?<sup>23</sup>

As another example, consider the 1941 Anglo-Soviet invasion of Iran, which secured a critically important route for transporting Allied supplies to the USSR. The Allies had misrepresented their true motives prior to the invasion, requesting only that Iran expel German citizens, who were claimed to pose a threat to Iranian sovereignty (Beaumont 1981). Suppose, however, that the Allies had instead requested permission to move supplies through Iran's territory. If Iran were to refuse, would the Allies then have a just cause for war?<sup>24</sup> What if Iran were to allow the transit of supplies moved by the Allies' own means, but refused to hand over

<sup>22</sup> *Qu.* 6.10, in Reichberg et al. 2006: 82.

<sup>23</sup> The recovery of property or collection of debts is also viewed as a just cause by Gratian, Raymond of Peñafort, Aquinas, Vitoria, Molina, Suárez, Pufendorf, Wolff, and Vattel. See, respectively, Reichberg et al. 2006: 113, 134, 177, 315, 336, 348, 457; Wolff [1749] 1934: §620; and Vattel [1758] 1916: bk. 3, ch. 1, §28.

<sup>24</sup> Vitoria argues that denial of passage through a foreign country can provide a just cause to those who are denied. He is, however, explicitly discussing the case where granting passage would bring no harm to those who control the territory in question ([1557] 2006: 300). By contrast, if Iran were to "breach ... their obligations as a neutral," that would "lay them open to the risk of retaliation by Germany" (Beaumont 1981: 217).

control of the Trans-Iranian Railway?<sup>25</sup> Or suppose that the Allies wished to station forces in Iran, in order to prevent the country from falling into German hands.<sup>26</sup> Would a refusal provide the Allies with a just cause for war?<sup>27</sup>

The characterization of just cause as a wrong or injustice sheds little light on these questions. In part, the problem is determining whether there are certain types of injustice that are unable to provide a just cause for war. The variations on the case involving Iran raise questions about whether failures to aid, including failures to allow others to use one's property, can provide a just cause.<sup>28</sup> Even if we assume that there is no restriction on the *type* of conduct that can provide a just cause, there is still the question of when an injustice is *severe enough* to provide a just cause. Failure to aid might be more or less wrongful. Even if *some* failures to aid provide a just cause for war, presumably not *all* do. Suppose, for example, that some state is just barely falling short of its foreign aid commitments. If *this* failure to aid does not provide a just cause for war, but some failures to aid *do*, then when is the threshold crossed? Again, we need a more detailed account of just cause.<sup>29</sup>

---

<sup>25</sup> This assumption would bring the case closer to that of Iceland, which had no capacity to resist incursions into its waters but refused to provide naval or air bases to the Allies. Britain invaded Iceland in May 1940 but encountered no resistance.

<sup>26</sup> Allied concern over this possibility waxed and waned, but it appears not to have been a major motivation for the actual invasion (Beaumont 1981: 221).

<sup>27</sup> Dinstein argues for the legality of military operations in the territory of a foreign state, without its consent, if needed to prevent cross-border attacks by terrorists or other non-state actors (2017: ch. 8, §2.A). By contrast, in the cases that I have in mind, one state would attack another *state*, and it would do so in the absence of *any* prior attack from the latter's territory, whether by state or non-state actors.

<sup>28</sup> These cases raise other questions as well, of course. It is not *merely* that Iran fails to aid in the imagined cases. Presumably, Iran would *also* use force to resist the Allies' attempt to use the resources in question. If there *is* a just cause for war, perhaps it is only because of this fact about Iran's conditional use of (presumably unjustified) force.

<sup>29</sup> In addition to the examples provided in this section—preventive war, humanitarian intervention, recovery of stolen property, collection of debts, and failure to aid—it may also be unclear whether there is a just cause in cases of cyberattacks, border disputes or incursions, or foreign electoral intervention.

### 3. Refinements

When is an injustice significant enough, or of the right type, to provide a just cause for war? In answering this question, writers in the early modern period were guided by their view of war as “an extension of the rule of law” (Reichberg 2008a: 22). For example, Vitoria writes that “not every or any injury gives sufficient grounds for waging war.” He continues:

The proof of this proposition is that it is not lawful to inflict cruel punishments such as death, exile, or confiscation of goods for all crimes indiscriminately... Therefore, since all the effects of war are cruel and horrible—slaughter, fire, devastation—it is not lawful to persecute those responsible for trivial offences by waging war upon them. ([1557] 2006: 314)

Citing Vitoria, Molina makes the very same points, concluding that “punishment ought to be proportionate to the crime” ([1593] 2006: 337). Suárez likewise draws on the analogy with criminal punishment, in order to support his claim that “not any cause whatsoever is sufficient to justify war, but only causes that are very serious and proportionate to the ravages of war” ([1621] 2006: 348).<sup>30</sup>

These points about proportionality are well taken, whether or not war should be understood as a form of punishment.<sup>31</sup> War involves the intentional infliction of severe harm, including death. One of the primary challenges facing just war theory is to explain how such acts can be just. The standard approach, at least among contemporary authors, is to argue that some

---

<sup>30</sup> In contrast to Molina and Vitoria, who each reference proportionality in criminal punishment, Suárez appeals to the idea that some wrongs should not be punished at all: “In like manner, a judge can punish, not all offences whatsoever, but only those that militate against the general peace and to the good of the commonwealth” ([1621] 2006: 348). Similarly, Grotius writes: “Yet, it is quite clear that wars should not be undertaken for any sort of delinquency. For even the vengeance of the laws, which is exercised in safety and only harms the guilty, does not follow upon every wrong” ([1625] 2006: 406).

<sup>31</sup> The authors just cited do not maintain that punishment is the only just cause, despite their reliance on the analogy with criminal punishment. They each allow that defense is a just cause, and likewise for the recovery of stolen property (Vitoria [1557] 2006: 310; Molina [1593] 2006: 336; Suárez [1621] 2006: 349; Grotius [1625] 2006: 402).

individuals are *liable* to be killed. If an individual is liable to suffer a given harm, then inflicting that harm on them would not infringe their rights, or wrong them. Liability is often understood in terms of the *forfeiture* of rights; the idea being that by acting in a certain way, an individual can lose their right against being harmed (e.g., McMahan 2013/2014: 7). The intentional infliction of harm will be reconciled with the demands of justice if those who are harmed are liable to this treatment, and hence have no right against it. For, as Grotius writes, “the use of force which does not violate the right of others is not unjust” ([1625] 2006: 394).<sup>32</sup>

There is always a limit to the harm that an individual is liable to suffer. A pickpocket is not liable to capital punishment. Nor are they liable to lethal defensive force. In both punishment and defense, the harm that an individual is liable to suffer depends on the magnitude of the wrong committed or threatened, as well as the degree to which the individual is responsible for the wrong committed or threatened. Harm that exceeds this limit—harm that is disproportionate to the crime or the threat, in light of the individual’s responsibility—is harm to which the individual is not liable.<sup>33</sup>

---

<sup>32</sup> I intend to remain neutral regarding the claim that an individual can be liable to the foreseen infliction of a given harm without being liable to the intentional infliction of the same harm. This claim appears to be entailed by the view that people have a right against being intentionally involved in the furtherance of others’ plans that is *independent* of other rights, such as the right against being harmed (e.g., Nelkin and Rickless 2014: §2); so that even if an individual has no right against being harmed, they may nonetheless have a right against being *intentionally* harmed. On the other hand, the claim in question appears to be incompatible with the view that intention only magnifies the severity of *other* rights infringements, such as infringements of the right not to be harmed (e.g., Quinn 1989: 346); so that if unintentionally harming someone does not infringe their rights, then neither does intentionally harming them. (The claim on which I intend to remain neutral is endorsed by McMahan [2013b: 161] and denied by Haque [2018: 386].)

<sup>33</sup> At least in the context of defense, the harm to which an individual is liable also depends on the degree to which the conduct that harms them would mitigate or avert the relevant threat. My comments ignore a number of questions that arise in connection with proportionality. For an overview of such issues in the context of punishment, see Walen 2016: §4.4. For an account of proportionality in the context of defense, see McMahan 2013/2014. On the various factors relevant to liability, see McMahan 2011: 548.

These considerations suggest an answer to the question of *which* injustices provide a just cause for war. In order for there to be a just cause, there must be an injustice that is sufficient to make one's adversary *liable* to be warred upon. Given the reality of war, this means that there must be an injustice that is sufficient to make one's adversary liable to be killed.<sup>34</sup>

I will go on to suggest that war can be justified even when the opposing combatants are not liable to be killed; and hence that war can be justified even when there is no just cause, according to the account of just cause sketched above. Before making that argument, a couple of objections should be considered.

#### **4. Narrow and Wide Proportionality**

In order to state the first objection, it will help to draw on a distinction that McMahan makes between two types of proportionality. He writes:

Narrow proportionality is a constraint on a liability justification for the infliction of harm. When the harm inflicted [on an individual] exceeds that to which he is liable, it is disproportionate in the narrow sense. Wide proportionality, by contrast, governs harms to which the victims are not liable. It is a constraint on a lesser-evil justification for the infliction of harm... When the infliction of harm on a nonliable person ... exceeds what can be justified in this way, it is disproportionate in the wide sense. (2018: 420, cf. 2013/2014: §3)

According to McMahan, “wide proportionality is a more stringent constraint than narrow proportionality,” as it is permissible to infringe an individual’s right to not be harmed only when doing so is “necessary to prevent a substantially greater amount of harm from being suffered by others” (2018: 420). It is typically impermissible, for example, to kill one innocent person as a

---

<sup>34</sup> More accurately, there must be one or more injustices that *together* are sufficient to make one's adversary liable to be killed. There may not be any one injustice that *by itself* is sufficient for such liability (McMahan 2014a: §2.1). McMahan (2005b, 2014a) argues for the view of just cause presented in this section. His account will be considered in detail below.

side effect of saving one or even two other innocent people.<sup>35</sup> By contrast, in many circumstances an attacker is liable to be killed, and can permissibly be killed, as a means of saving a single person.

The first objection can be stated in terms of the distinction between narrow and wide proportionality.

*First objection.* Vitoria and the rest are not requiring that the enemy be *liable* to be warred upon. For they recognize *one* proportionality relation rather than two (narrow and wide). Without distinguishing between harms to the liable and harms to the non-liable, these authors are simply requiring a proportion between the harms caused by war—*all* the harms caused by war, taken together—and the injustice that provides the basis for war.

The objection, in other words, is that the account of just cause from section 3 is a fundamentally modern one, as it is predicated on a distinction that earlier authors do not draw.

Contrary to the first objection, the authors in question do seem to tie just cause to narrow proportionality. Consider Vitoria for instance. First recall his claim that “since all the effects of war are cruel and horrible … it is not lawful to persecute those responsible for trivial offences by waging war upon them” ([1557] 2006: 314). Vitoria immediately follows this by saying that “the wicked man ‘shall be beaten according to his fault, by a certain number’ (Deuteronomy 25:2).”<sup>36</sup>

---

<sup>35</sup> The widespread acceptance of this fact is why cases involving runaway trolleys typically involve a choice between killing one person and failing to save *five* people. There would be no temptation to turn the trolley onto the side track, where the one is trapped, if there were only one person trapped on the main track. And for some, the temptation would not be especially strong even if two were trapped on the main track.

<sup>36</sup> In the King James Bible, the entire verse reads: “And it shall be, if the wicked man be worthy to be beaten, that the judge shall cause him to lie down, and to be beaten before his face, according to his fault, by a certain number.”

So Vitoria does seem to be concerned with narrow proportionality—concerned, that is, with whether the harms inflicted on those responsible for the relevant injustice are excessive in relation to their wrongdoing.<sup>37</sup>

In addition, in Vitoria's discussion of "what may be done in a just war," he writes that foreseeably killing the innocent is sometimes permissible. "This is proven," he says, "since it would otherwise be impossible to wage war against the guilty, thereby preventing the just side from fighting. Nevertheless ... care must be taken to ensure that the evil effects of the war do not outweigh the possible benefits sought by waging it" ([1557] 2006: 325). Hence, on Vitoria's view, possessing a just cause does not guarantee that the evils of war are proportionate to its benefits. How, then, should we interpret his claim that "not every or any injury gives sufficient grounds for waging war" (314)? Vitoria would be committed to a contradiction, were he claiming that possessing a just cause requires the evils of war be proportionate to its benefits. This supports the view on which Vitoria is instead claiming that possessing a just cause requires an injury sufficient to make the responsible parties liable to be warred upon.

Hence it seems that Vitoria may have narrow proportionality in mind after all. Much the same could be said of Molina, Suárez, and Grotius. Each suggests that possessing a just cause is compatible with war causing disproportionate harm to the innocent.<sup>38</sup>

This connection between just cause and liability is bolstered by Gregory Reichberg's account of medieval and early modern views on just cause. Reichberg begins by considering Aquinas:

---

<sup>37</sup> I return to the relationship between desert and liability later in this section, and to questions about collective vs. individual liability in the next section.

<sup>38</sup> In contrast to the passages quoted from Vitoria, these other authors focus on disproportionate harm to one's *own* community. See Reichberg et al. 2006: 338, 351, 414–415.

“Those who are attacked,” he wrote, “should be attacked because they merit it on account of some fault.” Commentators have generally assumed that Aquinas’s intent in uttering this phrase was to conceive of just cause in terms of desert... An alternative reading (anticipated in some measure by Vitoria) was, however, proposed by Luis de Molina. Drawing a distinction between material and formal wrongdoing ..., the Spanish Jesuit articulated a liability-based theory of just cause. (2013a: 158; dates omitted)

While formal wrongdoing involves culpability on the part of the wrongdoer, material wrongdoing does not. Molina writes, for example:

The second kind of just [offensive] war is to recover property which is due us, when it is held because of invincible ignorance, and there is no other way for us to recover it. For this sort of war no preexisting guilt is required, but it is enough that a material injury has occurred. ([1593] 2006: 336)<sup>39</sup>

Likewise, Grotius writes that “those persons who bring about injury in any way whatsoever are liable to prosecution in war,” including those who blamelessly (but wrongfully) cause harm ([1604] 1950: ch. 7, cor. to qu. 6, art. 2). Hence just war theorists prior to McMahan had a notion of the harm to which one’s adversary is liable. Indeed, they had a notion of liability *similar* to McMahan’s insofar as they did not take liability to require culpability. Reichberg thus writes that McMahan’s “liability account of licit killing in war ... provides a more systematic reiteration of the positions staked out in the sixteenth and seventeenth centuries by Vitoria, Molina, and Grotius” (2013a: 160).<sup>40</sup>

There are, of course, authors in the just war tradition who focus on culpability and desert rather than (mere) liability. Cajetan, for instance, “views just war as akin to the administration of

<sup>39</sup> Brackets inserted by the editors. Note that invincible ignorance is that which “result[s] from an interplay of factors that are beyond the agent’s control” and hence “may not be imputed to the agent as a personal fault” (Reichberg et al. 2006: 317). It need not be ignorance that is literally ineradicable.

<sup>40</sup> Reichberg does not include Suárez in this list. Yet Suárez suggests that those who are faultlessly ignorant can be subjectively justified in fighting for an objectively unjust cause. He writes, for example: “some persons reply that ... the war may incidentally be just for both sides. *Excluding cases of ignorance*, however, this seems impossible” ([1621] 2006: 366, emphasis added; cf. 348). It seems, then, that Suárez should agree with Vitoria, Molina, and Grotius that legitimate targets in war need not be culpable.

a penal sanction against a condemned criminal” (Reichberg 2013a: 167; cf. Cajetan [1524] 2006: 247–248). Yet if the enemy *deserves* to be warred upon as punishment, then they are also *liable* to be warred upon.<sup>41</sup> So even the authors who focus on desert rather than liability should agree that a just cause exists only when one’s adversary is liable to be attacked.

## 5. Individualism and Collectivism

The second objection agrees that the historically mainstream account of just cause requires the liability of one’s adversary, but it denies that liability was understood in the individualist manner common today.

*Second objection.* The discussion above has made the history of just war theory seem more continuous than it really is. In particular, the classical just war theorists appear to have seen combatants primarily as *parts* of the state to which they belong. By contrast, contemporary authors take seriously the rights that combatants have as individuals, and hence they would resist the suggestion that “the harms [combatants] suffer, including the harm of being killed, should be understood as non-lethal harms to the state.”<sup>42</sup>

We have seen that authors from Gratian to Vattel held that the recovery of stolen property or the collection of debts can provide a just cause for war.<sup>43</sup> How could an individual be liable to lethal attack on the basis of a debt that they owe or stolen property that they possess? Presumably they *cannot*, assuming that the rightful owner can live a decent life without collecting the debt or recovering the property. An individual cannot be liable to lethal force for the sake of mere

---

<sup>41</sup> Similarly, McMahan writes that “if a person deserves to be killed, it follows that he is liable to be killed” (2005a: 386).

<sup>42</sup> In the words of McMahan (2014b: 114), who himself resists such a view. Thanks to Alec Walen for pressing this objection to my historical account of just cause.

<sup>43</sup> See note 23.

property (as acknowledged above, when we said that a pickpocket is not liable to lethal force). If the classical just war theorists think otherwise, then they must be thinking of lethal harms to combatants simply as non-lethal harms to the state.

Contrary to the second objection, the classical just war theorists are not far from their modern-day counterparts with respect to collectivism and individualism. As Reichberg writes, “the general thrust of medieval just war doctrine … spontaneously conceived of war as directed at determinate individuals.”

Alien to this conception was the ancient Roman understanding, revived in modernity by classical international law and reflected in authors such as Rousseau, whereby the enemy had the character of a moral person, a collective whole, in which the composing individuals were united by formal bonds of affiliation. (2013a: 162)

If medieval just war theorists had a more expansive understanding of the conditions under which war was justified, it was not because they conceived of lethal harm to combatants simply as non-lethal harm to the state.

Essentially the same is true of early modern just war theorists, although they raise complications not obviously present in the writings of earlier authors. In particular, early modern theorists often allowed for a type of collective liability with respect to the seizure of property:

A state is liable for its subjects' wrongs to foreigners if it fails to compel them to compensate for their wrongs and fails to punish them. If the state fails in this way, other subjects of the state become liable to punitive and restitutive harm. The result is that innocent subjects of a state may sometimes be liable for the wrongs of other culpable subjects of the same state... The debt can then be forcibly collected from the treasury of the state or from any private citizen, who thereby acquires a reimbursement claim against the state. (Schwartz forthcoming: §3.1)

Although early modern authors such as Molina, Vitoria, and Suárez endorsed this collective or vicarious liability with respect to the seizure of property, they had no such view about liability to lethal force. Explanations for this difference vary. Suárez, for example, writes that there is a “difference existing between life and other possessions.”

For the latter fall under human dominion, and the commonwealth as a whole has a greater right to them than do individual persons. Hence, such persons may be deprived of [their] property because of the fault of the whole commonwealth. But life does not fall under human dominion, and therefore, no one may be deprived of his life save by reason of his own fault. (Suárez [1621] 2006: 364)<sup>44</sup>

Thus Suárez was an individualist about liability to lethal harm, as were most early modern just war theorists.

It may not be clear how this individualism is compatible with the claim that the recovery of property can provide a just cause for war. Assuming that the property in question is not needed for life itself, or even for a *decent* life, how can it be proportionate to kill for its recovery? How can it be proportionate, that is, unless we *are* thinking of killing individuals as simply inflicting non-lethal harm on their state?

I have two answers to this question. The first is that just war theorists from earlier periods may simply place the threshold for liability to lethal harm at a different place, as compared to contemporary authors. One can think that lethal force is a proportionate response to the theft of property, or to the nonpayment of debts, while still being an individualist about liability. There is no incoherence, for instance, in thinking that one may permissibly use lethal force to prevent the theft of a car.

My second answer is that permitting lethal force to be used in the service of recovering property is not, in fact, a view that is alien to the mindset of contemporary authors. Consider the case of conditional threats. A conditional threat will attempt to cause harm if (and *only* if) their demands are not met. An example is the mugger who offers you a choice between your wallet and your life. There is debate about whether it is permissible to kill those who pose a conditional

---

<sup>44</sup> Brackets inserted by editors.

threat of serious harm. Gerhard Øverland (2012) argues that sometimes this is indeed permissible, as did McMahan at one point.<sup>45</sup> McMahan, in particular, argued as follows:

One is ... entitled to take certain steps to resist the theft. Suppose, however, that the thief threatens to kill one if one resists. In that case one is permitted to create the conditions of one's own lethal self-defense. As soon as the thief structures the situation in such a way that to defend one's possessions automatically creates the need for self-defensive killing, one's right to self-defense is immediately activated. One is permitted to kill the thief even without first provoking him to attack by attempting a non-lethal defense of one's possessions. (1994: 196)

The same argument could be made regarding using force to recover stolen property or to collect debts. Suppose that some individual is currently in possession of your property and threatens to kill you if you attempt to collect the property. This individual has “structured the situation in such a way that to defend your possessions automatically creates the need for self-defensive killing.” So it would seem that by the logic of McMahan’s (1994) argument, you are permitted to use lethal force in order to collect your property. The point is not that this conclusion is *correct*. It may not be correct. Rather, the point is simply that such a view is not alien to the mindset of contemporary just war theorists. Hence there is not nearly as much distance between them and their predecessors as is suggested by the second objection. The classical just war theorists held that the recovery of stolen property or the collection of debts can provide a just cause for war, but this fact does not put them radically at odds with contemporary just war theory.<sup>46</sup>

## **6. Justification without Just Cause**

From its Roman foundations up to the present day, the mainstream view within the just war tradition has been that a just cause for war is an injustice sufficient to make one’s adversary

<sup>45</sup> McMahan now rejects this view. See his 2014b: §6.8.

<sup>46</sup> I assume that in the cases that the classical just war theorists had in mind, those in possession of others’ property *are* prepared to cause serious harm in order to maintain possession of this property.

liable to be warred upon. At least since the medieval period, liability to lethal harm has been predicated on an individual's wrongdoing.<sup>47</sup> Yet if just cause and liability are understood in these ways, then it seems that war can be justified in the absence of a just cause. This section provides four examples drawn from the work of McMahan.

Before turning to those examples, I consider another from McMahan that I suggest may not be a genuine example of a justified war in the absence of a just cause.

Suppose that country A is about to be unjustly invaded by a ruthless and more powerful country, B. A's only hope of successful defense is to station forces in the territory of a smaller, weaker, neighboring country, C, in order to be able to attack B's forces from prepared positions as they approach A along the border between B and C. A's government requests permission from the government of C to deploy its forces on C's territory for this purpose, but C's government, foreseeing that allowing A to use its territory in this way would result in considerable destruction, denies the request. (McMahan 2005b: 15; cf. 2009: 27–28)

McMahan suggests that the government of country C does not wrong country A by denying the latter's request. (After all, the government of C can permissibly give some preference to the interests of its own citizens over those of A.) Hence C is not liable to attack by A. Even so, McMahan continues, there could be a lesser-evil justification for A to forcibly take the territory in question from C. If that is correct, then A's war against C could be justified despite lacking a just cause.

It is not clear that such a case is possible, however. If A has a lesser-evil justification for waging war on C, then A's waging war prevents a much greater harm for A than it causes to C.<sup>48</sup> And if A's waging war prevents a much greater harm for A than it causes to C, then C's granting

---

<sup>47</sup> See the quote from Reichberg (2013a) in the previous section. Walzer (1977) and others who endorse the moral equality of combatants, thus divorcing liability from wrongdoing, are historically anomalous within the just war tradition. See Reichberg 2008b, 2013b, 2013c.

<sup>48</sup> The "much greater" clause basically reflects a rejection of consequentialism. See McMahan 2013/2014: 7.

of A's request without a fight would also prevent a much greater harm for A than it cause to C. (For granting the request would decrease the harm that comes to C without decreasing the harm that A prevents by using C's territory.) And if C's granting of A's request would prevent a much greater harm for A than it would cause to C, then it is not clear that C can refuse A's request without wronging A. Even if C can permissibly give *some* preference to its interests, or to the interests of its citizens, it is not clear that the permissible degree of preference is great enough to overcome the fact that A's interest in using the land is *much* greater than C's interest in A not using the land.

Nonetheless, there are *other* cases in which it seems that war is justified despite the fact that the opposing combatants are not liable to be killed. Some of these cases involve a lesser-evil justification for killing those who are not liable to any degree of harm. Other cases involve what McMahan calls a *combined justification*. In such cases, an individual is liable to some degree of harm less than death *and* there is a lesser-evil justification for inflicting more harm beyond that to which they are liable. As a result, there can be a justification for inflicting the harm of death, even when only part of this harm has a liability justification.

To illustrate the idea of a combined justification, consider McMahan's (2014b) discussion of *lesser aggression*—aggression in which “the threat to use military force is ... conditional,” insofar as “the aggressor will need to use military force only if the victims resist rather than capitulate” (112). In such cases, “the question is ... whether it can be permissible for the victims to go to war to defend the values or rights, such as rights to territory, resources, or political sovereignty, that are threatened by lesser aggression” (112). Suppose that there are many combatants participating in the lesser aggression, and hence that each would make no more than a small contribution to the threatened wrong. McMahan suggests that because each lesser

aggressor poses such a small threat, it may be that none are liable to be killed, especially if each is only minimally culpable for the threat that they pose.<sup>49</sup> It may also be that there is no lesser-evil justification for killing any of the lesser aggressors. For killing any *single* lesser aggressor would avert little or no harm. And while killing a sufficient number of the lesser aggressors *would* substantially or even entirely avert the threatened harm, it may be that the magnitude of the threatened harm is “not substantially greater than the aggregate harm that [the lesser aggressors] would suffer if the necessary proportion of them were killed”—meaning that there can be no lesser-evil justification for killing the necessary proportion of lesser aggressors (133). Nonetheless, it may be that a lesser-evil justification can be combined with a liability justification.

While none of the lesser aggressors is liable to be killed, each is nevertheless liable to some degree of harm as a means of defence against the aggression. Let  $x$  be the maximum degree of harm to which a lesser aggressor is liable and let  $d$  be the average degree of harm that people suffer in being killed... Finally, let  $y$  be the extent to which  $x$  is less than  $d$ ... It is possible that it is permissible to kill the lesser aggressor even though he is liable only to harm  $x$ . For ... the remainder of the harm, equivalent to  $y$ , might be justified as the lesser evil in relation to what would happen if he were not killed. (133–134)

When a combined justification is possible, an “act that is disproportionate in the narrow sense can nevertheless be permissible” (134).<sup>50</sup>

---

<sup>49</sup> McMahan suggests that there are important differences between this case and Derek Parfit’s case of the harmless torturers. See McMahan 2014b: 135–141; Parfit 1984: 80.

<sup>50</sup> After discussing this possibility, McMahan writes: “Yet it seems implausible to suppose that the prevention of some aggregate of tiny harms could justify the infliction of a major harm on a single individual on the ground that the major harm is the lesser evil but not on the ground that the individual is liable to the major harm as a means of preventing the many tiny harms” (2014b: 140). I am not sure whether this remark is simply meant to apply to cases in which the portion of harm that must be justified as the lesser evil is especially large—“a major harm”—or whether it instead expresses a more general skepticism about using a combined justification to justify the killing of lesser aggressors.

Yet there is no just cause if there is only a *combined* justification, and not a *liability* justification, for killing those whom it is necessary to kill as a means of preventing or remedying the relevant injustice. For as we have seen, there is a just cause for war only when one's adversary is liable to be warred upon, at least according to the mainstream account of just cause. Hence, defense against lesser aggression is the first example of a war that can be justified in the absence of a just cause (given the mainstream account of just cause).

The second example is a war fought, at least in part, against child soldiers. McMahan considers a hypothetical case in which a child is “brutalized and … indoctrinated” to the point that it would be plausible to think he is “incapable of morally responsible agency” (2010: 32). McMahan writes:

If, given his age and all that has been done to him, this child lacks the moral resources to resist the command to fight, then he is, in the circumstances, a nonresponsible threat. If I am right that a nonresponsible threat cannot be liable to defensive attack, the justification, if there is one, for attacking this child must, it seems, appeal to the idea that attacking him is dictated by necessity: the necessity of averting a much greater evil—for example, the greater evil that would be involved in the success of his leaders’ unjust cause. (33)

McMahan goes on to suggest that child soldiers, at least in most cases, have *diminished* responsibility rather than *no* capacity for responsible agency (33–34). In that case, it may be that a child soldier is liable to some degree of harm less than death, and that there is a combined justification for inflicting the full harm of death.<sup>51</sup>

In the same discussion of child soldiers, McMahan mentions another way that a combatant could be a nonresponsible threat.

Suppose that certain combatants are fighting in a war that is objectively unjust but that all the evidence accessible to them supports the belief that the war is just… Because of this, although they pose an unjust threat, or a threat of unjust harm, they are not morally

---

<sup>51</sup> The reason that a combined justification may be required, and that a liability justification may not available for killing the child soldier, is that the child soldier’s diminished responsibility may limit the harm to which he is liable. See McMahan 2010: 35, 2014b: 134.

responsible for doing so... It seems that they would be nonresponsible threats and would not, on the assumption that liability presupposes responsibility, be liable for their objectively wrongful action. (2010: 32)

Killing such combatants might nonetheless be justified as the lesser evil. Or, if there are combatants with a *low* but *non-zero* degree of responsibility for posing an unjust threat, there might be a combined justification for killing them. This is the third type of case in which war might be justified despite one's adversary not being liable to attack.<sup>52</sup>

Finally, consider the case of preventive war—that is, war intended to avert a non-imminent threat, or intended to prevent a threat from arising in the first place. Suppose that state A has learned of an unjust attack being planned by the leaders of state B, and that it would be much better for A to launch a preventive attack on B now rather than to wait for the latter's plans to progress. Suppose also that the members of B's military have no reason to believe that the leaders of their state plan to involve them in an unjust war (McMahan 2009: 183). In such a case, it is not clear that those attacked as part of a preventive war are liable to be killed. For it is not clear whether B's soldiers have yet *done* anything that makes them responsible for a threat of unjustified harm. McMahan suggests the following possibility:

When it is true that a soldier is strongly predisposed to fight if ordered to do so, and when his leaders are in fact planning to order him to fight in an unjust war, his presence in the military increases the objective risk faced by the potential victims of this potential unjust war. If he has chosen to join when these conditions were among the foreseeable possible outcomes of his joining, that seems to provide a basis of liability to preventive attack, as a means of reducing the risk to which his action has contributed. (2013a: 132–133; cf. 2009: 183–184)

---

<sup>52</sup> What McMahan says in the passage quoted just above seems to conflict with what he says elsewhere: "Even if unjust combatants ... act with subjective justification, there is a basis for the attribution of liability in their prior choice to join or to allow themselves to be conscripted into the military, knowing that there was a risk that they would be ordered to fight in an unjust war, and knowing as well that they might mistakenly regard that war as just" (2009: 183). What McMahan says about non-culpable ignorance in his (2010) paper on child soldiers is likewise in tension with his (2013) discussion of preventive war, considered below.

To continue with our example, an individual in state B's military has been made into a threat by the leaders of his country. He may not *know* that he poses a threat of unjustified harm to state A, but he does in fact pose such a threat. If he poses this threat as a result of voluntarily joining the military, then he is responsible for this threat, and he can be liable to attack on this basis.

This line of reasoning has several limitations, as McMahan notes. For one thing, there may be individuals who bear no responsibility for the fact that they are members of B's military, or who bear too little responsibility to make them liable to be killed, perhaps because they were "forcibly dragooned into the military" (2013a: 143). Hence the justification for A's preventive war against B may be a combined one. There might be a liability justification for killing some of B's combatants and a lesser-evil justification for killing others. Or it may be that some of B's combatants are liable to non-lethal harm and there is a lesser-evil justification for inflicting harm beyond that to which they are liable.<sup>53</sup>

We have now seen four cases in which it appears possible to have a justification for killing opposing combatants who are not liable to be killed—and hence in which it appears possible to have a justification for war without a just cause, given the account of just cause that we have been assuming. Respectively, the cases involve lesser aggression, child soldiers, non-culpable ignorance, and preventive war. In each case, the opposing combatants may not be responsible, or may not be *sufficiently* responsible, for posing a threat of unjustified harm. As a result, a liability justification for killing them may not be available, although a lesser-evil or

---

<sup>53</sup> McMahan suggests that there are two more categories of combatants who would not be liable to be attacked as part of a preventive war: those "who will no longer be in the military when the order to fight is given, or who will receive that order but conscientiously refuse to obey it" (2013a: 142–143). But even if such individuals would not take part in the unjust surprise attack that their leaders are planning, it may be that they would defend B against A's justified preventive war. If so, then it would seem that they pose a threat of unjustified harm to A's combatants and could be liable to attack on that basis—assuming that they had voluntarily joined B's military, and so could be said to be responsible for the threat that they pose.

combined justification might nonetheless be. In the next section, I consider several ways of responding to the suggestion that war can be justified in the absence of a just cause.<sup>54</sup>

## 7. Responses

### *First Response*

I begin this final section by considering McMahan's account of just cause in more detail.<sup>55</sup> Here is an early statement of his account:

There is just cause for war when one group of people ... is morally responsible for action that threatens to wrong or has already wronged other people in certain ways, and that makes the perpetrators liable to military attack as a means of preventing the threatened wrong or correcting the wrong that has already been done. (2005b: 8)

Defense against aggression is a just cause, for example, because the aggressors are responsible for action that seriously wrongs (or threatens to wrong) their victims. The aggressors thereby become liable to be warred upon, in the same way that an individual attacker (outside of war) is liable to be harmed in the defense of the attacker's victim.

McMahan refines his account of just cause in more recent work. He suggests that there is a just cause for war if, and only if,

(1) those whom it is necessary to attack or kill as a means of preventing or rectifying certain wrongs are responsible for those wrongs to a degree sufficient to make them potentially liable to be attacked or killed for that reason, and (2) the attacks and killings

<sup>54</sup> The range and plausibility of the cases we have considered casts doubt on McMahan's claim that "acts of war by unjust combatants are in practice very unlikely ever to be proportionate in the wide sense" (McMahan 2009: 27; cf. 2013b: 164). If an unjust combatant—i.e., a combatant not fighting for a just cause—has a lesser-evil or combined justification for killing, then this act *is* proportionate in the wide sense. Likewise, McMahan writes that justified wars without a just cause "are very rare" (2018: 422). McMahan's remarks are more plausible if we focus on the likelihood of there being a pure lesser-evil justification for war, or for killing in a war lacking a just cause; that is, if we ignore the possibility of a combined justification. Indeed, McMahan seems to be thinking of pure lesser-evil justifications in the discussions just cited.

<sup>55</sup> Among contemporary just war theorists, McMahan is the most detailed in his account of just cause and the most explicit in endorsing a liability-based account. Hence my focus on McMahan.

would occur in a context that would make them acts of war rather than acts of violence in some other form of conflict. (2014a: 431)<sup>56</sup>

This version of McMahan's account differs from the earlier one in its explicit inclusion of condition 2, as well as its reference to *potential* liability. An individual is potentially liable to be attacked or killed if the *agent-based* conditions of liability are satisfied to a sufficient degree; the *circumstance-based* conditions of liability need not be satisfied. The agent-based conditions are "that an individual poses a threat of unjustified harm and is morally responsible for doing so" (433). These conditions "arguably determine the upper limit to the harm to which a person may be liable" (433). Whether a person is *actually* and not merely potentially liable to a given harm depends on the circumstances at hand. For one thing, "a person cannot be liable to a certain defensive harm if there is an alternative means of achieving the defensive aim that would be better, all things considered" (433). In addition, "if effective defense requires harming a threatener in excess of the harm to which he is liable on the basis of the harm he will otherwise cause and the degree of his responsibility for it, the threatener cannot be liable to suffer that excess harm" (433).<sup>57</sup> These are the circumstance-based conditions of liability. Hence, on McMahan's account, an individual is potentially liable to be attacked or killed if they are "morally responsible to a sufficient degree for a threat of sufficiently serious harm to make them

---

<sup>56</sup> McMahan writes "it is sufficient for there to be a just cause" that these conditions are satisfied (2014a: 431), and that for there to be a just cause, condition 1 "must" be satisfied (432, 433). Hence, McMahan seems to endorse both the necessity and the sufficiency of these conditions (as confirmed in his 2018: 421).

<sup>57</sup> It is tautological that a person is not liable to suffer harm *in excess* of the harm to which he is potentially liable on the basis of the agent-based conditions of liability. I assume that McMahan *intends* to say that the second circumstance-based condition of liability is this: A person is not actually liable to a certain harm, even if he is potentially liable to it, if inflicting that harm would not avert the harm that he threatens.

liable to be attacked or killed *if*, but only if, the circumstance-based conditions of liability are also satisfied” (433–434).

In the previous section, I ignored the distinction between agent-based and circumstance-based conditions of liability. In other words, I ignored the possibility that possessing a just cause merely requires that one’s adversary be *potentially* liable to be killed, rather than *actually* liable to be killed. One might wonder, then, whether there could be a just cause for war in the cases considered in section 6. For one might wonder whether the individuals in question could be potentially liable to be killed, even if they are not actually liable to be killed.

In fact, the distinction between potential and actual liability should not affect my argument. In each of the cases from section 6, the *reason* that the relevant individuals are not liable to be killed is that they are not sufficiently responsible for a threat of sufficient magnitude.<sup>58</sup> Hence these are cases in which the relevant individuals are not even potentially liable to be killed.

There is an additional point that can be made here (although it is not necessary for my argument, given the point made in the previous paragraph). It is not clear that we *should* understand just cause in terms of potential rather than actual liability. McMahan writes:

As just war theory has traditionally distinguished morally between harming people who are not liable to be harmed as a means and harming them as a side-effect, there is a good case for reserving the notions of ‘just cause’ and ‘just war’ for wars in which those whom it is necessary to attack as a means of achieving the war’s ends are potentially liable to attack for that reason. (2014a: 432)<sup>59</sup>

---

<sup>58</sup> In these cases, the relevant individuals were (1) numerous enough that the contribution of each to the relevant threat was very small, (2) child soldiers with diminished responsibility, (3) non-culpably ignorant that their cause was unjust, or (4) soldiers who had been unwittingly made into unjust threats by their leaders’ plans to launch a future attack.

<sup>59</sup> The alternative, which McMahan is rejecting, is to say that a just war is one in which *all* those who are harmed—including those who are not intentionally harmed—are liable to the harms that they suffer (Neu 2012).

As McMahan says here, the traditional distinction is between harming as a means those who are *not liable* to harm, and harming such individuals as a side effect. The traditional distinction is not between harming as a means those who are not *potentially* liable to harm, and harming such individuals as a side effect. So if we are trying to keep with tradition, it may be best to reserve the terms ‘just cause’ and ‘just war’ for cases in which those whom it is necessary to attack are *actually* liable to attack. This conclusion is also supported by a more substantive consideration. As McMahan says, attacking those who are not liable to be attacked is unjust.<sup>60</sup> It is not merely that attacking those who are not *potentially* liable to be attacked is unjust. Justice, and hence just cause, seem to be tied to *actual* liability.<sup>61</sup>

### *Second Response*

Recall McMahan’s claim that there is a just cause only when “those whom it is necessary to attack or kill as a means of preventing or rectifying certain wrongs are responsible for those wrongs to a degree sufficient to make them potentially liable to be attacked or killed for that reason” (2014a: 431). This claim seems to entail that *every one* of those whom it is necessary to kill as a means of preventing or rectifying the relevant wrongs must be liable to be killed, in order for there to be a just cause.<sup>62</sup>

---

<sup>60</sup> “That there is a liability justification for harming people in pursuit of a certain aim is what makes the infliction of harm not merely permissible or morally justified but *just*” (McMahan 2014a: 431).

<sup>61</sup> On the other hand, McMahan may be right when he suggests that nothing substantive depends on whether we understand just cause in terms of actual or potential liability, as long as “we know that the harming and killing in a war are morally justified, and if we also know what kinds of justification there are for the different acts of harming and killing” (2014a: 442).

<sup>62</sup> This is contradicted by another statement that McMahan makes at one point: “In the absence of a just cause, none of those who are attacked in the war is liable to attack, at least according to the account of just cause sketched earlier” (2018: 422). This claim entails that a just cause exists if even *one* of those who is attacked in war is liable to attack.

One might reject such a view, adopting either a *threshold* or a *sliding-scale* view. On the former, a just cause requires only the liability of a certain portion of those who must be killed. (It is not clear what the relevant portion would be, or how one would find a non-arbitrary threshold.) On the latter view, the justice of a war varies according to how many of those who must be killed are liable to be killed, and there is nowhere on this scale where we find a bright line between *just* and *unjust*. (This would be a highly revisionary view relative to historical and contemporary just war theory, both of which draw a bright line around wars possessing a just cause.)

The cases described in section 6 could easily be ones in which many, most, or even all of those who must be killed are not liable to be killed. Hence they could easily be cases in which war is justified despite possessing no just cause on a threshold view, or despite being unjust to high degree on a sliding-scale view. So even if a threshold or sliding-scale view is correct, that would not derail the argument I have presented.<sup>63</sup>

### *Third Response*

McMahan considers the worry that the combatants engaged in a preventive war would infringe the requirement of discrimination. This worry arises because a preventive war would appear to involve the intentional killing of those who are not liable to attack. McMahan writes that “although this may seem to be an issue of discrimination, it is actually an issue of proportionality” (2013a: 139). He draws a comparison to a case in which a pilot bombs a group

---

<sup>63</sup> In conversation, Alec Walen notes a problem for McMahan that I have ignored. Given a wrong that is threatened or has already been done, it is doubtful that there *is* a set of individuals whom it is necessary to kill as a means of preventing or rectifying the wrong. Typically, there are *many* sets for which it is true that killing all of the individuals in the set would prevent or rectify the wrong. Indeed, there may be multiple such sets even after excluding consideration of all the options that are morally inferior to some other option—e.g., after ignoring the fact that one could prevent a threatened wrong by killing a large proportion of the civilians of the opposing state. I think McMahan should say that possessing a just cause requires only that there *be* a set of individuals such that killing them would prevent or rectify the wrong in question, and such that they are liable to be killed for that reason.

of enemy combatants, among whom a few civilians are hidden. He claims that the pilot need not intend to kill the civilians, and hence that the pilot's attack need not infringe the requirement of discrimination. In the same manner, McMahan suggests, preventive war need not involve the intentional killing of those who are not liable to attack. He writes that "the foreseen killing of unmobilized soldiers who are not liable to attack may be justifiable in the same way that the foreseen killing of innocent civilians as a side effect of military action can sometimes be justified in the course of just warfare" (2013a: 143).

If what McMahan says about discrimination is correct, then something similar may apply to just cause. Recall his suggestion that we "reserv[e] the notions of 'just cause' and 'just war' for wars in which those whom it is necessary to attack as a means of achieving the war's ends are potentially liable to attack for that reason" (2014a: 432). If non-liable combatants need not be killed as a means, because their killing can instead be a side effect of killing liable combatants, then a war can possess a just cause despite unavoidably involving the killing of non-liable combatants.

There is a response to McMahan's argument that we might be tempted to make at this point, but that we should not endorse. Suppose that a rifleman, R, shoots and kills an enemy combatant, E<sub>1</sub>. Because R intentionally kills E<sub>1</sub>, it would seem that if E<sub>1</sub> is not liable to be killed, then R intentionally kills a non-liable combatant. E<sub>1</sub> is not killed as a *side effect* of killing some other combatant, in contrast to a case where E<sub>1</sub> is killed by a bomb or missile targeted at another combatant. Indeed, suppose that R later shoots and kills a second enemy combatant, E<sub>2</sub>, who is liable to be killed. Assuming that R does not know which combatant is liable to be killed and which is not, then presumably R's mental state with respect to E<sub>1</sub> is the same as his mental state with respect to E<sub>2</sub>. If R intentionally kills the liable enemy combatant, then he intentionally kills

the non-liable enemy combatant. So how could it be that R does not intentionally kill a non-liable combatant? The answer is that R *does* intentionally kill a non-liable combatant, in the sense that he intentionally kills a combatant and this combatant is not liable to be killed. The question, however, is whether such a killing infringes a moral constraint on intentionally harming the innocent. In other words, the question is whether such a killing infringes a constraint that—perhaps among other things—speaks against intentionally harming those that one knows are not liable to the relevant harm. So we should not dismiss McMahan’s argument simply because the individuals in question will, at least in one sense, intentionally kill combatants who are not liable to be killed.<sup>64</sup>

Yet there is a different response to McMahan’s argument that we *should* endorse. First note that in each of the cases from section 6, preventing or remedying the relevant injustice *requires* that some of the non-liable combatants be killed as a means. It is not merely that some of those who kill non-liable combatants will *in fact* intend to kill them.<sup>65</sup> In the case of lesser aggression, for instance, the aggressors each make a (small) contribution to the unjust threat that their state poses. Killing them is a means of eliminating the contributions that they make. It may also be a means of convincing the aggressing state to abandon its unjust aims. Much the same is true regarding child soldiers, as well as for combatants who are non-culpably ignorant of the fact that they pose a threat of unjustified harm. In each case, some non-liable combatants must be killed as a means if the justified combatants are to protect themselves and achieve their war aims. Indeed, the same may be true in the case of preventive war—the case in which McMahan writes that a liability justification must be “supplemented by a necessity justification for the

---

<sup>64</sup> For elaboration, see my “The Responsibility Dilemma and the Relevance of Intentions.”

<sup>65</sup> This is relevant if we follow McMahan in saying that a just cause requires only that those who *must* be killed as a means are (potentially) liable to be killed.

unavoidable killing as a *side effect* of people, including some soldiers, who are not liable to attack” (2013a: 143, emphasis added). Consider the members of state B’s military who did not voluntarily join the military, and hence whose joining cannot be a basis for liability to attack. It may be necessary to kill some of these combatants in order to prevent them from defending against the preventive attack.<sup>66</sup> So in this case as well, war may require that non-liable combatants be attacked and killed as a means. This is the first step in my response to McMahan’s argument about liability and discrimination.<sup>67</sup>

A second step is necessary. From the fact that non-liable combatants must be killed as a means, it does not follow that these killings would infringe a moral constraint on intentionally harming the innocent. Indeed, McMahan holds that no such constraint is infringed if one believes that one’s target is liable to be killed.<sup>68</sup> So in the cases discussed above, one might think that the justified combatants need not infringe any constraint on intentionally harming the innocent, as they may believe that their targets are liable to be killed. In other words, one might think that while some non-liable individuals must be killed as a means, these killings would be morally *like* unintended killings. Hence one might conclude that such killings are irrelevant to whether there

<sup>66</sup> Would it be necessary to kill some of these combatants simply in order to prevent them from participating in the aggression that their leaders are planning? Presumably not, for winning the preventive war on state B should suffice to prevent the planned aggression. Hence I focus on the case in which some of B’s non-liable combatants must be killed in order to prevent them from defending against state A’s preventive attack.

<sup>67</sup> Suppose that it is necessary to kill some members of state B’s military in order to prevent them from defending against state A’s preventive attack. Given that these individuals will use force to resist A’s justified preventive attack, does it follow that they are liable to be killed by A’s combatants? Not necessarily. For if they did not voluntarily join B’s military, then it may be true that they have not yet *done* anything to make themselves liable—i.e., that they are not yet *responsible* for a threat of unjustified harm. Compare: it may be true that “students in a country with universal conscription … later will pose a threat,” but that does not mean they are liable to be attacked *now* (McMahan 2013a: 138).

<sup>68</sup> At one time McMahan held that this belief must be *reasonable* (2011: 557f), though he is no longer committed to this restriction (personal communication). For simplicity, I will speak in terms of belief *simpliciter*, though the argument to come could be restated in terms of reasonable belief.

is a just cause for war, but instead—like the unintended killing of non-combatants—is relevant only to whether war is necessary and proportionate. Further argument is needed to show that this is not the case.

Here, then, is the second step in my response to McMahan’s argument about liability and discrimination. Suppose for reductio that both of the following are true:

1. If one person believes that a second is liable to be killed, then the former’s intentional killing of the latter does not infringe a moral constraint on intentionally harming the innocent.
2. If the intentional killing of certain non-liable combatants would not infringe a moral constraint on intentionally harming the innocent, then the fact that some war must involve the intentional killing of these combatants is relevant only to whether the war is necessary and proportionate, and not to whether the war has a just cause.

Given these suppositions, consider the situation that results when state A preventively attacks state B, the leaders of which have been planning an unjust attack on state A. Presumably the combatants of B will believe that the combatants of A are liable to be killed. Indeed, the former will see the latter as unjustified aggressors. (These beliefs may even be entirely justified, given the information available to B’s combatants.) From 1, it follows that the killing of A’s combatants by B’s combatants would not infringe a moral constraint on intentionally harming the innocent. From 2, it then follows that although B’s war of defense against A’s preventive attack would necessarily involve the widespread intentional killing of A’s non-liable combatants, this fact has no bearing on whether B has a just cause for war against A. If we were being a bit careless, we might say that this means B has a just cause for war against A. More accurately, it means that B does not *need* a just cause for war, insofar as the role of a just cause is to provide a

liability justification for killing as a means, and since—given our assumption for reductio—all of the killing of A’s combatants by B’s combatants is morally equivalent to unintended side-effect killing. This is obviously the wrong result. State B *does* need a just cause if its war against state A is to be just, and it does *not* have a just cause. Hence we must reject the supposition that we made for reductio. So McMahan’s argument about liability and discrimination does not block the conclusion that the cases from section 6 are examples of justified wars that lack a just cause. Those who do the killing in those cases may believe that their targets are liable to be killed, but this fact does not mean that the killings are relevant only to proportionality and necessity rather than just cause. A war lacks a just cause if it must involve killing as a means individuals who are not liable to be killed, even if those who do the killing believe that their targets are liable to be killed.

#### *Fourth Response*

I have argued that war can be justified in the absence of a just cause *given* the mainstream understanding of just cause within the just war tradition. One might accept this argument and conclude that the mainstream understanding of just cause is mistaken. One might, in particular, suggest that a just cause requires only that there be a justification for attacking one’s adversary—not that there be a *liability* justification for doing so. On such a view, there can be a just cause for war even if there is only a combined liability/lesser-evil justification for attacking the opposing combatants. Indeed, there can be a just cause, on such a view, even if there is a pure lesser-evil justification for war.<sup>69</sup>

---

<sup>69</sup> Thanks to Alec Walen for pressing this type of response.

While a full evaluation of this suggestion is beyond the scope of this paper, it is worth noting the following. As we have seen, we will be departing from the historical mainstream if we say that war can be justified without a just cause. Presumably we wish to avoid such a departure. But we will be making a similar departure if we say that there can be a just cause for war without one's adversary being liable to attack. For on the traditional view, a just cause requires that one's adversary be liable to be warred upon, as argued in section 3.

In addition, it is not clear whether there is anything more than a verbal difference between the following positions: (i) a war against child soldiers (for instance) is justified despite lacking a just cause; (ii) a war against child soldiers has a just cause, although just cause does not require the liability of one's adversary. Either way, it seems that what matters is that achieving one's justified war aims requires killing as a means some who are not liable to be killed.<sup>70</sup> Hence it is not clear whether the fourth response provides a substantive alternative to the view that war can be justified in the absence of a just cause.

#### *Fifth Response*

The argument in section 6 relied on a particular view of liability. On this view, if a person is liable to be killed, they are liable in virtue of their responsibility for an injustice committed or threatened. One might reject such a view, endorsing either of two alternatives. First, one might endorse a form of collective liability, holding that an individual can be liable to be killed in virtue of the actions of the state to which they belong. Second, one might adopt a view on which committing or threatening an injustice can be sufficient for liability, even in the absence of

---

<sup>70</sup> Recall McMahan's suggestion that what matters is knowing "that the harming and killing in a war are morally justified, and ... what kinds of justification there are for the different acts of harming and killing" (2014a: 442).

responsibility. In particular, one might adopt a view on which posing a threat of unjustified harm can be sufficient for liability to lethal defensive force, even if one is not responsible for the fact that one poses such a threat. With respect to lethal harm, opting for collective liability would be a significant break from the mainstream of the just war tradition. (See section 5.) Opting for liability in the absence of responsibility would not be nearly so radical. Recall Grotius's claim that "those persons who bring about injury in any way whatsoever are liable to prosecution in war" ([1604] 1950: ch. 7, cor. to qu. 6, art. 2). After making this claim, Grotius adds that "law corrects not only voluntary but also involuntary sins." Hence he claims:

If any person threatens me with danger while he is dreaming ... [or] while he is insane ... there is no doubt but that I may rightly repel force with force, even to the point of slaying that person if no other way of ensuring my own safety is left open. Yet such an assailant is not "acting with wrongful intent," since at the time in question he is *non compos mentis*. It suffices that his act is in conflict with the Third Law.<sup>71</sup> ([1604] 1950: ch. 7, cor. to qu. 6, art. 2)

In other words, it suffices that the dreaming or insane individual poses a threat of wrongful harm.<sup>72</sup> Likewise, we might attribute liability to child soldiers, combatants operating under non-culpable ignorance of the fact that they pose a threat of unjustified harm, and soldiers whose leaders are planning unjust aggression. In each case, the relevant individuals can be said to pose a threat of significant and unjustified harm, and perhaps that suffices for liability to attack.<sup>73</sup>

This section has considered five responses to the suggestion that war can be justified in the absence of a just cause. Contrary to what we might have hoped, the first three responses each

<sup>71</sup> The third law is this: "Let no one inflict injury upon his fellow" ([1604] 1950: ch. 2).

<sup>72</sup> Thomson (1991) and Walen (2019) each hold a similar view.

<sup>73</sup> It is less clear whether such a claim could be made about those who participate in lesser aggression. For in that case, the primary consideration is not that the aggressors lack responsibility for posing a threat of unjustified harm. At least as important is the fact that each individual makes only a small contribution to the relevant threat.

failed to yield the conclusion that there is a just cause in the relevant cases. The fourth response was able to produce that conclusion, but at the cost of departing significantly from the traditional understanding of just cause. It was also unclear whether the fourth response presented a substantive rather than merely verbal alternative to the view that war can be justified in the absence of a just cause. The most adequate and least revisionary response is the fifth. This response is based on the claim that an individual can be liable to harm in virtue of the fact that they pose a threat of unjustified harm to others, even if they are not *responsible* for the fact that they pose such a threat. Such a claim is controversial. It is rejected by McMahan and most other contemporary authors who base liability in individual wrongdoing. Yet it is accepted by others within the just war tradition. It may be the best way of avoiding the conclusion that a justified war need not have a just cause. This would preserve the traditional understanding of a just cause as a basis upon which war can be justified, with its implication that no war can be justified in the absence of a just cause.

## References

- Beaumont, Joan. 1981. Great Britain and the rights of neutral countries: The case of Iran, 1941. *Journal of Contemporary History* 16.1: 213–228.
- Cajetan. [1524] 2006. *Summula*. In Reichberg et al. 2006: ch. 22.
- Dinstein, Yoram. 2017. *War, Aggression and Self-Defence*, 6th ed. Cambridge: Cambridge University Press.
- Grotius, Hugo. [1604] 1950. *Commentary on the Law of Prize and Booty*. Translated by Gwladys L. Williams, with the collaboration of Walter H. Zeydel. Oxford: Clarendon.
- . [1625] 2006. *On the Law of War and Peace*. In Reichberg et al. 2006: ch. 32.
- Haque, Adil Ahmad. 2018. Human shields. In *The Oxford Handbook of Ethics of War*, edited by Seth Lazar and Helen Frowe. New York: Oxford University Press.
- McMahan, Jeff. 1994. Innocence, self-defense and killing in war. *Journal of Political Philosophy* 2.3: 193–221.

- . 2005a. The basis of moral liability to defensive killing. *Philosophical Issues* 15: 386–405.
- . 2005b. Just cause for war. *Ethics and International Affairs* 19.3: 1–21.
- . 2009. *Killing in War*. New York: Oxford University Press.
- . 2010. An ethical perspective on child soldiers. In *Child Soldiers in the Age of Fractured States*, edited by Scott Gates and Simon Reich. Pittsburgh: University of Pittsburgh Press.
- . 2011. Who is morally liable to be killed in war. *Analysis* 71.3: 544–559.
- . 2013a. The conditions of liability to preventive attack. In *The Ethics of Preventive War*, edited by Deen K. Chatterjee. New York: Cambridge University Press.
- . 2013b. War crimes and immoral action in war. In *The Constitution of the Criminal Law*, edited by R.A. Duff, Lindsay Farmer, S.E. Marshall, Massimo Renzi, and Victor Tadros. Oxford: Oxford University Press.
- . 2013/2014. Proportionate defense. *Journal of Transnational Law and Policy* 23: 1–36.
- . 2014a. Proportionality and just cause: A comment on Kamm. *Journal of Moral Philosophy* 11.4: 428–453.
- . 2014b. What rights may be defended by means of war? In *The Morality of Defensive War*, edited by Cécile Fabre and Seth Lazar. New York: Oxford University Press.
- . 2018. Proportionality and necessity in jus in bello. In *The Oxford Handbook of Ethics of War*, edited by Seth Lazar and Helen Frowe. New York: Oxford University Press.
- Molina, Luis de. [1593] 2006. *On Justice and Law*. In Reichberg et al. 2006: ch. 28.
- Nelkin, Dana Kay and Samuel C. Rickless. 2014. Three cheers for double effect. *Philosophy and Phenomenological research* 89.1: 125–158.
- Neu, Michael. 2012. Why McMahan's *just* wars are only *justified* and why that matters. *Ethical Perspectives* 19.2: 235–255.
- Øverland, Gerhard. 2012. Conditional threats. *Journal of Moral Philosophy* 7.3: 334–345.
- Parfit, Derek. 1984. *Reasons and Persons*. New York: Oxford University Press.
- Quinn, Warren S. 1989. Actions, intentions, and consequences: The Doctrine of Double Effect. *Philosophy and Public Affairs* 18.4: 334–351.
- Reichberg, Gregory M. 2008a. Jus ad bellum. In *War: Essays in Political Philosophy*, edited by Larry May with the assistance of Emily Crookston. New York: Cambridge University Press.
- . 2008b. Just war and regular war: Competing paradigms. In *Just and Unjust Warriors: The Moral and Legal Status of Soldiers*, edited by David Rodin and Henry Shue. New York: Oxford University Press.
- . 2013a. Culpability and punishment in classical theories of just war. In *Just War: Authority, Tradition, and Practice*, edited by Anthony F. Lang Jr., Cian O'Driscoll, and

- John Williams. Washington, DC: Georgetown University Press.
- . 2013b. The moral equality of combatants – A doctrine in classical just war theory? A response to Graham Parsons. *Journal of Military Ethics* 12.2: 181–194.
- . 2013c. Second response to Parsons. *Journal of Military Ethics* 12.4: 370–372.
- Reichberg, Gregory M., Henrik Syse, and Endre Begby, eds. 2006. *The Ethics of War: Classic and Contemporary Readings*. Malden, MA: Blackwell.
- Russell, Frederick H. 1975. *The Just War in the Middle Ages*. Cambridge: Cambridge University Press.
- Suárez, Francisco. [1621] 2006. Disputation 13: On war. In Reichberg et al. 2006: ch. 29.
- Schwartz, Daniel. forthcoming. *Jus post bellum* in the late scholastics: Statism and individualism. In *A Companion to the Spanish Scholastics*, edited by Harald E. Braun and Erik De Bom. Leiden: Brill.
- Thomson, Judith Jarvis. 1991. Self-defense. *Philosophy and Public Affairs* 20.4: 283–310.
- Vattel, Emer de. [1758] 1916. *The Law of Nations*. Translated by Charles G. Fenwick. Washington, DC: Carnegie Institution.
- Vitoria, Francisco de. [1557] 2006. *On the Law of War*. In Reichberg et al. 2006: ch. 27.
- Walen, Alec D. 2016. Retributive justice. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, winter 2016 edition.  
<https://plato.stanford.edu/archives/win2016/entries/justice-retributive/>.
- . 2019. *The Mechanics of Claims and Permissible Killing in War*. New York: Oxford University Press.
- Walzer, Michael. 1977. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. New York: Basic Books.
- Wolff, Christian. [1749] 1934. *The Law of Nations Treated According to a Scientific Method*. Translated by Joseph H. Drake. Oxford: Clarendon.

## The Responsibility Dilemma and the Relevance of Intentions

In various articles and his (2009) *Killing in War*, Jeff McMahan challenges the once-dominant position represented by Michael Walzer's (1977) *Just and Unjust Wars*. Walzer claims that combatants on both sides of a war may permissibly kill their enemies. Allegedly, all combatants pose a threat to the opposing side and hence are liable to be killed.<sup>74</sup> McMahan responds that posing a threat to another is not sufficient for being liable to harm. A police officer who poses a lethal threat to a dangerous criminal is not necessarily liable to being harmed by the criminal, for example. Rather, liability to harm is grounded in responsibility for a threat of *unjustified* harm—or, more generally, responsibility “for a grievance that provides a just cause for war” (McMahan 2004: 722).

Following McMahan, “Let us say that those who fight in a just war are ‘just combatants,’ while those who fight in a war that is unjust because it lacks a just cause are ‘unjust combatants’” (2004: 693).<sup>75</sup> There is, McMahan claims, an asymmetry between just and unjust combatants. Unjust combatants will often be responsible for threats of unjustified harm, thus becoming liable to be killed. By contrast, just combatants typically will not be responsible for threats of unjustified harm, and hence will not be liable to be killed or otherwise harmed.

Seth Lazar (2010) contends that McMahan has inadvertently committed himself to either

<sup>74</sup> If someone is liable to harm, then they have “acted in such a way that to [harm them] would neither wrong [them] nor violate [their] rights” (McMahan 2005: 386). This is typically understood in terms of the person *forfeiting* rights that they previously had.

<sup>75</sup> A *just cause* is often understood simply as a good reason for war. McMahan in particular believes there is a just cause when “those whom it is necessary to attack as a means of preventing or correcting a wrong are morally liable to be attacked as a result of their responsibility for that wrong” (2013/2014: 15; see McMahan 2014 for details). The following remarks may not apply to combatants who fight in a war with a just cause, but that is unjust because it fails to satisfy one or more of the other conditions of a just war (such as proportionality or necessity).

(i) rejecting the existence of justified wars, or (ii) embracing indiscriminate attacks upon civilian populations. For according to Lazar, when it comes to responsibility for threats of unjustified harm, many civilians are *as responsible* as many unjust combatants. So if liability is grounded in responsibility for such threats, as McMahan claims, then either (i) unjust civilians are rarely liable to be killed *but many unjust combatants are not liable to be killed either*, or (ii) nearly all unjust combatants are liable to be killed *but so are many unjust civilians*.<sup>76</sup>

Two issues should be clarified at this point. First, there is Lazar's claim about the relative responsibility of unjust combatants and unjust civilians. Lazar writes that many unjust combatants make "small and unnecessary" contributions to the threats their side poses (2010: 190). He cites cooks, mechanics, and engineers as examples, as well as combatants who are simply ineffective at performing their duties (190–191). Crucially, civilians also make small contributions to unjust wars, for example by paying taxes, voting, and contributing to the economy, as well as by raising and educating those who go on to become combatants (192).

Second, there is Lazar's claim that the non-liability of unjust combatants would commit McMahan to a contingent form of pacifism—contingent because no *actual* wars will be justified, even if war can *in principle* be justified. Just combatants will often be unable to tell which of their opponents are non-liable, especially when attacking from a distance, as in air or artillery attacks (Lazar 2010: 187). So if unjust combatants are often not liable to be killed, then it is inevitable that just combatants will intentionally kill people who turn out to be innocent (i.e., not liable to the harm in question). McMahan believes that intentionally killing the innocent is much harder to justify than foreseeably killing the innocent as a side effect of attacking a military

---

<sup>76</sup> By 'unjust civilian', I simply mean a civilian whose state fights "a war that is unjust because it lacks a just cause" (to borrow McMahan's definition of an unjust combatant; 2004: 693).

target. Indeed, he believes that the former is a grave *pro tanto* wrong that can be justified only in the rarest of circumstances. So if many unjust combatants are not liable to be killed, then McMahan must accept that all actual wars are unjustified because they involve the widespread intentional killing of innocent people. Or so Lazar argued.

Hence, McMahan appears to face the *responsibility dilemma*: either unjust combatants are not responsible enough (for threats of unjustified harm) to make war permissible, or civilians are responsible enough to make them legitimate targets. McMahan (2011) responds primarily by arguing that, contra Lazar, most unjust combatants are liable to lethal harm while most unjust civilians are not.

I develop a different response to the responsibility dilemma. This response accepts that many unjust combatants are not liable to be killed, but it denies that the intentional killing of non-liable unjust combatants always infringes the moral constraint on intentionally harming the innocent. Absent infringements of that constraint, there can be a lesser-evil justification for making a decision (such as initiating or continuing a war) that leads to the killing of many non-liable individuals. Contingent pacifism is avoided if there are cases where the resort to war is justified in this manner.<sup>77</sup> Before developing this response, I attempt to clarify the nature of the responsibility dilemma.

---

<sup>77</sup> A disclaimer: In his 2010 paper, Lazar was primarily arguing that McMahan's pre-2010 work did not contain the resources to deal with the responsibility dilemma. Lazar may well be correct in that regard. My aim is to consider a response to the responsibility dilemma briefly mentioned by McMahan (2011: 557–558; Lazar 2010: 202n55), and to develop what I take to be the most defensible version of this response. In order to do this, I will draw on McMahan's work post-2010 as well as the work of other philosophers. None of this should be read as a criticism of Lazar 2010, or as an attempt to reconstruct how the dialectic unfolded between Lazar and McMahan in subsequent publications.

## 1. Clarifying the Dilemma

McMahan, I suggest, does not just face a choice between two unattractive positions. Although Lazar does not make this point, I believe that Lazar's remarks suggest McMahan is committed to embracing *both* pacifism and widespread civilian liability.

First, note that McMahan embraces a *comparative* account of liability. On such an account, even *slight* responsibility for a threat of unjustified harm can ground liability if one is *more* responsible than those who would otherwise be harmed. Suppose, for example, that a boulder is about to fall on Abe, though he could redirect the boulder so that it falls on Bea. If Abe is not at all responsible for the impending harm, while Bea is very slightly responsible, and there are no other relevant differences between Abe and Bea, then McMahan would say that Abe may permissibly redirect the boulder so that it kills Bea, who is liable to that fate. Given that Bea is more responsible than Abe for the impending harm, it is more just if Bea suffers that harm than if Abe does.

So if civilians on the unjust side bear even a small degree of responsibility for the threats posed by their state, then according to McMahan's account they can be liable to be killed. As Lazar says, "the whole point of [McMahan's] model of self-defense is to enable small differences [in responsibility]"—such as the difference in responsibility between just combatants and unjust civilians—"to make all the difference in the allocation of unavoidable harms" (204). This leads Lazar to believe that McMahan is forced toward the civilian-liability horn of the responsibility dilemma (personal communication). McMahan's comparative account of liability guarantees that a very slight degree of responsibility can be sufficient for liability. So McMahan cannot avoid the commitment to civilian liability by embracing pacifism and raising the degree of responsibility required for liability.

What Lazar does not say, but should, is that McMahan cannot avoid pacifism by embracing widespread civilian liability. Lazar writes that many unjust combatants have *no* responsibility for threats of unjustified harm: “Whether through fear, disgust, principle, or ineptitude, many combatants are wholly ineffective in war,” or are even “a positive hindrance” to their side (2010: 190). Unjust combatants may also have their fighting days behind them, with the result that they no longer contribute to threats of harm, though still be located in positions vulnerable to attack (191). Since none of these combatants bear any responsibility for current or future threats of unjustified harm, none of them will be liable to harm on an account that grounds liability in responsibility for such threats. Hence McMahan cannot avoid pacifism simply by reducing the degree of responsibility required for liability.

So it is not that “McMahan must choose between two unpalatable options,” as Lazar says (2010: 181). Rather, Lazar in fact provides reason to believe that McMahan is committed to *both* of the unpalatable options.

McMahan *could* face a genuine dilemma if (1) Lazar were wrong that many unjust combatants bear no responsibility for threats of unjustified harm, and (2) McMahan's comparative account of liability did not commit him to widespread civilian liability. For even if (contra Lazar) nearly all unjust combatants are responsible to some degree for threats of unjustified harm, many civilians are *as responsible* as many combatants (or so Lazar claims). So either those combatants are not liable, or those civilians are. But McMahan does not face a *choice* between these two options unless he can respond to the claim that he is committed to *both* the widespread liability of civilians *and* the widespread innocence of unjust combatants.<sup>78</sup>

---

<sup>78</sup> Lazar (personal communication) suggests that McMahan still faces a choice insofar as adjustments in McMahan's view meant to deal with one horn of the dilemma will make it harder for McMahan to avoid the other horn (see Lazar 2010).

I hope to have clarified the dialectical position that McMahan is in. I will now suggest that he may not be committed to a contingent form of pacifism, even if a significant number of unjust combatants bear *no* responsibility for threats of unjustified harm and hence are not liable to be killed. This paper does not address the claim that McMahan's comparative account of liability commits him to widespread civilian liability.<sup>79</sup>

## **2. Intentionally Harming the Innocent**

McMahan denies that (a) intentionally killing individuals who are in fact innocent, always counts as (b) infringing the moral constraint on intentionally harming the innocent. If Agent kills Victim, mistakenly believing that Victim is liable to be killed, then Agent may not have infringed the constraint in question. At one point, McMahan suggested that Agent must *reasonably* believe Victim is liable, if Agent is to avoid infringing the constraint on intentionally harming the innocent (McMahan 2011: 557–558). McMahan now suspects that even an *unreasonable* belief in Victim's liability would prevent Agent from infringing the constraint (personal communication). Either way, McMahan claims that (on his account) war will not be as difficult to justify as Lazar suggests; war may involve the killing of non-liable unjust combatants, but this killing need not infringe the stringent constraint on intentionally harming the innocent.<sup>80</sup>

To be clear, the issue is not whether an agent who kills, reasonably believing that her target is liable, intentionally harms an innocent person. Such an agent does intentionally harm an

<sup>79</sup> But see McMahan 2011: §2.2 and §3 for discussion of that claim.

<sup>80</sup> Note that McMahan is concerned with objective or fact-relative permissibility, and not subjective or belief-relative permissibility, despite his appeal to agents' beliefs about liability. There is no contradiction here, since fact-relative permissibility is permissibility relative to *all* the facts, including facts about agents' mental states (Lazar 2015a: 96). I follow McMahan in focusing on fact-relative permissibility throughout.

innocent person *in the sense* that the agent intends harm to a person who is, in fact, innocent.<sup>81</sup>

The question is whether intentionally harming an innocent person in *that* sense infringes the intuitive moral constraint that—perhaps among other things—speaks against intentionally harming those known to be innocent. McMahan thinks not.

McMahan's (2011) argument in favor of his view is flawed, however. That argument depends upon comparing three cases. In the first case, a pilot drops a bomb on a group of combatants despite the fact that a civilian is seen to be present and will foreseeably be killed. The second case is the same, except that the civilian has been disguised as a combatant, so the pilot cannot visually distinguish between the combatants and the civilian. Still, the pilot knows that the group of combatants contains one disguised civilian. The third case is just like the second, except that the pilot's weapon is a gun rather than a bomb, so that the pilot must shoot each individual dressed like a combatant (including the one who is really a civilian). McMahan argues that (a) in the first case the pilot does not infringe the constraint on intentionally harming the innocent, (b) the pilot's intentions need not change as we move to the second and third cases, and hence (c) the pilot need not infringe the constraint in the third case, where the pilot intentionally kills an individual who turns out to be non-liable.<sup>82</sup> But premise (b) is mistaken, since in case 3 (but not in case 1) the pilot *de re* intends to kill an innocent person (i.e., the pilot has an intention to kill that is directed at a person who is in fact innocent). Of course, McMahan thinks such a *de re* intention is not morally relevant in itself, but that view must be argued for.

---

<sup>81</sup> ...and puts this intention into action by harming the person in question.

<sup>82</sup> I have simplified the cases by supposing that in each case the agent is a pilot, rather than sometimes a pilot and sometimes a helicopter crew. My reconstruction of McMahan's argument is based largely on the following passage: "I have suggested that if, as most believe, the pilot need not intend to kill the civilian in [case] 1, he need not do so in [case] 2 either. The difference between [case] 2 and [case] 3 is that in 3 the basis of the crew's inability to discriminate is epistemic only, rather than both physical and epistemic. This difference does not necessitate a difference in intention" (2011: 558).

While McMahan's *argument* in favor of his view is flawed, I will nonetheless assume that the *view itself* is correct—in other words, that intentionally killing a person you (reasonably) believe is liable does not infringe the moral constraint on intentionally harming the innocent.

There are three reasons that I make this assumption.

First, it is worthwhile seeing whether this view of McMahan's even *could* provide a response to the responsibility dilemma, and what the *best* form of that response would be. In order to do these things, I will assume McMahan's interpretation of the moral constraint in question.

Second, that interpretation is plausible. To see this, consider two cases. In both cases, Hunter shoots and kills Victim, mistaking Victim for a deer. In the first case, Hunter aims at Victim. In the second case, Hunter aims at an actual deer, foreseeing that the bullet will not only kill the deer but will also kill Victim, who is behind the deer (and who Hunter mistakenly believes is a second deer). In the first case, Hunter *de re* intends to kill an innocent person—that is, Hunter intends to kill some thing, where that thing *in fact* is an innocent person. In the second case, Hunter merely foresees that he will kill some thing, where that thing in fact is an innocent person. Is the first killing intuitively morally worse than the second? I think not. If my judgment here is correct, this would support the view that there is no special moral constraint on *de re* intending to harm an innocent person.<sup>83</sup>

It is possible that there is a special moral constraint against *de re* intending to harm an innocent person *when you know they are a person* (but are unaware they are innocent), even if

---

<sup>83</sup> It is possible that such a constraint is operative in some cases but not in others—that it is sometimes but not always worse to *de re* intend to kill an innocent person (believing that they are not an innocent person) than to merely foreseeably kill an innocent person (believing the same). (See Kagan 1988 on the additive assumption and the ubiquity thesis.) But it is not clear why such a constraint would be operative in war and not in the case of Hunter, apart from the possibility considered in the next paragraph.

there is no special constraint against *de re* intending to harm an innocent person *when you do not know they are a person*. In other words, it is possible that the combatants McMahan and Lazar are discussing *do* violate the moral constraint on intentionally harming the innocent, even if the hunter does *not*. So the Hunter case does not provide decisive support for McMahan's interpretation of the moral constraint in question.

The case nevertheless provides some support for McMahan. For the case challenges the natural suggestion that the constraint in question speaks against *de re* intending harm to the innocent. And it is plausible that being (non-negligently) unaware that your target is a person is morally like being (non-negligently) unaware that your target is non-liable. For both are ways of being non-negligently ignorant of your target's moral status. So it is plausible that combatants who are unaware of their targets' innocence are like Hunter insofar as they do not infringe the moral constraint on intentionally harming the innocent.

The third reason that I will proceed on the assumption that McMahan's interpretation of this moral constraint is correct is that Lazar himself is committed to this interpretation. So if we are wondering whether the responsibility dilemma should drive us from McMahan's account of killing in war to Lazar's own account—presented at length in Lazar 2015b—it is only fair to grant McMahan the assumptions to which Lazar himself is committed.

Lazar believes that killing non-liable combatants can often be justified, and that part of the justification derives from the fact that refusing to fight against non-liable combatants may allow harm to come to one's family, friends, or comrades in arms. Lazar (2013) defends this view by reference to a case where, allegedly, it is permissible to kill a non-liable person in order to save a family member. One drawback of this argument is that the central case involves (1) unintentionally but foreseeably killing an individual one knows is non-liable, while the wartime

killings that Lazar hopes to justify involve (2) intentionally killing individuals one mistakenly believes are liable to be killed. Because of this discrepancy, Lazar (2016a) clarifies his position that 1 is relevantly similar to 2. In this way, Lazar hopes to support conclusions about wartime killing with his example involving unintentional killing.

In particular, Lazar (2016a) considers a case in which ten militants “have grabbed [a] civilian and hauled him into their bunker,” and “you’re unable to tell … apart” the militants and the civilian (§4). Lazar claims that if you have a sniper rifle, “you may permissibly kill each person in sequence,” even though this will involve intentionally killing a person (the civilian) who turns out to be non-liable. Lazar claims that “this is a kind of intentional killing of the innocent which is qualitatively similar to the … unintentional killing of the innocent which [Lazar has] argued can be justified by appeal to associative duties” (*ibid.*).

It is common ground between McMahan and Lazar that there is a special moral constraint on intentionally harming the innocent,<sup>84</sup> and that (i) unintentional (but foreseeable) killings of the non-liable do not infringe this constraint. And we have just seen that Lazar claims that (ii) such unintentional killings are “qualitatively similar” to intentional killings where each victim is reasonably believed to be liable. Hence Lazar is committed to the claim (which follows from i and ii) that the constraint on intentionally harming the innocent is not infringed by intentional killings where each victim is reasonably believed to be liable. In other words, Lazar and McMahan are committed to the same interpretation of the constraint on intentionally harming the innocent.<sup>85</sup> Because of this fact, and because the interpretation in question is a plausible one, I will assume that this interpretation is correct and consider whether and how it helps McMahan

---

<sup>84</sup> Lazar writes, for example: “Intended rights-violations are worse than unintended ones” (2015a: 97).

<sup>85</sup> To be more specific, Lazar is committed to McMahan’s 2011 version of that constraint, which required one’s belief in the liability of one’s target to be *reasonable*.

with the responsibility dilemma.<sup>86</sup>

### **3. Lesser-Evil Justification**

It might be thought that intentionally killing a non-liable unjust combatant can easily be permissible, given the assumption that such an act does not infringe the constraint on intentionally harming the innocent. The thought here is that since just combatants cannot tell which unjust combatants are liable to be killed, it is necessary for just combatants to kill both liable and non-liable unjust combatants in order to achieve their war aims. And if those aims are important enough, then there is a lesser-evil justification for killing non-liable unjust combatants. For while it is an evil to kill non-liable people, it may be a greater evil to fail to achieve one's just war aims.

This view has been suggested by McMahan in conversation. It also seems to be suggested by McMahan's remark that "the killing of non-liable unjust combatants is a foreseeable but unintended effect of the effort to kill those who are liable," and hence "is merely a part of the larger problem of 'collateral damage'" (2011: 557). Similarly, Bazargan (2013) suggests that killing non-threatening unjust combatants can be "necessary to prevent the harms imposed by effective unjust combatants," since killing the former is "unavoidable if just combatants are to engage in defense against effective unjust combatants" (183–184). Indeed, this assumption is crucial to Bazargan's solution to the responsibility dilemma.

The act of intentionally killing an unjust combatant, who turns out to be non-liable, *may* be justified. That could be the case, for example, if the death results from an artillery or air attack

---

<sup>86</sup> Lazar (2010: 202f, 2016b: 301f) presents a number of objections to McMahan's suggestions regarding how to interpret the constraint on intentionally harming the innocent. I believe that most or all of these objections do not have force against the particular way that McMahan's view is presented in this paper, but I do not have space to argue the issue.

that also kills liable unjust combatants. For if a (reasonable) belief in the non-liable combatant's liability prevents the killing from infringing the constraint on intentionally harming the innocent, then the attack is justified in roughly the same circumstances in which the collateral killing of a civilian would be justified.

On the other hand, a rifleman who targets, shoots, and kills a non-liable unjust combatant likely acts impermissibly. For shooting a non-liable unjust combatant is unlikely to avert a greater evil, let alone be *necessary* for averting a greater evil. After all, a non-liable combatant is not responsible for a significant threat of unjustified harm. (If they *were* responsible for such a threat, then they would be *liable* to be killed.) And killing liable unjust combatants will typically be sufficient for protecting the just combatants and achieving their war aims. Hence shooting a non-liable unjust combatant is unlikely to be either necessary or proportionate, and thus is likely to be impermissible (in the objective or fact-relative sense that is our focus throughout).

In response, it might be claimed that shooting a non-liable unjust combatant *can* produce good effects, insofar as casualties may contribute to a decision by the unjust state to cease hostilities. I am skeptical of this response for at least three reasons. First, the contribution that a single casualty makes to such a decision by the unjust state would seem to be extremely small—too small to make killing an innocent person proportionate. Second, even if shooting a non-liable unjust combatant can be proportionate, such an act will still be unnecessary if liable unjust combatants could have been killed instead. Third, the just combatant who shoots a non-liable unjust combatant would seem to violate the “requirement of justificatory intent” (Haque 2007: 637). For the just combatant’s motivation would likely be to eliminate a threat that the unjust combatant poses (or contributes to); while the normative reason that supports the killing would

be entirely different, since the unjust combatant makes little or no contribution to such a threat.<sup>87</sup>

I will assume, then, that shooting a non-liable unjust combatant is usually unjustified. If a significant proportion of unjust combatants are not liable to be killed, then war inevitably involves widespread violations of the *jus in bello* principles of necessity and proportionality. This result does not necessarily commit us to contingent pacifism, however.

Contingent pacifism is avoided if there are cases where the resort to war is justified. This means that our focus is *jus ad bellum*. Even if there are violations of *jus in bello*, as long as the principles of *jus ad bellum* are satisfied we can respond to the pacifist that going to war (or remaining at war) was the right thing to do. Violations of *jus in bello* are relevant here only insofar as they present obstacles to the satisfaction of the requirements of *jus ad bellum*—in particular, the requirement of *ad bellum* proportionality.

Suppose that a head of state is considering whether to initiate (or continue) a war in order to achieve some valuable goal. Among other issues, she should consider whether the war would be a proportionate means to that goal. She knows that wars typically involve many instances in which civilians are unintentionally killed, often as a side effect of attacks on military targets. The fact that such killings will occur counts against the war's being proportionate: unintentional infringements of the right to life are evils that must be weighed against the evils (if any) that would result from declining to wage war. If the latter evils are sufficiently grave, then the resort to war may be justified. In other words, the head of state may have a lesser-evil justification for

---

<sup>87</sup> Is killing a non-liable unjust combatant as a means of influencing his government especially hard to justify because it is an instance of “opportunistic” rather than “eliminative” killing (Quinn 1989: 344)? Perhaps not, insofar as opportunistic and eliminative killing are typically understood to be objectionable insofar as they are different ways of infringing the constraint on intentionally harming the innocent, and I have suggested that killing a non-liable unjust combatant need not infringe that constraint. It is possible, of course, that the opportunistic/eliminative distinction is morally relevant even in cases that do not involve the infringement of the constraint on intentionally harming the innocent.

acting in a way that she knows will result in the killing of innocent people.<sup>88</sup>

In the same manner, a war can be justified even if it involves many instances in which non-liable unjust combatants are killed. Such killings will often be impermissible because unnecessary or disproportionate. But the resort to war might nonetheless be permissible. If McMahan's views entail that unjust combatants are often not liable to be killed, then fewer wars will meet the requirement of *ad bellum* proportionality. That is not to say, however, that no wars will be proportionate—and hence it is not to say that McMahan is committed to a form of pacifism.

Note the role being played here by the assumption that the moral constraint on intentionally harming the innocent is not infringed by intentionally killing combatants who are mistakenly believed to be liable. Killings that infringe this constraint are *especially* objectionable. As a result, there might be no lesser-evil justification for initiating a war in which many killings occur that infringe the constraint on intentionally harming the innocent. Even so, there might be a lesser-evil justification for initiating war if the wrongful killings that result do not infringe that constraint.

In addition to considerations of *ad bellum* proportionality, the number of non-liable individuals who will be killed is also relevant to *ad bellum* necessity—that is, relevant to whether there is a morally superior means of achieving the just cause. (For example, the number of non-liable individuals who will be killed in war might make the difference to whether war is morally superior to a sanctions regime that itself would lead to many deaths.) But, again, contingent pacifism does not follow from the assumption that the killing of non-liable unjust

---

<sup>88</sup> This does not mean that *ad bellum* justification (or the head of state's deliberation) is *purely* a matter of lesser-evil justification. For the head of state can have a liability justification for causing (by initiating war) the deaths of those who are responsible for a threat of unjustified harm to others.

combatants presents an obstacle to satisfying *ad bellum* necessity.

It is important to note that in initiating war, a head of state is not herself infringing the constraint on intentionally harming the innocent. The head of state has the *de dicto* intention to win the war, and she may also have the *de dicto* intention to kill many of the enemy's combatants (or some similar intention). But there is no reason to believe that she has the (rather perverse) *de dicto* intention to kill non-liable enemy combatants. And presumably the head of state does not have a *de re* intention that infringes the moral constraint on intentionally harming the innocent, for presumably the head of state does not have *any* intentions directed at specific enemy combatants. So the head of state does not infringe the moral constraint in question. She neither intends *de dicto* nor intends *de re* that innocent people be killed.<sup>89</sup>

How much will killing of the non-liable increase if unjust combatants are often not liable to be killed, as Lazar claims? On average, it seems that about 50% of war-related deaths are civilians (Eckhardt 1989; contra the frequently cited figure of 90%, which has been debunked by Roberts 2010).<sup>90</sup> Suppose a war results in the death of 100,000 civilians and 100,000 combatants, where 50,000 of the latter group are unjust combatants. If all (and only) unjust combatants are liable to be killed, then the war results in the death of 150,000 non-liable people (100,000 civilians and 50,000 just combatants). If instead 50% of unjust combatants are not liable to be

---

<sup>89</sup> The *de dicto* intention to kill an innocent person is an intention to <kill an innocent person>, while the *de re* intention to kill an innocent person is an intention to <kill Victim>, where Victim is in fact an innocent person. The *de dicto* intention is not directed at any particular person. It is satisfied if and only if the agent kills at least one—any one—innocent person. The *de re* intention is satisfied if and only if the agent kills the particular innocent person at whom the intention is directed.

<sup>90</sup> Eckhardt's (1989) figure of 50% civilian deaths appears to exclude deaths due to war-related hunger and disease. For the deaths recorded in Eckhardt's table 1 (which are used to establish the civilian share of war-related deaths) are low enough that they cannot include the hunger- and disease-related deaths in table 6. For example, table 1 records 9,655,000 war-related deaths in the 19th century, while table 6 records 18,000,000 civilian deaths in the Taiping Rebellion alone, noting that most of these deaths were “caused by famine and disease.” Including hunger- and disease-related deaths—so that more than 50% of war-related deaths are civilians—would not weaken the argument I am about to make.

killed, then the war results in the death of 175,000 non-liable people (100,000 civilians, 50,000 just combatants, and 25,000 unjust combatants). This increase in the number of non-liable deaths is relatively modest. It will render some wars disproportionate or unnecessary, but not all.

The non-liable people killed specifically by the *just* side will increase by the same number—25,000—although this number will represent a greater percentage change. If we assume that each side in our example kills an equal number of civilians, then the number of non-liable people that the just side kills will increase from 50,000 (all civilians) to 75,000 (including 25,000 non-liable unjust combatants). This is still not an increase of enormous proportions. And the increase could be considerably less if we do not assume that a full 50% of unjust combatants are non-liable.

Indeed, what really matters is not the percent of unjust combatants who are non-liable, but the percent of unjust combatants *who are killed* who are non-liable. And the latter percentage should be lower than the former, given the relative safety of most of the roles Lazar cites as contributing little to threats of unjustified harm (including cooks, mechanics, and engineers). In other words, those unjust combatants who occupy the most dangerous jobs (such as infantry) are more likely to be liable to be killed. This consideration should help to limit the number of non-liable unjust combatants whose deaths McMahan must justify.

Lazar (2016b: 297f) suggests that the responsibility dilemma makes trouble for McMahan even if the first horn is not contingent pacifism, but rather the claim that fewer wars are justified than we had thought. But I do not see that as a particularly implausible implication of McMahan's view, especially if only *moderately* fewer wars are justified than we had thought. And the considerations discussed in the last few paragraphs suggest that McMahan's views will not have radical implications in this regard. McMahan need not claim that the Allies were

unjustified in fighting World War II, for example, or even that America's involvement in the Korean War was unjustified.

Two more points are relevant before moving on, although they do not bear directly on the responsibility dilemma.

First: Just as a head of state can have a lesser-evil justification for waging a war that involves the killing of non-liable individuals, so too an individual can have a lesser-evil justification for enlisting in a war despite that act setting the individual on a course that results in him killing non-liable individuals at a later time. The combatant will probably not have a lesser-evil justification for shooting non-liable individuals, when that time comes. The act of enlistment may nonetheless be justified if it is necessary to achieve some good end, such as helping to achieve a just cause.<sup>91</sup>

Second: The ideas developed in this section provide a response to Statman (2011). Whereas Lazar claims that just combatants inevitably kill non-liable unjust combatants, Statman argues that (on McMahan's view) even the killing of otherwise *liable* unjust combatants will often be impermissible because the killings will be unnecessary or ineffective. But the impermissibility of many acts of killing in war does not entail the impermissibility of the decision to wage war. Nor does it entail the impermissibility of the decision to participate in war. Statman is too fast to declare that "if the necessity condition and success condition apply to *jus in bello*, we seem to be on the safe road to pacifism" (442). It must be shown that there is never a lesser-evil justification for waging a war containing unnecessary and ineffective acts of violence,

---

<sup>91</sup> I intend to remain neutral with respect to the actualism/possibilism debate. It will be even easier to justify enlistment if we say that the individual should enlist-and-kill-only-the-liable, and if we take this fact to entail that the individual should enlist (despite the fact that enlisting will not *actually* lead to the individual killing only the liable).

or for participating in one.<sup>92</sup>

#### **4. Objections**

Before concluding, I consider several objections to McMahan's views on liability and the relevance of intentions. First, there is a worry raised for McMahan by Saba Bazargan (2013). Bazargan claims that a just combatant “*does* violate the constraint against killing intentionally,” if “her reasonable belief that the target is liable is not what explains why she kills her target” (181n6). And Bazargan worries that just combatants would kill their targets even in the absence of the belief that their targets are “effective”—that is, even in the absence of the belief that their targets “contribute substantially to the war’s aims” (180). But it is doubtful that many combatants believe liability in war is grounded in individual contributions to war aims. As Bazargan himself writes, it is likely that “many combatants see the enemy uniform as a sufficient reason to kill” (181n6). So it is not clear why the killing of unjust combatants must counterfactually depend upon the belief that unjust combatants are effective. It would seem that a just combatant’s act could be explained by the belief that her target is liable, even if her act is not explained by the belief that her target is effective.<sup>93</sup>

---

<sup>92</sup> Compare: You know that if you become a doctor, then you will end up making some medical mistakes and doing some things that are wrong, objectively speaking. That does not mean it is impermissible to become a doctor. I ignore the complication that Statman, unlike McMahan, is concerned with *subjective* permissibility—and, in particular, with combatants’ *reasonable beliefs* about necessity and effectiveness. This complication should not affect the effectiveness of my response to Statman. (Regarding McMahan’s focus on objective or fact-relative permissibility—despite his reference to combatants’ beliefs about their enemies’ liability—see my note 80.)

<sup>93</sup> I have said it is doubtful that many combatants believe liability in war is grounded in individual contributions to war aims. However, Bazargan is assuming (with McMahan 2011) that it is relevant whether a combatant’s belief in her target’s liability is *reasonable*. And Bazargan may be thinking that in order for belief in the target’s liability to be reasonable, one must believe that liability in war is grounded in individual contributions to war aims. It must be argued, however, that there is such a requirement on the reasonableness of beliefs about liability.

Next, there is a set of objections from Lazar (personal communication). First, Lazar worries that on McMahan's view, if A kills B while mistakenly believing that B is liable to be killed, then A has not infringed B's right to life. And Lazar finds it implausible that A's mistake could deprive B of his right to life. Contrary to Lazar's suggestion, however, McMahan is not committed to saying that A has not infringed B's right to life. McMahan is only committed to saying that A has not infringed the especially stringent constraint on intentionally harming the innocent. And one can infringe someone's right to life without infringing the constraint on intentionally harming the innocent—as happens whenever a non-liable civilian is killed as an unintended side effect of attacking a military target.

Second, we have seen that Lazar claims (1) McMahan is committed to accepting widespread civilian liability. And I have been developing McMahan's view that (2) one does not infringe the constraint on intentionally harming the innocent when one believes (or reasonably believes) that one's victim is liable. Lazar worries that 2 interacts with 1 in a way that makes McMahan's view even more permissive regarding killing civilians. For if McMahan is correct, then it would appear that just combatants can justifiably believe enemy civilians are liable to be killed (given 1); which means that if they attack enemy civilians, they do not infringe the constraint on intentionally harming the innocent (given 2). So 2 has the effect of further lowering the moral barriers, on McMahan's view, to attacking civilians.

Contrary to the objection just sketched, however, the acceptance of McMahan's views should not lead just combatants to believe that civilians who they might attack are liable to be killed. For in the absence of special evidence regarding the liability of enemy civilians, a belief in their liability would be reasonable only if the average (adult) unjust civilian is liable to be killed. And McMahan is not committed to the majority of such civilians being liable to lethal

harm, even if he is committed to many being liable (see McMahan 2011: §2.2, §3). Indeed, Lazar (2016b) argues only that McMahan should accept the “Overlap Hypothesis,” according to which many civilians are as liable as many combatants—which does not entail that most unjust civilians are liable to lethal harm.<sup>94</sup>

The previous objection claimed that 2 combines with 1 in a way which makes McMahan’s view on killing civilians even more permissive than it otherwise would be. A related objection focuses more narrowly on 2 itself. Given 2, those who intentionally kill civilians, mistakenly believing they are liable to lethal harm, do not infringe the constraint on intentionally harming the innocent. On McMahan’s view, and other things being equal, such killings would have something like the moral status of foreseen but unintended (“collateral”) killings. And that may seem to underrate how objectionable such killings are.

I take this to be one of the more forceful objections to McMahan’s view of the constraint on intentionally harming the innocent. Nonetheless, there is much to be said in McMahan’s defense. First, collateral killings can be grievously wrong, as when there are failures of proportionality or necessity. So saying that, on McMahan’s view, intentional killings of civilians can have the moral status of collateral killings is by no means to dismiss them as objects of moral concern. Second, McMahan could limit the counterintuitiveness of his view at this point by repudiating the idea that *any* belief in the liability of one’s victim is relevant, instead embracing the idea that only a *reasonable* belief in the liability of one’s victim can prevent a killing from infringing the constraint on intentionally harming the innocent. This move would limit the

---

<sup>94</sup> A caveat: claim 2 may indeed make McMahan’s view more permissive about killing *certain types* of civilians, namely, civilians that can be reasonably assumed to be liable. (Perhaps certain types of government workers can be reasonably assumed to be liable, when they work for an unjust belligerent.) But this is not nearly as disturbing as the possibility that the combination of claims 1 and 2 make McMahan’s view more permissive with respect to *indiscriminately* killing civilians—which, I assume, is the objection Lazar has in mind.

number of killings that McMahan must say have something like the moral status of collateral killings. Whether there is a residual problem for McMahan may depend upon whether the perpetrators of paradigm instances of terrorism reasonably believe their victims are liable. For if the perpetrators of such killings do reasonably believe their victims are liable, and if we nonetheless are confident that such killings are morally worse than unjustified collateral killings, then we should reject the claim that reasonable belief in liability is relevant in the way McMahan suggests. It is, however, beyond the scope of this paper whether terrorists often have reasonable beliefs in the liability of their victims—reasonable, for example, in light of the testimony by their family, community, or religious leaders.<sup>95</sup>

Finally, there is a question about how McMahan's view accommodates degrees of belief. On McMahan's view, it matters tremendously whether an agent *believes* her target is liable to be harmed. What degree of belief, or credence, is required for an agent to count as having such a belief? However we answer that question will lead to a problem. For suppose combatant A intentionally kills target 1 (who turns out to be non-liable), while combatant B intentionally kills target 2 (who also turns out to be non-liable). Combatant A has just enough credence to count as reasonably believing that target 1 is liable to lethal harm. Combatant B is just shy of the credence needed to count as reasonably believing that target 2 is liable to lethal harm. Otherwise, combatants A and B are similarly situated. Without revision, McMahan's view would seem to entail that combatant B acts much more wrongly than combatant A, since only combatant B infringes the constraint on intentionally harming the innocent. Yet it may seem implausible that there is a large moral difference here, given that these two combatants have very similar credences in their respective targets being liable to the relevant harm.

---

<sup>95</sup> For his part, McMahan thinks that such beliefs are often not reasonable (2011: 558).

In response to this problem, one possibility would be to say that while combatant B does indeed infringe the constraint in question, this infringement is not seriously wrongful in light of B's relatively high credence in his target being liable. This response, however, gives up much of the traditional character of the moral constraint on intentionally harming the innocent. For this constraint has usually been thought of as a bright moral line which it is seriously *pro tanto* wrong to cross. Indeed, the response just sketched would move McMahan quite close to Lazar's view that, other things being equal, a killing (whether intentional or merely foreseen) is morally worse the likelier it is that one's victim is innocent, where likelihood is understood in terms of one's credences (Lazar 2015a, 2019).

In any case, the problem at hand is not unique to McMahan's account. Rather, the need to accommodate degrees of belief is faced by anyone who claims that there can be a significant moral difference between possessing and lacking certain beliefs.<sup>96</sup> So while I am uncertain about the best way for McMahan's view to accommodate degrees of belief, I am hopeful that some accommodation will be available.

## **5. Concluding Summary**

I have argued that McMahan faces a serious risk of being committed to both horns of the responsibility dilemma, contrary to Lazar (2010). I then developed one response to the contingent pacifist horn of the dilemma. This response depends upon a specific view of the moral constraint on intentionally harming the innocent. I suggested that while McMahan's own argument for that view is flawed, the view itself is plausible; and, moreover, that Lazar himself is

---

<sup>96</sup> Essentially the same issue will also arise for anyone who claims that there can be a significant moral difference between possessing and lacking certain intentions—at least this is true if intentions, like beliefs, come in degrees (as argued in Goldstein 2016 and Shpall 2016).

committed to that view. I have not fully addressed all of the potential issues with McMahan's position, such as the need to accommodate degrees of belief. Nevertheless, I hope to have clarified the dialectical situation in which McMahan finds himself, and to have made progress in developing one response to the responsibility dilemma.<sup>97</sup>

## References

- Bazargan, Saba. 2013. Complicitous liability in war. *Philosophical Studies* 165.1: 177–195.
- Eckhardt, William. 1989. Civilian deaths in wartime. *Bulletin of Peace Proposals* 20.1: 89–98.
- Goldstein, Simon. 2016. A preface paradox for intention. *Philosophers' Imprint*. 16.14: 1–20.
- Haque, Adil Ahmad. 2007. Torture, terror, and the inversion of moral principle. *New Criminal Law Review* 10.4: 613–657.
- Kagan, Shelly. 1988. The additive fallacy. *Ethics* 99.1: 5–31.
- Lazar, Seth. 2010. The responsibility dilemma for *Killing in War*: A review essay. *Philosophy and Public Affairs* 38.2: 180–213.
- . 2013. Associative duties and the ethics of killing in war. *Journal of Practical Ethics* 1 (1): 3–48.
- . 2015a. Risky killing and the ethics of war. *Ethics* 126.1: 91–117.
- . 2015b. *Sparing Civilians*. New York: Oxford University Press.
- . 2016a. The associativist account of killing in war. In *Global Political Theory*, edited by David Held and Pietro Maffettone. Cambridge: Polity Press.
- . 2016b. Liability and the ethics of war: A reply to Strawser and McMahan. In *The Ethics of Self-Defense*, edited by Christian Coons and Michael Weber. New York: Oxford University Press.
- . 2019. Risky killing: How risks worsen violations of objective rights. *Journal of Moral Philosophy* 16.1: 1–26.
- McMahan, Jeff. 2004. The ethics of killing in war. *Ethics* 114.4: 693–733.
- . 2005. The basis of moral liability to defensive killing. *Philosophical Issues* 15: 386–405.
- . 2006. On the moral equality of combatants. *Journal of Political Philosophy* 14.4: 377–393.
- . 2009. *Killing in War*. New York: Oxford University Press.

---

<sup>97</sup> For discussion of some complications that I have ignored, see the appendix following the references.

- \_\_\_\_\_. 2011. Who is morally liable to be killed in war. *Analysis* 71.3: 544–559.
- \_\_\_\_\_. 2013a. The conditions of liability to preventive attack. In *The Ethics of Preventive War*, edited by Deen K. Chatterjee. Cambridge University Press.
- \_\_\_\_\_. 2013b. War crimes and immoral action in war. In *The Constitution of the Criminal Law*, edited by R.A. Duff, Lindsay Farmer, S.E. Marshall, Massimo Renzi, and Victor Tadros. Oxford: Oxford University Press.
- \_\_\_\_\_. 2013/2014. Proportionate defense. *Journal of Transnational Law and Policy* 23: 1–36.
- \_\_\_\_\_. 2014. Proportionality and just cause: A comment on Kamm. *Journal of Moral Philosophy* 11.4: 428–453.
- Quinn, Warren S. 1989. Actions, intentions, and consequences: The Doctrine of Double Effect. *Philosophy and Public Affairs* 18.4: 334–351.
- Roberts, Adam. 2010. Lives and statistics: Are 90% of war victims civilians? *Survival* 52.3: 115–136.
- Shpall, Sam. 2016. The Calendar Paradox. *Philosophical Studies* 173.3: 801–825.
- Statman, Daniel. 2011. Can wars be fought justly? The necessity condition put to the test. *Journal of Moral Philosophy* 8.3: 435–451.
- Walzer, Michael. 1977. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. New York: Basic Books.

## **Appendix: Remaining Difficulties**

On McMahan's behalf, I have developed one line of response to the responsibility dilemma.

Before concluding I note several difficulties that remain for McMahan's views on liability and the relevance of intentions, as well as how those difficulties might be addressed.

First, I have simplified McMahan's position in the foregoing. His 2011 paper actually suggests that those who mistakenly believe their victim is liable do not infringe the constraint on intentionally harming the innocent only if *both* (i) their mistaken belief is reasonable, *and* (ii) their mistaken belief about liability is due to a mistake of fact rather than a mistake about morality (2011: 538). Believing a civilian is a combatant is a mistake of fact, for example, while a mistake about morality would be believing that a pro-war attitude can, by itself, make one liable to be killed.

Holding that condition (ii) is necessary to avoid infringing the constraint in question would prevent McMahan from claiming that just combatants who kill non-liable unjust combatants typically do not infringe that constraint. After all, the relevant mistake that just combatants are making is likely a mistake about morality, rather than a mistake about non-moral fact. In particular, the reason why just combatants believe their targets are liable to be killed is likely *not* that they (*a*) accept the revisionary view that liability in war is grounded in responsibility for a threat of unjustified harm, and (*b*) believe (sometimes mistakenly) that the unjust combatants they kill are responsible for threats of unjustified harm. Rather, the reason why just combatants believe their targets are liable to be killed is likely that they accept the orthodox view that *all* combatants—or at least all *unjust* combatants—are liable to be killed, regardless of whether they are responsible for a threat of unjustified harm. If accepting this orthodox view is a mistake (as McMahan claims), it is a mistake about morality rather than about non-moral fact. Since McMahan claims that just combatants typically do not infringe the relevant moral constraint by killing non-liable unjust combatants, he is not in a position to claim that only mistakes of fact are relevant to whether that constraint is infringed.

McMahan can simply abandon the claim that mistaken beliefs about liability must be based on mistakes of non-moral fact. Indeed, McMahan now claims that believing one's target is liable is sufficient to avoid infringing the constraint on intentionally harming the innocent, whether or not one's belief is reasonable and whether or not one's mistaken belief in liability is based on a mistake of non-moral fact (personal communication). There are, however, problems that remain.

For one thing, it seems McMahan must give up the claim that unjust combatants infringe the principle of discrimination by killing just combatants (2006: 379; 2009: 15–16; 2013b: 164).

For McMahan takes the principle of discrimination to be “a moral constraint on the intentional killing of innocent people” (2013a: 139). And McMahan now holds that combatants do not intentionally kill innocent people (in the morally relevant sense) if they believe their targets are liable to be killed. It follows that unjust combatants do not infringe the principle of discrimination if, as seems likely, they believe that their targets are liable to be killed.

Indeed, McMahan may have to give up the claim that unjust combatants infringe the principle of discrimination, whether or not a belief in the liability of one’s target is sufficient to avoid infringing the constraint on intentionally harming the innocent. For unjust combatants may not only *believe* that just combatants are liable, but may *reasonably* believe so. As I said above, the reason why just combatants believe their targets are liable to be killed is likely that they accept the orthodox view that *all* combatants—or at least all *unjust* combatants—are liable to be killed, regardless of whether they are responsible for a threat of unjustified harm. So just combatants’ belief that their targets are liable to be killed is reasonable only if it is reasonable for them to accept an orthodox view about liability. And if it is reasonable for just combatants to accept an orthodox view about liability, a large part of *why* it is reasonable will be that their peers, authorities, and society in general believe and profess an orthodox view of liability. But unjust combatants are also told by their society that all combatants are liable to be killed—or that their enemies are unjust combatants, and that all unjust combatants are liable to be killed. So if just combatants reasonably believe that their targets are liable to be killed, it may be that unjust combatants do as well—in which case unjust combatants typically do not infringe the principle of discrimination.

Importantly, acknowledging that unjust combatants do not infringe the principle of discrimination would not prevent McMahan from saying that unjust combatants act

impermissibly when they kill just combatants. For such killings will typically fail to satisfy the requirement of proportionality: unjust combatants typically achieve no good aims that could weigh against the harm they do to just combatants (McMahan 2004: §5).

Next, while McMahan now rejects his 2011 claim that the mistaken belief in the liability of one's target must be reasonable, he may be mistaken. Haque (2012), for one, believes that the mistake must be reasonable if an agent is to avoid infringing the moral constraint on intentionally harming the innocent. And if Haque is correct, then McMahan faces another problem. McMahan (2006) writes:

[W]hat combatants are considering is whether to kill people of whom they have no personal knowledge and what they need to know is whether that would be permissible. The standard for whether a belief about this matter is reasonable is high. For a belief about this matter to count as reasonable, the person whose belief it is must have done some epistemic work—for example, by investigating the facts and engaging in serious moral reflection. (390)

It is doubtful that a significant proportion of just combatants do this epistemic work. Hence it is doubtful that just combatants generally have the reasonable belief that their targets are liable to be killed. (And hence it is doubtful that just combatants who kill non-liable unjust combatants avoid infringing the constraint on intentionally harming the innocent, if Haque is correct that mistaken beliefs in liability must be reasonable.) Of course, McMahan could give up the claim that combatants must do some “epistemic work” in order to have reasonable beliefs. At least, if foreign forces invade your country, then perhaps you need not do much epistemic work before deciding that it is permissible to kill the invading combatants. (Other causes for war, such as those found in cases of humanitarian intervention and preemptive or preventive strikes, may be another story.)

## The Modal Fog of War

When is it permissible to harm others? By which I mean, roughly: when is it morally permissible to physically harm non-consenting persons, such as occurs in war and self-defense? This question motivates the following discussion, though I do not attempt to identify the conditions under which harming is permissible. Instead, I present a problem for any plausible account of these conditions. On any plausible account, the permissibility of an act can depend not only on the harm it causes, but also on the harm it prevents. Yet it is not clear how to draw the distinction between *preventing* harm and merely *refraining from causing* harm. This presents an obvious problem, as well as a less obvious one. The obvious problem is that if we cannot tell *whether* some act would prevent harm, then we certainly cannot tell *what* harm it would prevent, and this may leave us in the dark about the act's permissibility. The less obvious problem is that if we cannot tell whether some act would prevent harm, this threatens to leave us in the dark about the harm that would be prevented by the agent's *other* options. After presenting this latter problem, I consider several possible responses. I endorse one and note its limitations.

The structure of the paper is as follows. Section 1 discusses the conditions on an act's preventing harm. It argues that non-circular necessary and sufficient conditions are not forthcoming; but that in any particular case, we can typically identify considerations that lead us to the correct conclusion. Section 2 focuses on the problem cases, in which harm is counterfactually related to some act but it is unclear whether the act should be said to prevent the harm. The extent to which this unclarity undermines moral judgment can be limited but not eliminated.

Although my focus is on the harm caused or prevented by an act, I do not assume that these are the only considerations relevant to permissibility. I assume only that they are often

among the considerations in virtue of which an act is permissible or impermissible. These considerations may sometimes involve foreseen or expected consequences, or something similar, though for convenience I refer to actual consequences in the following.

## 1. Conditions on Preventing Harm

### 1.1 Accounting for Cases

Abe dislodges several rocks at the top of a hill, putting them on a collision course with Bea. Bea sits at the bottom of the hill, wearing noise-cancelling headphones and engrossed in a book. Luckily you are driving by with a truckload of bowling balls. You throw one into the path of the rocks, the rocks are deflected, and Bea is unharmed. For brevity, call your act *the throw*. What harm does the throw prevent?

A simple answer is that the throw prevents whatever harm would have occurred, had the throw not occurred. This is not the right answer, at least not for a humanitarian such as yourself. Had the throw not occurred, you would have found some other way to save Bea. You do have a truck, after all, as well as a whole truckload of bowling balls. Because there is no harm that would have occurred had the throw not occurred, our first answer entails that the throw prevents no harm at all. This is the wrong result. The throw *actually* prevents rocks from colliding with Bea, even if another act otherwise *would have*.<sup>98</sup>

---

<sup>98</sup> Kvanvig's (1986: 10–11) analysis of prevention gives us the “simple answer” just considered, as does the initial account offered by Dowe (2000: 132–34). The final version of Dowe’s account faces problems similar to those discussed below; see Koons 2003 (246–47). Counterfactual conditionals should be interpreted in a non-backtracking manner throughout. In the present case, one should suppose that had the throw not occurred *and had nothing gone differently up until that point*, you would have done something else to save Bea from harm.

According to a second answer, the throw prevents whatever harm would have occurred, had you done nothing at all.<sup>99</sup> Clearly there will be questions about what *counts* as doing something. (Breathing? Remaining perfectly still through great effort?) But the second answer also faces problems that are more serious, or at least more relevant to our discussion. For there are details of the case that have not yet been mentioned. Just after throwing the first bowling ball, you grab a second and roll it into the path of the rocks. Call this act *the roll*. The throw deflects some of the rocks heading for Bea and the roll deflects the rest. Had you done nothing, then neither the throw nor the roll would have occurred, and all the rocks would have collided with Bea. Hence our second answer entails that the throw prevents the harm that would have resulted from these collisions—*including* the collisions prevented, in actuality, by the roll. Hence our second answer overcounts the harm prevented by the throw.<sup>100</sup>

This leads us to a third answer: the throw prevents whatever harm would have occurred, had the throw not occurred and had you done nothing other than what you do in actuality. Again there will be questions of detail. (What counts as doing the *same* thing as you do in actuality?) Putting such questions aside for the moment, the third answer appears to be inadequate. Had the throw not already deflected some of the rocks, the roll would have deflected *all* of them. Or so we can suppose.<sup>101</sup> In that case, no harm would have occurred, had the throw not occurred and had you done nothing other than you do in actuality. So our third answer entails that the throw

<sup>99</sup> Presumably this means: had you done nothing from the time the throw actually occurs until whenever it is too late to prevent the threatened harm.

<sup>100</sup> Note that the problem posed for this answer would equally affect the following one: the throw prevents whatever harm would have occurred, had you done nothing to the rocks or Bea.

<sup>101</sup> This assumption simply focuses our attention on a specific case. If our answer leads to the wrong conclusion in this case, then we have reason to revise the answer again.

prevents no harm. This is the wrong conclusion, given that the throw actually prevents some of the rocks from colliding with Bea.

There is an objection that might be made at this point. According to the argument just offered, our third answer goes wrong because the roll would have deflected all the rocks, had the throw not occurred and had you done nothing other than what you do in actuality. One might object as follows:

*(The objection)* A possible world where the roll deflects all the rocks is a world where you do something *other* than you do in actuality. For in such a world, but not in actuality, you roll a single bowling ball into all the rocks. So it cannot be true that the roll would have deflected all the rocks, had the throw not occurred and had you done nothing other than what you do in actuality.

In short, the claim is that <the roll's deflecting all the rocks> is incompatible with <your doing nothing other than you do in actuality><sup>102</sup>.

Even if this objection is correct, our third answer still goes wrong—it just goes wrong in a different way. For ease of reference, I will use the following as shorthand:

*C* is the condition requiring that you do nothing other than what you do in actuality.

*Throw-deflected rocks* are rocks deflected by the throw in actuality.

*Roll-deflected rocks* are rocks deflected by the roll in actuality.

Now suppose that the objection is correct. In other words, suppose that C is incompatible with the roll's deflecting rocks other than those it deflects in actuality. Then in any world where C is satisfied, the roll does *not* deflect rocks other than those it deflects in actuality. In particular, in any world where C is satisfied, the roll does not deflect the *throw-deflected* rocks. And, crucially,

---

<sup>102</sup> Angle brackets are included simply for readability.

in any *nearby* world where C is satisfied *and* the throw does not occur, the roll deflects *neither* the throw-deflected *nor* the roll-deflected rocks. This is because the throw-deflected rocks are located between you and the roll-deflected rocks, in much the way that bowling pins up front stand between a bowler and pins further back. If the throw were not to occur, then the roll would have to push the throw-deflected rocks out of the way in order to reach the roll-deflected rocks, violating condition C. So if C had been *satisfied* and the throw had not occurred, then the roll would not have reached the roll-deflected rocks. Those rocks would not have been deflected and would have continued on to collide with Bea. Our third answer entails that the throw prevents the harm that would have resulted. (For it is among the harm that would have occurred, had the throw not occurred and had you done nothing other than what you do in actuality.) This overcounts the harm prevented by the throw, crediting it with harm actually prevented by the roll. That is the conclusion we reach, at least, if we assume that the objection is correct. So our third answer goes wrong even if we grant the objection.<sup>103</sup>

We have now considered three answers to the question of what harm the throw prevents. The details of the case were introduced in stages, and one of our three answers was offered at each stage. The first answer appeared to be adequate given the initial description of the case, but it proved to be inadequate as more information was disclosed; and likewise for each of the other two answers. Rather than continue to introduce revisions and offer counterexamples, I turn to

---

<sup>103</sup> Note that the problems posed for this answer would equally affect the following one: the throw prevents whatever harm would have occurred, had the throw not occurred and had you done nothing to the rocks or Bea other than what you do in actuality.

slightly more theoretical considerations, and to the lessons that might be drawn from the foregoing.<sup>104</sup>

### *1.2 Taking Stock*

We have been working with a counterfactual account of harm prevention. The basic idea is that an act prevents whatever harm does not *actually* occur but *would have*, had certain conditions been met.<sup>105</sup> We considered several possibilities for what these conditions might be, as follows:

1. The act does not occur.
2. The act does not occur, and the agent does nothing.
3. The act does not occur, and the agent does nothing other than what they do in actuality.

None of these three possibilities proved to be correct, as each was prone to undercount or overcount the harm prevented by the relevant act.

It may appear that we were zeroing in on the conditions we need. After all, there is a wide range of cases in which our first answer leads to the wrong conclusion.<sup>106</sup> And our second answer goes wrong in a similarly wide range of cases.<sup>107</sup> By contrast, it may seem that our third answer gets things just about right. There appear to be counterexamples, but perhaps only in a narrow range of cases that involve preemption. (Above, the roll's deflection of certain rocks was preempted by the throw.) One might think that we can continue to refine our account, weeding out the remaining problem cases.

<sup>104</sup> But see the appendix for one more epicycle. For skepticism about the use of counterfactuals in a related context, see Walen 2019: ch. 2, §4; ch. 3, §4.2; ch. 9, §2.1.

<sup>105</sup> No harm *actually* comes to Bea in the case considered above. This allowed us to equate the harm prevented with the harm that would have occurred, had the relevant conditions been met.

<sup>106</sup> Namely, any case in which the harm prevented by the act would have been prevented by other means, had the act not occurred.

<sup>107</sup> Namely, any case in which the agent performs more than one act that prevents harm.

Unfortunately, the prospects look dim for improving on our third answer. Consider an arbitrary act, ACT. Say that *ACT-prevented harm* is any harm prevented by ACT in actuality. We need conditions such that had they been satisfied, all ACT-prevented harm would have occurred. (Otherwise, our account will compare the actual world with one in which some ACT-prevented harm does not occur, and hence will undercount the harm prevented by ACT.) In particular, we need conditions such that had they been satisfied, the agent would not have prevented any ACT-prevented harm. Our arbitrarily selected ACT could turn out to be any act whatsoever, performed in any circumstances, and so we need to cast a wide net. But we cannot simply adopt conditions that require the agent not to act at all. (That is the mistake made by our second answer.) Nor can we adopt conditions that require the agent to do nothing other than what they do in actuality. (That is the mistake made by our third answer.)

What we need is a description of the acts to be ruled out—the acts that our conditions should require the agent *not* perform. The description should capture all and only the acts to be ruled out.<sup>108</sup> Yet the only discernible feature shared by all and only the acts to be ruled out is this: they prevent harm that is prevented, in actuality, by ACT. The ways that such harm might be prevented are numerous and varied, and a description that captures them all will capture much else as well.<sup>109</sup>

---

<sup>108</sup> For we must neither undercount nor overcount the harm prevented by ACT, as our first and second answers did, respectively. (Our third answer either undercounted or overcounted, depending on whether we granted the objection at the end of section 1.1.)

<sup>109</sup> Perhaps all of the acts to be ruled out are things that the agent does to the threat or potential victim. But we cannot simply adopt conditions requiring that the agent does nothing to the threat or potential victim, or that the agent does nothing to the threat or potential victim *other* than what they do in actuality. See notes 100 and 103. See the appendix for criticism of a related answer: ACT prevents whatever harm would have occurred, had ACT not occurred and had the agent's other acts produced *all and only* the effects that they do in actuality.

It is not clear what we can do, other than to say that ACT prevents whatever harm does not actually occur but would have, had ACT-prevented harm not been prevented.<sup>110</sup> But how are we to know what harm would have occurred, had ACT-prevented harm not been prevented? Presumably, we must imagine a scenario in which the agent performs only acts that do *not* prevent ACT-prevented harm, and imagine how events unfold in such a scenario. This seems to require knowing *which acts* would prevent ACT-prevented harm. (How else could we be confident that the imagined scenario is one in which the agent performs no such acts?) Knowing which acts would prevent ACT-prevented harm, in turn, seems to require knowing *what harm* this is. In other words, it seems that we must know what harm *is* prevented by ACT. But if we *do* know what harm is prevented by ACT, then there is *no need* to go through this whole process, the point of which was to learn what harm is prevented by ACT.<sup>111</sup>

There may be a temptation to admit that we lack non-circular necessary and sufficient conditions on an act's preventing some harm, but to deny that this matters. For it may seem as though there are relatively few cases in which our account will go wrong. Perhaps we can settle for the claim that an act prevents whatever harm does not actually occur but would have, had the act not occurred and had the agent done nothing other than what they do in actuality. This answer misled us in the case of Abe and Bea, but only because the second bowling ball would have deflected all the rocks, had the first bowling ball not already deflected some of them. This case may seem contrived. In the real world, one might think, backup preventers are not waiting in the wings. This may lead us to conclude that our account will be adequate for most purposes.

---

<sup>110</sup> ...and had the agent prevented all of the *other* harm that they prevent in actuality.

<sup>111</sup> This paragraph expresses a very plausible worry that the reader is likely to have at this point. The next section will suggest a way in which this worry is misguided, however.

We should resist this temptation, however. For one thing, backup preventers often *are* waiting in the wings. There are at least two types of cases in which this is true. First, there are cases in which multiple actions are directed at achieving the same goal (or closely related goals). Consider an airstrike on an enemy position that has already been weakened in ground combat. In such a case, it is likely that the airstrike would have destroyed some of the targets actually destroyed by the ground assault, had the latter not occurred. Second, there are cases in which an action has wide-ranging effects. Perhaps one state attacks another to deter continued rights infringements of some sort; later the foreign policy of the first state changes in some way, though not with the aim of deterring any rights infringements; and the policy change would have prevented some of the infringements actually deterred by the war, had the latter not occurred. Cases involving backup preventers, or preempted prevention, are not uncommon. So we should not settle for an answer that goes wrong in such cases.

In section 1.1, we saw the difficulty of finding conditions on an act's preventing some harm. The present section offered a diagnosis of that difficulty. To identify the harm that some ACT prevents, we cannot look to a world in which ACT does not occur but some *other* act prevents the ACT-prevented harm.<sup>112</sup> But we also cannot look to a world in which the agent *fails* to prevent harm that another act of theirs prevents in actuality.<sup>113</sup> Because of this latter fact, we cannot simply look to the harm that would have occurred, had the agent done nothing. We must instead find a description of the acts to be ruled out, the acts we do *not* want in the possible

---

<sup>112</sup> This would undercount the harm prevented by ACT.

<sup>113</sup> This would overcount the harm prevented by ACT.

world that we compare to actuality. Yet the only description that is fully adequate seems to be, “acts that prevent ACT-prevented harm,” a description that appears to be unhelpful.<sup>114</sup>

### *1.3 Intermediate Conclusions*

The arguments of the previous section should not lead us to reject a counterfactual approach to thinking about the prevention of harm. Those arguments cast doubt on our finding a non-circular counterfactual conditional that is *fully general*—a counterfactual that leads to the correct conclusion for *any* act whatsoever. Prevention is nonetheless closely tied to counterfactuals. It is simply that different counterfactuals are relevant in different cases. In any particular case, the harm prevented *is* picked out by some counterfactual or other.

Consider a variation on the case of Abe and Bea. If it were true that *all* you do is throw the one bowling ball into the path of the rocks, then it *would* be true that the throw prevents whatever harm would have occurred had you done nothing.<sup>115</sup> If Bea would have been killed, then the throw prevents her death. If she would only have suffered a broken bone, then the throw prevents the fracture. If the rocks would have missed Bea completely, then the throw prevents no harm at all.

In the actual version of the case discussed in section 1.1, the relevant counterfactuals are more complicated. In that version of the case, the throw deflects some of the rocks and the roll

<sup>114</sup> Klocksiem’s (2012) defense of the counterfactual comparative account of harm provides an instructive comparison. His response to putative counterexamples relies on the notion of “the nearest *relevant* possible world in which [event] *e* does not occur” (288, my emphasis). He does not provide conditions on a world’s being relevant, however. That may not be a problem for Klocksiem’s argument, but it is indicative of the difficulty of specifying, in general terms, *what* we should be comparing to the actual world. (Klocksiem is concerned with a world’s relevance to whether an event *constitutes harm*, but similar issues arise when we ask whether an act prevents harm.)

<sup>115</sup> I assume that had *you* not prevented the rocks from harming Bea, then *no one* would have. Without this assumption, a different counterfactual would be needed.

deflects the rest; but had the throw not occurred, the roll would have deflected all of the rocks. If the case is not revealed to involve any further complications, then the throw prevents whatever harm would have occurred, had the throw-deflected rocks *not* been deflected and the roll-deflected rocks *still* been deflected. Equivalently, the throw prevents whatever harm would have been caused specifically by the throw-deflected rocks, had they not been deflected.<sup>116</sup>

Better yet, the harm prevented by the throw can be specified without referencing the rocks that were *actually* deflected by the throw. In order to more easily illustrate the idea, I will make the following simplifying assumption: you had no way to intervene *other* than throwing the first bowling ball, rolling the second bowling ball, or doing both. Now take the harm that would have occurred, had neither the throw nor the roll occurred. Call this *harm<sub>A</sub>*. This is the harm prevented by the conjunction of the throw and the roll.<sup>117</sup> Next take the harm that would have occurred, had the roll not occurred. Call this *harm<sub>B</sub>*. This is the harm prevented by the roll.<sup>118</sup> The *difference* between *harm<sub>A</sub>* and *harm<sub>B</sub>* is the harm prevented by the throw.

For example, suppose that if neither the throw nor the roll had occurred, and hence all of the rocks had collided with Bea, then Bea would have suffered one broken leg and one broken arm; and that if the roll had not occurred, and hence only the roll-deflected rocks had collided

<sup>116</sup> I assume the harm that would result if both sets of rocks were to collide with Bea is *equal* to the sum of the harm that would result if only the one set were to collide with Bea and the harm that would result if only the other set were to do so. Without this assumption, the harm prevented by the throw could still be specified by counterfactuals, but we would first have to answer questions such as the following: If the conjunction of two acts prevents a harm, but neither act would have done so without the other, then what should we say about *each* act? Should we say that *act<sub>1</sub>* prevents the harm *and* that *act<sub>2</sub>* prevents the harm? Or that neither act prevents the harm (though their conjunction does)?

<sup>117</sup> This is true *given* our “simplifying assumption” (and the assumption that no one will intervene if you do not).

<sup>118</sup> See the previous note. Also note that if the roll had not occurred, the throw would still have occurred—in accordance with the stipulation in note 98 about how to interpret our counterfactual conditionals.

with Bea, then Bea would have suffered only the broken arm.<sup>119</sup> Suppose, in other words, that  $\text{harm}_A$  consists in the broken leg and the broken arm, and that  $\text{harm}_B$  is just the broken arm. The difference between  $\text{harm}_A$  and  $\text{harm}_B$  is the broken leg. According to what was said in the previous paragraph, this is the harm prevented by the throw. And this appears to be the correct answer. For the rocks were initially on course to break both an arm and a leg, and after the throw only the arm was in danger. The throw deflected the rocks that would have broken Bea's leg, but not the rocks that would have broken her arm.

In the cases just discussed, the harm prevented by the relevant act *does* correspond to the harm that would have occurred, had certain conditions been met. (Or, in the most recent case, the harm prevented corresponds to the *difference* between the harm that would have occurred, had one set of conditions been met, and the harm that would have occurred, had a second set of conditions been met.) Even if we have no account that applies to *all* cases, in any *particular* case the harm prevented is tied to some counterfactual or other.

Furthermore, we can often *tell* which counterfactuals are relevant in a given case. In a variation on the case of Abe and Bea, suppose that we know the following: (i) the throw deflects *all* the rocks, and (ii) you *could* have simply pushed Bea out of the rocks' path. Then we know better than to assume that the throw prevents whatever harm would have occurred, had the throw not occurred. For we know that if the throw prevents any harm at all, then that harm probably would have been prevented by a push, had the throw not occurred. More generally, we can tell that a counterfactual is not a good choice if, *for all we know*, the harm prevented by the relevant act would have been prevented in some other way, had the counterfactual's antecedent been true. We can also tell that a counterfactual is not a good choice if, for all we know, the agent's *other*

---

<sup>119</sup> As always, Bea is unharmed in actuality.

acts prevent harm that would *not* have been prevented, had the counterfactual's antecedent been true. Hence if you throw one ball at the rocks *and* roll a second ball into their path, we know better than to focus on what would have happened, had you not acted at all. For we know that if the throw prevents harm, then the roll likely does so as well (and that we should hold fixed the effect of the roll in order to identify the effect of the throw).

On the other hand, a counterfactual *is* a good choice if it does not raise the worries just noted. As mentioned, if the *only* thing you do is throw the first bowling ball, then we *can* ask what would have happened, had you done nothing at all. We know that the harm prevented by the throw, if any, would *not* have been prevented by another act of yours had you done nothing at all. And we know that there are no *other* acts of yours that prevent harm, and hence that no such harm would have occurred, had you done nothing.

We can know these things, moreover, without yet knowing *what harm* is prevented by the throw. This helps to answer a concern raised in the previous section. Recall the claim that our arbitrary ACT prevents whatever harm does not actually occur but would have, had ACT-prevented harm not been prevented.<sup>120</sup> This claim may appear to be unhelpful because of its circularity. Determining what would have happened, had ACT-prevented harm not been prevented, requires knowing *which acts* would have prevented such harm; and this may seem to require knowing *what harm* this is—the very thing we had hoped to learn. In many cases, however, we can know which acts would have prevented ACT-prevented harm *without* yet knowing what harm this is. If the throw prevents any harm at all, then pushing Bea out of the rocks' path would likely have prevented the same harm had the throw not occurred. We can know this, without knowing what harm was actually prevented by the throw. We can likewise

---

<sup>120</sup> ...and had the agent prevented all of the *other* harm that they prevent in actuality.

know that chewing gum, checking your cell phone, or walking away would probably *not* have prevented any throw-prevented harm.<sup>121</sup>

At this point we can state a few intermediate conclusions. In particular, section 1 has been intended to establish the following claims:

1. The harm prevented by an act corresponds to the difference between the harm that actually occurs and the harm that would have occurred, had certain conditions been met.
2. A non-circular and fully general statement of the relevant conditions is not forthcoming.
3. Nonetheless, we can often tell which conditions are relevant in a given case. In particular, we can often identify conditions under which the act-prevented harm would not have been prevented.

The following discussion will rely on each of these claims to one degree or another. But before moving on, a few potential objections should be answered.

#### *1.4 Objections*

The first objection comes from Jeff McMahan and Robert McKim (1993).<sup>122</sup> They write:

[O]ne should consider whether counterfactual speculation is really necessary in order to determine what effects may be attributed to war. We believe that it is not. In order for it to be true that an agent's action prevents an evil from occurring, it is sufficient that there be a causal sequence that will result in the evil if it continues uninterrupted and that the agent intervenes in the sequence so that it will not result in the evil. It need not be the case that the causal sequence would actually result in the occurrence of the evil if this particular intervention did not occur[.] (509)

---

<sup>121</sup> It is *possible* that walking away would have prevented the same harm as the throw actually prevents. (Perhaps you would have stepped on a switch that would have activated some mechanism.) It is also possible that pushing Bea out of the rocks' path would *not* have prevented any throw-prevented harm. (Perhaps there is *no* throw-prevented harm, as the rocks would have stopped short of Bea regardless.) But certainty is not required to make an informed judgment about which counterfactuals to use, and hence to make an informed judgment about the harm prevented by the throw.

<sup>122</sup> A very similar objection was offered by Alec Walen (personal communication).

I agree with the last sentence here. The relevant question is not—or not *always*—what would happen if this particular intervention did not occur. But I disagree with the conclusion that McMahan and McKim draw from this fact. For it does not follow that “counterfactual speculation is [not] really necessary.” Rather, what follows is that the particular counterfactual in question is inadequate. Consider the example that McMahan and McKim offer:

[A] child is in the path of a speeding truck. A passerby snatches the child out of the way. It is not necessary to know what would have happened had the passerby not acted in order to know that she saved the child. Indeed, suppose that someone else would have snatched the child out of the way if the passerby had not. It remains true that the passerby saved the child—that her act prevented its death—even though it is not true that the child would have been killed if she had not acted. (*ibid.*)

The passerby does save the child, of course. But that would not be true unless there were conditions under which the child *would have* died. McMahan and McKim acknowledge as much in the first passage quoted above, where they speak of “a causal sequence that will result in the evil if it continues uninterrupted.” This seems to be interchangeable with speaking of a causal sequence that *would* result in the evil if it *were* continued uninterrupted. And this is a counterfactual conditional.<sup>123</sup> Hence the objection from McMahan and McKim seems to confirm rather than refute the counterfactual nature of harm prevention.<sup>124</sup>

---

<sup>123</sup> It is, at least, a *subjunctive* conditional. For our purposes, nothing substantive turns on whether or not the antecedent is truly contrary to fact. What matters is simply that the harm that would be prevented by an act corresponds to the difference between the harm that would occur were one set of conditions met and the harm that would occur were another set of conditions met.

<sup>124</sup> Have we at least found a *general* account of the harm prevented by an act? Can we say that an act prevents whatever harm would have occurred, had the relevant causal sequence not been interrupted? No, as that would often overcount the harm prevented by the act in question. Suppose that  $\text{act}_1$  and  $\text{act}_2$  each (partially) interrupt some threatening causal sequence, independently reducing the harm it would cause. The harm prevented by  $\text{act}_1$  does not correspond to the harm that would have occurred, had the sequence not been interrupted. For that harm includes both the harm prevented by  $\text{act}_1$  and the harm prevented by  $\text{act}_2$ . There will also be serious problems regarding what *counts* as interrupting a sequence—as illustrated in section 2.2.

The next objection claims that I have assumed a counterfactual account of harm. Take the first “intermediate conclusion” offered at the end of section 1.3, for instance.<sup>125</sup> This conclusion looks very similar to the claim that the harm caused by an event corresponds to the difference between one’s actual welfare and what one’s welfare would be, had the event not occurred (and had certain other conditions been met). Because of this similarity, one might think that I have assumed a counterfactual account of harm. Such an assumption would be objectionable to some, including those who endorse a noncomparative account of harm.<sup>126</sup>

I have not, however, assumed a counterfactual account of harm. In particular, I leave open the possibility that harm is noncomparative in nature. Deflecting all the rocks prevents the harm that would have resulted, had you not intervened. This is compatible with it being true that had you not intervened, the resulting state would have been one of *harm* in virtue of considerations unrelated to counterfactuals. For instance, if the rocks collide with Bea, perhaps the resulting pain is a harm because it is a *bad* state to be in, not because it is a *worse* state than some alternative. Yet even if harm is not comparative, prevention is. If a harm is *prevented*, then it does not actually occur, and so we cannot identify it by looking to the *actual* state of the victim, and to how bad it is to be in such a state. We must consider the state that the victim *would have* been in, had the harm not been prevented. If the pain resulting from a collision would have been severe, then deflecting the rocks prevents a lot of harm. If instead the pain would have been moderate, then deflecting the rocks prevents less harm. And if all the rocks

<sup>125</sup> That is: “The harm prevented by an act corresponds to the difference between the harm that actually occurs and the harm that would have occurred, had certain conditions been met.”

<sup>126</sup> According to such an account: To harm someone is not to make them worse off than they *would be*, nor is it to make them worse off than they *were*. Rather, to harm someone is to cause them to be in a state that is bad *in itself*, independently of any comparison to a prior or counterfactual state. See Shiffrin 1999, 2012; Harman 2004, 2009.

would have missed Bea, then deflecting the rocks prevents no harm. These are all counterfactual conditionals, and they do not assume that pain (or any other harm) is comparative in nature.

For similar reasons, I have not assumed a counterfactual account of causation. Again, deflecting all of the rocks prevents the harm that would have resulted, had you not intervened. This is compatible with it being true that had you not intervened, the rocks would have *caused* the resulting harm in virtue of considerations unrelated to counterfactuals. If the rocks collide with Bea, perhaps the causation here is wholly a matter of what *actually* occurs—the interlocking chain of spatiotemporally contiguous events, extending from Abe's dislodging of the rocks to the rocks' colliding with Bea; the energy that is transferred to Bea on impact; and the bodily state that results. But while causation while that may be true for some cases of causing harm, it is not true for our cases of preventing harm. When you deflect the rocks that Abe sends down the hill toward Bea, the chain of spatiotemporally contiguous events leads from you to the rocks, by way of the bowling balls. This transfers energy to the rocks, changing their physical state. What does this have to do with Bea? Only that the rocks *would have* collided with her, had you not intervened.

At the very least, it seems clear that we *rely* on counterfactual conditionals when making judgments about the harm prevented by an act. For simplicity, suppose that the *only* thing you do is throw the one bowling ball, and that this deflects all the rocks. Perhaps we are given information about the size and trajectory of the rocks, the constitution of Bea, and the relevant physical laws. This information will not tell us what harm the throw prevents, unless we can *use* it to determine what would have happened, had the rocks collided with Bea. By contrast, if medical experts tell us that Bea would have been killed, then we know immediately that the

throw prevents Bea's death. We know what harm is prevented when we learn what harm the rocks *would have* caused, and not before that point.<sup>127</sup>

## 2. Problems of Application

### 2.1 Operation Protective Edge

Hamas fires rockets into Israel. Israel bombs and then invades the Gaza Strip, aiming to prevent further rocket fire. The Israeli military operation is known as Operation Protective Edge (OPE). What harm does OPE prevent?<sup>128</sup> Given what has been said above, OPE prevents whatever harm does not *actually* occur but *would have*, had certain conditions been met. We have seen some of the difficulties involved in finding a non-circular and fully general statement of these conditions. But we have also seen that such a statement is not needed. In any particular case, we can typically identify conditions that lead to the right conclusion. The question, then, is whether we can identify such conditions in the case of OPE.

The relevant comparison is not the harm that would have occurred, had OPE not occurred. For had OPE not occurred, Israel would have used other means to prevent some of the OPE-prevented harm.<sup>129</sup> These means might have included alternative military operations, civil

<sup>127</sup> Perhaps what would have happened is *determined*, in one sense or another, by what actually happens, or by what actually happens in conjunction with physical laws. That is compatible with the claim that the harm prevented by an act corresponds to the difference between the harm that actually occurs and the harm that would have occurred, had certain conditions been met. It is likewise compatible with the claim that we rely on counterfactuals to identify the harm that an act prevents.

<sup>128</sup> OPE destroyed some of Hamas's rockets, among other resources, so presumably there were harms that the operation prevented. But I make no assumptions about whether OPE resulted in a decrease in the *net* amount of harm done by Hamas (or by all parties taken together).

<sup>129</sup> Following the convention from section 1, say that *OPE-prevented harm* is any harm prevented by OPE in actuality.

defense measures, or diplomacy. We will undercount the harm prevented by OPE if we compare the actual world to one in which some OPE-prevented harm is prevented by other means.<sup>130</sup>

Nor is the relevant comparison the harm that would have occurred, had Israel refrained from acting. Perhaps we suppose it is the entire Israeli government that refrains from acting, or perhaps we suppose it is just the military that does so. Either way, the following problem arises. If Israel refrains from all action, then Israel prevents no harm at all during the relevant period of time (whatever period that may be). As a result, the proposal overcounts the harm prevented by OPE. Suppose that OPE prevents harm  $h_1$  and other actions by Israel prevent harm  $h_2$ . If we compare the actual world to one in which Israel refrains from all action, then we compare the actual world to one in which both  $h_1$  and  $h_2$  occur, and so our account will credit OPE with preventing both  $h_1$  and  $h_2$ .<sup>131</sup> Perhaps  $h_2$  is unrelated to the threat posed by Hamas's rockets. (Perhaps it is a traffic accident prevented by the Transport Ministry.) In that case, we could avoid overcounting the harm prevented by OPE by considering only the harm that *Hamas's rockets* would have done, had Israel refrained from acting. But this solution will not work more generally, as  $h_2$  could be harm that Hamas's rockets would have done. (Perhaps it is a casualty prevented by Israel's civil defense system, which includes early warning and missile defense components.) We wish to know the harm prevented by Israel's bombardment and invasion of the Gaza Strip. Because these were not the only defensive measures taken, we cannot simply consider the harm that would have occurred had Israel refrained from acting.<sup>132</sup>

---

<sup>130</sup> The same problem is faced by the proposal that we compare the actual world to one in which OPE does not occur and Israel otherwise “goes on as before.” For Israel, going on as before would include responding to rocket attacks in some way.

<sup>131</sup> ...unless there is a backup preventer that would have prevented  $h_2$  had Israel not acted.

<sup>132</sup> This problem would equally affect the following proposal: OPE prevents whatever harm does not occur but would have, had Israel done nothing to Hamas, Hamas's rockets, or the civilians at risk. Likewise, the problem posed for the next proposal would equally affect the following one: OPE prevents

Next consider the harm that would have occurred, had OPE not occurred and had Israel done nothing other than what it does in actuality. This is still not the harm prevented specifically by OPE. For had OPE not occurred, other events that *occur* in actuality would have prevented *more harm* than they do in actuality. In particular, some of these events would have prevented OPE-prevented harm. (Had OPE not occurred, for example, some of the rockets actually destroyed by OPE would have been fired, increasing the amount of harm prevented by the activation of Israel's warning sirens during rocket attacks.) So if OPE had not occurred and Israel had done nothing other than what it does in actuality, then not all of the OPE-prevented harm would have occurred. Hence the current proposal undercounts the harm prevented by OPE.<sup>133</sup>

We are looking for conditions that will lead us to the right conclusion about the harm prevented by OPE. We have just considered the three sets of conditions discussed in section 1: (1) the act does not occur; (2) the act does not occur, and the agent does nothing; (3) the act does not occur, and the agent does nothing other than what they do in actuality. Each of these sets of conditions will lead to the right conclusion in some cases, though not in all. None will lead to the correct conclusion in the case of OPE, as we have just seen.

What we *need* to do is compare the actual world to one in which all of the OPE-prevented harm occurs. Otherwise we will undercount the harm prevented by OPE. Yet we cannot simply look to a world in which Israel prevents no harm at all. Indeed, we must look to a world in which Israel prevents *all* of the harm that it prevents, in actuality, by means other than OPE. Otherwise

whatever harm does not occur but would have, had Israel done nothing to Hamas, Hamas's rockets, or the civilians at risk other than it does in actuality.

<sup>133</sup> One might object that the following are incompatible: (i) Israel's warning sirens preventing more harm than they do in actuality; (ii) Israel's doing nothing other than what it does in actuality. But this objection will lead to a problem similar to that posed for the analogous objection considered at the end of section 1.1

we will overcount the harm prevented by OPE. Hence we must identify the specific actions that would prevent OPE-prevented harm, were OPE not to occur. And we must consider a world in which no such actions are performed (or in which they *are* performed but prevent no more harm than they do in actuality). In particular, we must consider a world in which the following is true: OPE does not occur; Israel does not use alternative military operations, civil defense measures, or diplomacy to prevent OPE-prevented harm; and, more generally, Israel does nothing else that prevents the harm that is prevented, in actuality, by OPE. Contrary to the circularity worry expressed in section 1.2, the problem is not that this requires us to already know what harm *is* prevented by OPE. Rather, the problem is that it seems to require us to know what *counts* as preventing harm. We consider this problem next.

## 2.2 What Counts as Preventing Harm?

Suppose that it is summer 2014, just before the start of OPE.<sup>134</sup> One of Hamas's goals is to secure the release of certain prisoners held by Israel. Assume that Hamas will fire fewer rockets and cause less harm if these prisoners are released than if they are not.<sup>135</sup> Does this mean that releasing the prisoners would prevent harm?

The case for saying yes is obvious. After all, I have *stipulated* that Hamas will cause less harm if the prisoners are released. And this would not be a coincidence. If the prisoners are released, Hamas will cause less harm *because* they are released. These facts have corollaries, however: Hamas will cause *more* harm if the prisoners are not released, and it will do so *because* the prisoners are not released. If the first pair of facts suggests that releasing the prisoners would

<sup>134</sup> This assumption will simplify the conditionals in this section.

<sup>135</sup> Other things being equal. In particular, this means that Hamas will cause less harm if the prisoners are released and OPE occurs than if the prisoners are not released and OPE occurs. (And likewise, replacing both instances of "occurs" with "does not occur.")

*prevent* harm, then the second pair suggests that continuing to hold the prisoners would *cause* harm. In that case, releasing the prisoners would be to *refrain* from causing harm. And to refrain from causing harm is not sufficient for preventing harm. So the initial case for saying yes to our question is counterbalanced by an analogous case for saying no.

Recall the example of Abe and Bea. Suppose that if you taunt Abe, then he will roll even more rocks down the hill toward Bea. Refraining from taunting Abe is not to prevent him from harming Bea. It is merely to refrain from causing Abe to harm Bea. The same is true in a case in which you begin taunting Abe but stop quickly enough that he does not dislodge any more rocks. By stopping, you do not prevent Abe from harming Bea. Rather, you abandon a course of action that would have caused Abe to harm Bea. In a similar manner, perhaps releasing the Palestinian prisoners would be merely to refrain from causing harm. For perhaps releasing the prisoners would be to cease a course of action that would have caused harm if continued.<sup>136</sup>

One might be tempted to deny that there is any real difference between preventing harm and refraining from causing harm. There are at least two ways that such a denial could be motivated. First, one might think that to prevent harm is to cause harm not to occur, and that causes come cheap. In particular, one might think that if an agent refrains from causing harm to occur, then they cause harm not to occur, and hence they prevent harm.<sup>137</sup> Second, one might think that we are merely probing linguistic intuitions about which description is more apt for a

---

<sup>136</sup> More precisely, perhaps releasing the prisoners would be to cease a course of action that would have caused *Hamas* to cause harm if continued. I often leave such qualifications implicit in what follows.

<sup>137</sup> The contrapositive is occasionally endorsed, typically when discussing harm caused (because not prevented) by the Queen. Illari and Williamson, for instance, make the following claim in passing: “Just as my failure to water your plants caused their death, so the Queen’s failure to water them caused their death, and so on” (2011: 832).

given case. (“Prevent harm” or “refrain from causing harm?”) And one might think that such intuitions are of little moral relevance.

There is nonetheless a distinction here that is worth drawing. If Israel’s releasing the prisoners is to prevent harm, then continuing to hold the prisoners is to refrain from preventing harm; in other words, it is to *allow* harm. On the other hand, if releasing the prisoners is to refrain from causing harm, then continuing to hold the prisoners is to cause harm; in other words, it is to *do* harm. So if there is no difference between preventing harm and refraining from causing harm, then there is no difference between doing and allowing harm.<sup>138</sup>

The example of Israel’s releasing the prisoners is not unique. To continue with the present case, one of Hamas’s primary goals is to end Israel’s blockade of Gaza. Suppose that Hamas will fire fewer rockets and cause less harm if the blockade is ended than if it is not. Does this mean that Israel’s ending the blockade would prevent Hamas from causing harm? Or would it be merely to refrain from causing Hamas to cause harm? Or suppose that Hamas will cause less harm if Israel scales back settlement in the Palestinian territories, or releases tax funds that had been collected on behalf of the Palestinian Authority but subsequently withheld. The same type of questions arise in these cases as well.

Each of these examples involves a course of action that extends over time, with the agent having the option to either continue or change course. Israel has the option of *continuing* to imprison certain individuals who were previously arrested, for example. This is not an essential feature of the case, however. For we could just as well ask questions such as the following: If

---

<sup>138</sup> The notion of causation relevant to moral evaluation is often narrower than that relevant in other contexts, such as in metaphysics or science. There is undoubtedly a sense in which withdrawing life-sustaining treatment is to cause death, for example. But that does not settle whether doing so is to cause death in the sense relevant to the doing/allowing distinction, as opposed to merely refraining from preventing death. See McMahan 1993 (277–278) and Bronner 2018 (§2).

Israel goes on to make *new* arrests, and this results in more rockets being fired than otherwise would be, is this to cause Hamas to cause harm? Or is this to refrain from preventing Hamas from causing harm, insofar as declining to make such arrests would be to prevent Hamas from causing harm?

One might think that the answer depends on human psychology. Perhaps if Hamas experiences further arrests as a *provocation*, then Israel's carrying out such arrests is to cause harm (and refraining is neither to cause nor to prevent harm). On the other hand, if Hamas experiences the absence of further arrests as a *concession*, then perhaps Israel's refraining from such arrests is to prevent harm (and carrying them out is neither to cause nor to prevent harm). This proposal leaves us with a number of unanswered questions, however (beyond familiar questions about the ascription of attitudes to *groups*). What if Hamas were replaced with a computer that fires a certain number of rockets depending on the number of persons who are arrested, without experiencing anything *as* a provocation or a concession? Or what if Hamas experiences further arrests as a provocation *and* the absence of further arrests as a concession? Should we then say that carrying out further arrests is to cause harm *and* that refraining from such arrests is to prevent harm? If so, this would be a case in which refraining from causing harm is sufficient for preventing harm, and in which declining to prevent harm is sufficient for causing harm. This would be an odd result. It is typically possible for an agent to take a course of action that neither causes nor prevents harm.<sup>139</sup>

---

<sup>139</sup> The exceptions to this generalization have features lacking in the present case. This is not, for instance, a version of the case in which Israel has *already* arrested certain individuals and now has the choice of either releasing or continuing to hold them. In *that* version of the case, one might think that the initial arrest will cause Hamas to cause harm unless Israel prevents this by releasing the prisoners.

One might suggest that the key issue is whether it is *permissible* for Israel to imprison the individuals in question, blockade Gaza, etc., or whether Israel has a *right* to engage in such conduct. But consider the following. We wish to know what harm is prevented by OPE—or, at least, to know how we might go about determining what harm is prevented by OPE. In section 1, it was argued that the harm prevented by an act corresponds to the difference between the harm that occurs in the actual world and the harm that occurs in a possible world where none of the act-prevented harm is prevented. Suppose that if OPE does not occur and Israel does not use alternative military operations to achieve the same ends, then Israel will release some of the prisoners in question. Would such a world be one in which OPE-prevented harm is prevented? Answering this question requires that we know whether releasing the prisoners would be to prevent harm at all. And whether releasing the prisoners would be to prevent harm does not depend on whether it is permissible to continue holding the prisoners, or whether Israel has a right to do so. Indeed, essentially the opposite is true: whether it is permissible to continue holding the prisoners, or whether Israel has a right to do so, depends on what harm is thereby caused or prevented (among other factors).

### *2.3 Counterfactuals and Proportionality*

Our problem is similar to one that has been discussed by others, and one might think that responses to the latter could resolve the former. This section argues against such a resolution.

Discussions of war and self-defense typically assume that the harm one does must be *proportionate* to the good one thereby achieves. Roughly speaking, harm is proportionate if it is

not excessive in comparison to the good achieved. Killing an attacker in defense of one's life is often proportionate, for instance, while killing a pickpocket is typically not.<sup>140</sup>

In order to identify the harm caused or the good achieved, it seems we must compare the outcome of war or defense with a possible world in which the relevant conduct does not occur. Yet "it is not obvious *which* counterfactual situation we are to consider in making our comparison" (Kavka 1991: 24). Possibilities include "the one resulting from the nation in question doing nothing, from its doing what it would most likely have done if it did not fight, or from its doing what ... would have maximized utility" (*ibid.*). Several authors have discussed the question of which counterfactual (if any) should be used.

First consider McMahan (2013/2014), who takes up the issue in the course of discussing opportunity costs. To use one of McMahan's examples, suppose that State A could save 10,000 lives by conducting a war of humanitarian intervention, but doing so would cost enough money that State A would then be unable to pursue a disease eradication program that would save 100,000 lives. Fighting the war comes with the opportunity cost of foregoing the saving of 100,000 people. Yet McMahan does not think that these deaths should count against the proportionality of war. He writes:

Proportionality, it seems, takes account only of good and bad effects that defensive action *causes*. It does not take account either of those effects it allows to occur, because engaging in the defensive action prevents the agent from being able to prevent them, or of those effects it allows not to occur because the defensive action prevents the agent from being able to cause them. Proportionality, in other words, does not require a comparison between entire possible worlds. It requires only a comparison between the relevant bad effects that defensive action (including war) would cause, either directly or indirectly, and the relevant good effects it would cause—in particular, the prevention of harms that would otherwise be caused by others. What a person or state would or could do if he or it

---

<sup>140</sup> As typically interpreted, proportionality is more than a matter of whether the harm inflicted is greater than the harm averted. Various deontological factors are relevant as well. See especially Hurka 2005 and McMahan 2013/2014.

were not to engage in some defensive action is irrelevant to the determination of whether the defensive action is proportionate. (4–5)

This all sounds plausible, but it does not address our problem. For if what is relevant is “the prevention of harms that would otherwise be caused by others,” then we need to know *what harms* those would be. It is not clear how to identify those harms, at least in the case that we have been discussing. McMahan is not considering such cases.

Next consider Thomas Hurka (2008), who, like McMahan, wishes to exclude consideration of opportunity costs from the evaluation of proportionality. Hurka’s suggestion is that “we should compare the net effect of war with that of the least beneficial alternative that is morally permitted” (130). This means, for example, that we should not compare a war of humanitarian intervention with an alternative in which a supererogatory disease eradication program is pursued—and hence that war is not disproportionate simply because it precludes such a program. Yet which alternatives *are* morally permitted will depend (in part) on the amount of harm that each alternative prevents. Which of the permissible alternatives is the *least beneficial* will also depend (in part) on the amount of harm that each alternative prevents. So, following Hurka’s proposal requires that we already have a solution to our problem in hand.

Finally, consider David Mellow’s (2006) account. He puts forward two possibilities:

The appropriate counterfactual baseline is the nonwar course of action that the leaders would undertake if they did not engage in actions that immorally furthered the harms or injustices that underlie the [just cause for war]. (444)

The appropriate counterfactual baseline is that course of action the leaders would undertake if they neither engaged in action that immorally furthered the harms or injustices that underlie the [just cause] nor undertook action that was directed at achieving the [just cause]. (445)

Mellow is “agnostic” as to which of these baselines is the correct one (446). In order to make use of the former, we must be able to tell *which* actions would immorally further the harms or

injustices that underlie the just cause for war. Furthering such harms or injustices, whether immorally or not, would involve causing them (or causing another agent to cause them). Hence in order to use Mellow's first counterfactual baseline, we must be able to distinguish between causing harm and merely refraining from preventing harm.<sup>141</sup> Using Mellow's second baseline requires that *and* the ability to tell which actions would be "directed at achieving" the just cause—which actions, that is, would "go some ways toward achieving" the just cause (444). Such actions would involve preventing harms or injustices. Hence Mellow's second baseline requires us to distinguish between preventing harm and merely refraining from causing harm. So Mellow's proposal, like Hurka's, requires that we already have a solution to our problem.

#### *2.4 Doing and Allowing*

Two distinctions were just mentioned when discussing Mellow's account. First is the distinction between causing harm and merely refraining from preventing harm. Second is the distinction between preventing harm and merely refraining from causing harm. These are two sides of the same coin. If we focus on Israel's *releasing* the relevant prisoners, for instance, then the question is whether this would be to prevent harm or merely to refrain from causing harm. If we instead focus on Israel's continuing to *hold* the prisoners, then the question is whether this would be to cause harm or merely to refrain from preventing harm.

To refrain from preventing harm is to allow harm, and to cause harm is to do harm.<sup>142</sup> So we might try to resolve our questions by applying an account of the doing/allowing distinction.

<sup>141</sup> And likewise for injustice. Distinguishing between causing injustice and merely refraining from preventing injustice will be no easier than making the parallel distinction regarding harm. After all, the harms under discussion typically *are* injustices.

<sup>142</sup> Given the relevant notions of prevention and causation; see note 138. More precisely, to refrain from preventing harm is *no more* than to allow harm. It is not *even* to allow harm if the threatened harm does not eventuate. I mostly leave this qualification implicit.

The best account of this distinction is some variation on Philippa Foot's (1967, 1984, 1985).<sup>143</sup>

According to Foot,

The difference we should have in mind is roughly this: in the case of *allowing*, a train of events must already have started or be on the horizon; an agent who could stop or prevent it does not do so, and therefore allows it to go on. In the other case it is he who *initiates* a sequence, as for instance by shooting someone who dies of a gunshot wound, or by pushing into the water someone who then drowns. (1985: 24)

Foot adds an important note: "In some cases the agent will not start a sequence but rather keep it going when it would otherwise have come to an end. Then he initiates a new stage of the sequence rather than the sequence itself" (37). So an agent can do harm either by setting a threatening sequence in motion or by sustaining an existing one. On the other hand, an agent allows harm by refraining from curtailing a threatening sequence.

Such an account leads to questions like the following: Is continuing to hold the prisoners simply to refrain from curtailing a threatening sequence of events? Or is it instead to keep such a sequence going? I find these questions no easier than those we started with. We might view the continued holding of the prisoners as sustaining or amplifying a sequence that leads from Hamas to the injury or killing of certain individuals. We might instead see the release of the prisoners as intervening to curtail such a sequence, and hence the continued holding of the prisoners as merely refraining from such intervention. Both views are plausible, and it is not clear which is correct.

We might be tempted to say something like the following: if the conduct in question is essentially an *act*, then it is to *intervene* in the relevant sequence of events; while if it is more like an *omission*, then it is merely to *refrain* from intervening. In the case of the previously withheld

---

<sup>143</sup> See, for example, how McMahan (1993) uses what is essentially a modified version of Foot's account to accommodate cases that other accounts struggle with. See also the use of such an account by Kamm (1996) and Woollard (2008, 2013).

tax funds, perhaps there is a bank transfer that requires approval, in which case the funds' release might be seen as an act. Or perhaps there is a bank transfer that will automatically execute unless a temporary hold is placed each week, in which case the release might be seen as an omission.

The current proposal identifies acting with doing, however, and omitting with allowing.<sup>144</sup> It entails, for instance, that if continuing to withhold the funds is an act, then it is to cause harm; while if it is an omission, then it is merely to refrain from preventing harm—that is, to allow harm. And matters are not nearly as straightforward as such an identification would suggest. The upshot of an *act* can be that one *allows* something to happen, and an *omission* can be sufficient for *doing* something. In particular, one can prevent harm by omitting an act, as in a case where you refrain from moving so that an oncoming rock will hit you rather than someone further down the hill (Howard-Snyder 2011: §4). Perhaps releasing the tax funds by omitting to prevent their release would likewise be to prevent harm. More generally, we cannot assume that there is a straightforward correspondence between acting and doing, and between omitting and allowing. Indeed, our sense of uncertainty about the case—*my* sense of uncertainty, at least—is not dispelled by the assumption that a bank transfer requires approval, nor by the assumption that a transfer automatically executes unless a hold is placed. Would releasing the funds prevent Hamas from causing harm? Or would it be merely to refrain from causing Hamas to cause harm? I still see no clear answer to such questions.<sup>145</sup>

---

<sup>144</sup> The proposal at the beginning of this paragraph, that is—not Foot's proposal.

<sup>145</sup> Foot herself is optimistic about our ability to distinguish between doing and allowing harm: “If there are difficulties when it comes to detail, some of these ideas may be best left unsolved, for there may be an area of indefiniteness reflecting the uncertainty that belongs to our moral judgments in some complex and perhaps infrequently encountered situations” (1984: 178). Yet the sort of case now under discussion is not infrequently encountered, and for that reason it is not unusually complex either.

One might think that the key issue is the ownership of the relevant resources. For example, releasing previously withheld medicine to an ailing individual is to provide aid and hence to prevent harm if the medicine is yours; but it is merely to refrain from further harming the individual if the medicine is theirs. The present case is not like this, however. The relevance of the withheld tax funds is not that they would be *used* to prevent harm if they were released (perhaps by purchasing medicine). Rather, the tax funds are relevant to our discussion simply because we have assumed that Hamas will fire fewer rockets if they are released. The question of who owns the funds is not obviously relevant.<sup>146</sup> In any case, not all of our examples involve the release or withholding of resources in any familiar sense. Consider the decision to refrain from arrests that would otherwise be made, or to release individuals who were arrested previously. These individuals are not *resources* analogous to money or medical supplies. (Indeed, we can even suppose that the individuals are not *useful* to Hamas at all.)

### *2.5 Forgetting About Prevention*

At this point, one might suggest that we forget about what harm an act *prevents* and simply consider what harm each party *causes* in various scenarios. If OPE occurs, then Israel will kill  $n$  civilians and Hamas will kill  $m$  civilians; if OPE does not occur, then Israel will kill  $n'$  civilians and Hamas will kill  $m'$  civilians; and so forth. More is relevant than simply the number of civilians killed by each party, of course. And each option available to Israel will have to be spelled out much more fully; Israel has more than two options. But perhaps information of this *sort* is sufficient for deciding whether OPE is permissible.<sup>147</sup>

---

<sup>146</sup> Suppose that A refuses to pay a debt which B claims is owed to them, and that this provokes either B or a third party to violence. Must we know whether the debt was genuine before we can say whether A has caused the harm or has merely refrained from preventing it?

<sup>147</sup> OPE is temporally extended and contains many individual actions. I intend to remain neutral as to whether we can (or should) evaluate the permissibility of OPE *as a whole*, as opposed to merely

Adopting this proposal would leave us unable to draw certain distinctions, however.

Those who take these distinctions to be morally relevant have reason to reject the proposal.

Three examples follow.

First is the distinction between doing and allowing harm. If releasing the relevant prisoners is to prevent harm, then continuing to hold them is merely to refrain from preventing harm; in other words, it is merely to allow harm. On the other hand, if releasing the prisoners is merely to refrain from causing harm, then continuing to hold them is to cause harm; in other words, it is to do harm.<sup>148</sup> So if we cannot tell whether releasing the prisoners is to prevent harm or is instead merely to refrain from causing harm, then we cannot tell whether continuing to hold the prisoners is merely to allow harm or is instead to do harm. And likewise for Israel's ending the blockade of Gaza, scaling back settlement in the Palestinian territories, releasing previously withheld tax funds, and any other conduct whose causal/preventive status is unresolved.<sup>149</sup>

The second distinction is between causing harm that would occur anyway, were you not to prevent it, and causing harm in other cases. Consider Judith Jarvis Thomson's case of a violinist with kidney failure. The Society of Music Lovers has "kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract the poisons from his blood as well as your own" (1971: 49). Thomson writes that unplugging

evaluating *discrete actions* such as decisions to authorize, initiate, or continue OPE. The difficulties now under discussion would arise either way.

<sup>148</sup> It is not *generally* true that if performing some act is merely to refrain from causing harm—that is, to refrain from causing harm *and* not to prevent harm—then refraining from the act is to cause harm. Releasing the relevant prisoners is special insofar as it is counterfactually related to harm in the manner stipulated in section 2.1.

<sup>149</sup> This also means that the proposal in question is self-undermining. The proposal is that we forget about what harm an act *prevents* and simply consider what harm each party *causes* in various scenarios. But as we have just seen, if we do not know whether some conduct would prevent harm, then we do not know whether refraining from that conduct would cause harm. (This is true, at least, if we understand 'cause harm' to mean 'do harm' rather than something broader. See note 138.)

yourself from the violinist would be to kill him (57). This claim may be mistaken, for it may be that withdrawing life-sustaining aid is a case of letting die rather than killing.<sup>150</sup> Nonetheless, we can imagine a variation of the case in which you *do* have good reason to kill the violinist. Suppose that simply unplugging yourself would somehow damage your health, and that in order to avoid this result, you must administer a lethal injection to the violinist and unplug yourself only after he is dead. If you do not kill the violinist, he will die unless your body continues to prevent his death. So if you *do* kill the violinist, you only deprive him of life that he would have had with your continued aid. This fact may lessen the objection to killing him.<sup>151</sup> If we give up on determining which harms are prevented by which acts, then we will be unable to identify cases like this—cases in which the harm caused by an agent would occur anyway, were the agent not to prevent it.

Suppose, for example, that Israel has three options:

*War.* OPE occurs and the Israeli bombardment of Gaza kills a civilian, C.

*Diplomacy.* OPE does not occur, Israel releases the prisoners specified by Hamas, and C survives.

*Status Quo.* OPE does not occur, Israel continues to hold the prisoners, and one of Hamas's rockets kills C.<sup>152</sup>

---

<sup>150</sup> Kuhse 1987: 49–50; McMahan 1993; Kamm 1996: 28–30; Jensen 2011; McGee 2014, 2015; Bronner 2018. The variation of the case considered next is similar to one discussed by Kamm (1996: 27), as well as one mentioned by McMahan (2002: 379).

<sup>151</sup> Here I follow Kamm 1996 (27, 29, 37, 53, among other places).

<sup>152</sup> There were civilians killed by Hamas's rockets. For example, Amnesty International “concluded that a Palestinian rocket that fell short was the cause of an explosion in the Shati refugee camp in Gaza in which 13 Palestinian civilians were killed” (Isabel Kershner, “Amnesty International sees evidence of Palestinian war crimes in 2014 Gaza conflict,” *New York Times*, March 25, 2015). More generally, the Israeli military alleged that a significant number of Hamas's rockets fell short and landed in Gaza (William Booth, “Here’s what really happened in the Gaza war [according to the Israelis],” *Washington Post*, September 3, 2014).

First assume that releasing the prisoners would be to prevent Hamas from causing harm. Then C survives in Diplomacy because Israel's releasing the prisoners prevents C from being killed (as C is in Status Quo). So if OPE occurs and Israel kills C—that is, if War occurs—then C's death is harm that would have occurred anyway, had Israel not prevented it. Assume instead that releasing the prisoners would not be to prevent harm, but would be merely to refrain from causing Hamas to cause harm. Then C survives in Diplomacy *not* because Israel's releasing the prisoners prevents C from being killed, but simply because Israel does not cause Hamas to fire the rocket that kills C. So if OPE occurs and Israel kills C, then C's death is *not* harm that would have occurred anyway, had Israel not prevented it. Hence, if we give up on determining whether releasing the prisoners is to prevent harm, then we cannot say whether Israel's killing C is to cause harm that would occur anyway, were Israel not to prevent it.

The third distinction is between killing and preventing people from being saved (Hanser 1999). Consider an example from McMahan:

Suppose that I am about to be killed by a culpable threatener. If the only way I can defend myself is through defensive action ... that will kill two innocent bystanders as a side effect, then I am not permitted to engage in effective defensive action. ... But suppose that ... I can kill [the attacker] without killing anyone else. But I know that he is a celebrated surgeon who is scheduled to perform two life-saving surgeries tomorrow that no one else in the world is capable of performing. If I kill him today, I will prevent him from saving two innocent people tomorrow. But in this case it is intuitively permissible ... for me to kill him in self-defense. (2014: 111)

In this case, killing two innocent people appears to be more objectionable than preventing two from being saved. We will not be able to draw such a distinction if we forget about what harm an act prevents.

Suppose that if Israel releases the prisoners in question, some will later be targeted for killing by some agent. Perhaps the agent is an individual or other entity within the Israeli government, or perhaps the agent is a foreign state engaged in a conflict in the region. Either

way, the agent's options include the following: (1) do nothing now, in which case the prisoners will be released, some will be targeted for killing, and civilians will be killed as a side effect of the targeted killings; (2) prevent the release of the prisoners, whether by offering incentives to the relevant decision-making authority, making threats, or some other means. First assume that Israel's releasing the prisoners *would* be to prevent harm—and, in particular, would be to save certain civilians from being killed by Hamas. Then the choice between 1 and 2 is a choice between killing civilians (as in 1) and preventing civilians from being saved (as in 2). Next assume that Israel's releasing the prisoners would *not* be to prevent harm, but would be merely to refrain from causing harm. Then the choice between 1 and 2 is *not* a choice between killing and preventing people from being saved; it is instead a choice between killing (as in 1) and preventing Israel from refraining from causing harm (as in 2). In other words, it is a choice between killing and causing another agent to cause people to be killed. Hence, if we give up on determining whether releasing the prisoners is to prevent harm, then we cannot say whether the choice between 1 and 2 is a choice between killing and preventing people from being saved. And likewise for other conduct whose causal/preventive status is unresolved.

Ethicists commonly draw one or more of the distinctions mentioned above: doing vs. allowing harm; causing harm that would vs. would not occur anyway, were you not to prevent it; and killing vs. preventing people from being saved. The current proposal is a relatively radical one insofar as it would leave us unable to draw these distinctions. It would be preferable to leave open the possibility that such distinctions should (and can) be made.

## 2.6 Argument by Cases

Briefly, here is the state of things. We wish to know the harm prevented by Operation Protective Edge (OPE), or at least how to go about identifying this harm. According to the argument in

section 1, the harm prevented by OPE corresponds to the difference between the harm that occurs in the actual world and the harm that occurs in a world where OPE does not occur and certain other conditions are met. Among these conditions is that nothing *else* prevents the harm that is prevented, in actuality, by OPE. Suppose that if OPE had not occurred and Israel had not used other military means to achieve the same ends, then Israel would have released some of the prisoners requested by Hamas. Suppose also that some of the rockets actually destroyed by OPE would have never been fired, had the prisoners been released. Does this mean that the prisoners' release would have prevented OPE-prevented harm—prevented, that is, some of the harm that is prevented in actuality by OPE? We have not yet answered this question. And it appears that we *must* have an answer in order to know whether we should compare the actual world to one in which the prisoners are released. Yet this appearance is deceiving, as I explain in this final section.

First assume that releasing the relevant prisoners *would* be to prevent harm, and hence would be to prevent OPE-prevented harm in particular. Then we cannot compare the actual world to one in which Israel releases the prisoners. For that would be to compare the actual world, in which OPE occurs, to a world in which not all of the OPE-prevented harm occurs. This would undercount the harm prevented by OPE.

Next assume that releasing the prisoners would *not* be to prevent harm, but would be merely to refrain from causing harm. Then continuing to *hold* the prisoners would be to *aggravate* the threat posed by Hamas. In this case as well, we cannot compare the actual world to one in which Israel releases the prisoners. For that would again be to compare the actual world to one in which not all of the OPE-prevented harm occurs. Consider the following case for comparison. I put two substances in your coffee: one is a poison and the other is the antidote.

You drink the coffee and are unharmed. What harm does the antidote prevent? It prevents whatever harm would have occurred, had you drank the poison and not drank the antidote.<sup>153</sup> We will undercount the harm prevented by the antidote if we compare the actual world to one in which I refrain from adding the poison to your coffee. In other words, we will undercount the harm prevented by the antidote if we compare the actual world to one in which I refrain from initiating the threat that the antidote averts. Likewise, if Israel's continuing to hold the prisoners would be to aggravate the threat posed by Hamas, then we will undercount the harm prevented by OPE if we compare the actual world to one in which Israel refrains from aggravating the threat posed by Hamas.

So whether or not releasing the relevant prisoners would be to prevent Hamas from causing harm, we must compare the actual world to one in which Israel does not release the prisoners. More precisely, we must compare the actual world to one in which Israel does not release *more* prisoners than in actuality, or release them *earlier* than in actuality. For that would be to undercount the harm prevented by OPE—either because it would be to compare the actual world with one in which Israel uses other means to prevent OPE-prevented harm, or else because it would be to compare the actual world with one in which Israel does not generate some of the threats that OPE averts. Hence in order to determine the harm prevented by OPE, we need not determine whether releasing the prisoners would be to prevent harm, or would be merely to refrain from causing harm. And likewise for Israel's ending the blockade of Gaza, scaling back settlement in the Palestinian territories, releasing previously withheld tax funds, and any other conduct whose causal/preventive status is unresolved.

---

<sup>153</sup> Assuming that there were no back-up preventers.

Given what has just been said, one might conclude that there was never a genuine *problem* here to begin with, and hence that this entire discussion was unnecessary. There are two reasons why this would not quite be true.

First, while we may be able to determine the harm prevented by OPE—or at least say how to go about identifying this harm—the discussion has revealed several other problems that have not been resolved. We have not determined whether Israel’s releasing the prisoners would be to prevent Hamas from causing harm. Without making this determination, we will be unable to apply the distinctions considered in section 2.5. We will be unable to say whether continuing to hold the prisoners would be to do harm or merely to allow harm. For some individuals who are harmed by OPE, we will be unable to say whether they would have suffered such harm anyway, had Israel not prevented them from being harmed. And we will not know whether a choice between killing and preventing the release of the prisoners is a choice between killing and preventing people from being saved. And likewise for the other conduct mentioned above (Israel’s ending the blockade of Gaza, etc.).<sup>154</sup>

Second, our discussion has suggested a procedure for identifying the harm prevented by an act. This section has argued that whether or not releasing the prisoners would be to prevent Hamas from causing harm, we must compare the actual world to one in which Israel does not release more of the prisoners than in actuality.<sup>155</sup> And likewise for the other conduct mentioned

---

<sup>154</sup> Does this mean that we are no better off now than we were at the end of section 2.5—that we might as well have settled for the solution proposed in that section? No, for that proposal was to forget about determining the harm prevented by acts *generally*. Determining the harm prevented by OPE, in particular, will allow us to bring deontological distinctions to bear on OPE itself.

<sup>155</sup> By analogous reasoning, the relevant world is also one in which Israel does not release *fewer* of the prisoners than in actuality. Otherwise we would *overcount* the harm prevented by OPE.

above. This is *almost* to say that OPE prevents the harm that does not occur but would have, had OPE not occurred and had Israel otherwise done just as it does in actuality.<sup>156</sup>

This is different than the procedures suggested by others. Mellow and Hurka each endorse a moralized counterfactual comparison. Both of Mellow's proposals direct us to a world in which the agent does not *immorally* contribute to the relevant threat. Hurka asks us to consider what would have happened, had the agent pursued the least beneficial *morally permissible* alternative to the act in question. On the other hand, McMahan suggests that we need not consider counterfactuals at all.<sup>157</sup> Each of these authors is discussing the requirement of proportionality, which is typically characterized in terms of a comparison between the harm caused by an act and the good achieved.<sup>158</sup> The good achieved often includes the prevention of harm. As we have seen, it *is* necessary to consider counterfactuals when determining the harm prevented by an act. The relevant counterfactuals are not moralized, however. For the harm prevented by an act does not depend on what would have happened, had the agent acted permissibly. For example: Had I acted permissibly, I would have put neither the antidote nor the poison in your coffee, and you would not have died. Putting the antidote in your coffee

<sup>156</sup> But only *almost*. For if OPE had not occurred and Israel had otherwise done just as it does in actuality, then some actions that *occur* in actuality would have prevented *more harm* than they do in actuality. (See section 2.1.) So we must consider a world in which these actions do not prevent any more harm than they do in actuality—or at least in which they do nothing more to avert the threat posed by Hamas's rockets than they do in actuality. There may be no general formula for how to pick out such worlds, but it is often possible to do so in particular cases. For an example, see the third, fourth, and fifth paragraphs in section 1.3.

<sup>157</sup> See section 1.4, which discusses McMahan and McKim 1993; and section 2.3, which discusses Mellow 2006, Hurka 2008, and McMahan 2013/2014.

<sup>158</sup> “Proportionality in defense is a relation between the good and bad effects of a defensive act” (McMahan 2013/2014: 1); “the good achieved by the war must be proportionate to the harm caused” (Mellow 2006: 439); “a war … is wrong if the destruction it causes is excessive, or out of proportion to, those benefits” (Hurka 2008: 127).

nevertheless prevents your death.<sup>159</sup> We should consider what other options I had, of course, and what their effects would have been. But doing so is not part of determining the effects of a given act, and hence is not part of the proportionality calculation.

## References

- Bronner, Ben. 2018. Two ways to kill a patient. *Journal of Medicine and Philosophy* 43.1: 44–63.
- Dowe, Phil. 2000. *Physical Causation*. Cambridge: Cambridge University Press.
- Foot, Philippa. 1967. The problem of abortion and the doctrine of the double effect. *Oxford Review* 5: 5–15. Reprinted in *Killing and Letting Die*, edited by Bonnie Steinbock and Alastair Norcross, 2nd ed., New York: Fordham University Press, 1994.
- . 1984. Killing and letting die. In *Abortion: Moral and Legal Perspectives*, edited by Jay L. Garfield and Patricia Hennessey. Amherst: University of Massachusetts Press.
- . 1985. Morality, action, and outcome. In *Morality and Objectivity: A Tribute to J. L. Mackie*, edited by Ted Honderich. London: Routledge and Kegan Paul.
- Hanswer, Matthew. 1999. Killing, letting die and preventing people from being saved. *Utilitas* 11.3: 277–295.
- Harman, Elizabeth. 2004. Can we harm and benefit in creating? *Philosophical Perspectives* 18.1: 89–113.
- . 2009. Harming as causing harm. In *Harming Future Persons: Ethics, Genetics and the Non-Identity Problem*, edited by Melinda A. Roberts and David T. Wasserman. Dordrecht: Springer.
- Howard-Snyder, Frances. 2011. Doing vs. allowing harm. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, winter 2011 edition. <https://plato.stanford.edu/archives/win2011/entries/doing-allowing/>.
- Hurka, Thomas. 2005. Proportionality in the morality of war. *Philosophy and Public Affairs* 33.1: 34–66.
- . 2008. Proportionality and necessity. In *War: Essays in Political Philosophy*, edited by Larry May with the assistance of Emily Crookston. Cambridge: Cambridge University Press.
- Illari, Phyllis McKay and Jon Williamson. 2011. Mechanisms are real and local. In *Causality in the Sciences*, ed. Phyllis McKay Illari, Federica Russo, and Jon Williamson. New York: Oxford University Press.

---

<sup>159</sup> This is most clearly true if we suppose that I still would have put the poison in your coffee, had I not put the antidote in; and hence that you would have died, had I not added the antidote.

- Jensen, Steven J. 2011. Killing and letting die. In *The Ethics of Organ Transplantation*, edited by Steven J. Jensen. Washington, DC: Catholic University of America Press.
- Kamm, F. M. 1996. *Morality, Mortality, Vol. 2: Rights, Duties, and Status*. New York: Oxford University Press.
- Kavka, Gregory S. 1991. Was the Gulf War a just war? *Journal of Social Philosophy* 22.1: 20–29.
- Kuhse, Helga. 1987. *The Sanctity-of-Life Doctrine in Medicine: A Critique*. Oxford: Oxford University Press.
- Klocksiem, Justin. 2012. A defense of the counterfactual comparative account of harm. *American Philosophical Quarterly* 49.4: 285–300.
- Koons, Robert C. 2003. Review of *Physical Causation*, by Phil Dowe. *Philosophy and Phenomenological Research* 67.1: 244–248.
- Kvanvig, Jonathan L. 1986. *The Possibility of an All-Knowing God*. New York: St. Martin's Press.
- McGee, Andrew. 2014. Does withdrawing life-sustaining treatment cause death or allow the patient to die? *Medical Law Review* 22.1: 26–47.
- . 2015. Acting to let someone die. *Bioethics* 29.2: 74–81.
- McMahan, Jeff. 1993. Killing, letting die, and withdrawing aid. *Ethics* 103.2: 250–279.
- . 2002. *The Ethics of Killing: Problems at the Margins of Life*. New York: Oxford University Press.
- . 2013/2014. Proportionate defense. *Journal of Transnational Law and Policy* 23: 1–36.
- . 2014. Self-defense against justified threateners. In *How We Fight: Ethics in War*, edited by Helen Frowe and Gerald Lang. New York: Oxford University Press.
- McMahan, Jeff and Robert McKim. 1993. The just war and the Gulf War. *Canadian Journal of Philosophy* 23.4: 501–541.
- Mellow, David. 2006. Counterfactuals and the proportionality criterion. *Ethics and International Affairs* 20.4: 439–454.
- Shiffrin, Seana Valentine. 1999. Wrongful life, procreative responsibility, and the significance of harm. *Legal Theory* 5.2: 117–148.
- . 2012. Harm and its moral significance. *Legal Theory* 18.3: 357–398.
- Thomson, Judith Jarvis. 1971. A defense of abortion. *Philosophy and Public Affairs* 1.1: 47–66.
- Walen, Alec D. 2019. *The Mechanics of Claims and Permissible Killing in War*. New York: Oxford University Press.
- Woollard, Fiona. 2008. Doing and allowing, threats and sequences. *Pacific Philosophical Quarterly* 89.2: 261–277.
- . 2013. If this is my body...: A defense of the doctrine of doing and allowing. *Pacific*

*Philosophical Quarterly* 94.3: 315–341.

### Appendix: Another Answer for Abe and Bea

In section 1.1, we considered three answers to the question of what harm was prevented by your throwing a bowling ball into the path of the rocks heading for Bea. Each of those answers proved to be inadequate, including the third:

(Third answer) The throw prevents whatever harm would have occurred, had the throw not occurred and had you done nothing other than what you do in actuality.

We might try another answer:

(Fourth answer) The throw prevents whatever harm would have occurred, had the throw not occurred and had your other acts produced *all and only* the effects they do in actuality.<sup>160</sup>

This fourth answer will not work either, however.

There are a few more details of the case that have not yet been mentioned. The rocks are all deflected by time  $t_1$ . Just to be safe, you push Bea out of the rocks' former path at  $t_2$ .<sup>161</sup> And under no condition would the rocks have reached Bea before  $t_3$ .

What harm does the throw prevent, in light of these facts? Had the throw not occurred, the push would have moved Bea out of the path of the oncoming rocks and she would not have been harmed. Given this fact, some of our *previous* answers might entail that the throw prevents

<sup>160</sup> If the roll had produced all and only the effects that it produces in actuality, then it would have deflected all and only the rocks that it deflects in actuality, solving the problem that was posed for the third answer.

<sup>161</sup> The roll deflects the rocks *to the north* and the push moves Bea *to the south*. So the push takes Bea further away from danger, rather than putting her back in danger.

no harm.<sup>162</sup> According to the *fourth* answer, however, we ask what would have happened had the throw not occurred *and* had your other acts produced all and only the effects they do in actuality. In actuality, the push does *not* move Bea out of the path of oncoming rocks. So the fourth answer does not entail that the throw prevents no harm.

The fourth answer is nonetheless inadequate. According to this answer, the harm prevented by the throw is the harm that occurs in some world—perhaps the closest world—in which certain conditions are satisfied. These conditions include:

1. At  $t_2$ , you push Bea out of the rocks' original path.<sup>163</sup>
2. You do not push Bea out of the path of oncoming rocks.<sup>164</sup>
3. The throw does not occur, and you do not deflect or stop the rocks in some other way.
4. The rocks do not reach Bea's initial location before  $t_3$ .<sup>165</sup>

Consider what happens in a world—in *every* world—where condition 1 is satisfied. If you push Bea out of the rocks' *original* path, then either you push Bea out of the path of *oncoming* rocks, contrary to condition 2; or you have *stopped or deflected* the rocks, contrary to condition 3; or the rocks have *already* reached Bea, contrary to condition 4. The only other possibility is that something or someone *other* than you has stopped or deflected the rocks—in which case no harm occurs and our fourth answer then entails that the throw prevents no harm. So either the

<sup>162</sup> That would be the wrong conclusion. After all, the throw *actually* deflects (some of) the rocks—and it does so *before* you push Bea. Quite literally, you do not move Bea out of harm's way. We can even stipulate that the rocks are deflected before you decide to move Bea at all.

<sup>163</sup> The rocks' *original* path is the path that they had been on, in actuality, before they were deflected. If condition 1 is not met, then the push does not have *all* the effects that it has in actuality.

<sup>164</sup> If condition 2 is not met, then the push does not have *only* the effects that it has in actuality.

<sup>165</sup> If condition 4 is not met, then the stipulations introduced above are not all satisfied.

fourth answer produces the wrong conclusion, or else the conditions provided by the answer cannot all be satisfied.<sup>166</sup>

What we *want* to do is focus on a world in which none of your acts prevent *more harm* than they do in actuality. It is clear how to do that in the present case, at least with respect to your pushing Bea. We simply rule out worlds in which you push Bea out of the path of oncoming rocks. (We rule out such worlds, that is, but we do not *also* rule out worlds in which you fail to push Bea out of the rocks' *original* path. Our fourth answer went wrong by ruling out both sets of worlds.) But this solution is tailored to the details of the present case, and so it is not the sort of generalizable answer that we had been looking for.

---

<sup>166</sup> It may be that we have no choice but to take the second horn of this dilemma, at least if counterfactual conditionals are being interpreted in a non-backtracking manner. The first horn seems to involve backtracking, insofar as it seems to involve reasoning such as the following: *had the push at  $t_2$  not moved Bea out of the way of oncoming rocks, then something must have stopped or deflected the rocks prior to  $t_2$ .*