

# **Authority and Harm in War**

**Jonathan Parry**

Thesis Submitted for the of  
Degree of Doctor of Philosophy  
(Philosophy)

June 2014

Department of Philosophy  
The University of Sheffield



## **Abstract**

Two features of warfare are particularly ethically striking. Firstly, and most obviously, war involves killing and maiming on a grand scale. Secondly, war involves large numbers of individuals obeying the commands of those who claim to possess authority over them, such as their superior officers and political leaders.

Whereas leading contemporary work on the ethics of war focuses on the former aspect, this thesis investigates the relevance of the latter. The project is orientated around a discussion of the traditional just war requirement that wars be waged by a 'legitimate authority', and makes three central contributions.

Firstly, I argue that the authority criterion plays a much more important role within just war theory than is commonly supposed. Rather than simply imposing a necessary condition for justifying the resort to war, the criterion also plays a crucial role in judgements about the permissibility of individuals' conduct in war. More specifically, the criterion captures the idea that individuals who fight on behalf of a certain type of entity are subject to a more extensive range of permissions to cause harm than those that apply to private actors.

Secondly, I assess the extent to which the commands of authorities are capable of affecting the deontic status of acts of harming. Drawing on a 'service-based' account of the justification of authority, I argue for the controversial conclusion that individuals may be morally required, all-things-considered, to obey commands to cause serious harm to others, even in cases where doing so would be straightforwardly morally unjustified in the absence of the command. This argument has important revisionary implications for our understanding of both the ethics of harm and the moral limits of the duty to obey.

Thirdly, I employ this argument to provide a qualified defence of the authority criterion against powerful objections raised by an influential 'reductivist' approach to just war theory. While they diverge in important respects, I argue that my defence is compatible with the basic commitments that motivate a reductivist view.



## **Acknowledgements**

I owe thanks to many people for their help in producing this thesis. Four persons deserve special mention. My supervisors Daniel Viehoff and Jimmy Lenman have been fantastic throughout. I also greatly benefitted from the input of Jeff McMahan, who went above and beyond the call of duty in taking on the role of unofficial supervisor. I am also grateful to Helen Frowe, who first introduced me to the topics addressed in the thesis as an undergraduate and encouraged me to pursue them further.

For written comments on parts of the thesis, which are incredibly time-consuming to produce, I am extremely grateful to Saba Bazargan, Yitzhak Benbaji, Ian Fishback, Jeff McMahan, Helen Frowe, Seth Lazar, Michael Neu, Jonathan Quong, Daniel Statman, and Victor Tadros.

I presented material from the thesis at the following conferences and workshops: MANCEPT Annual Workshops in Political Theory (Manchester, 2011); Brave New World (Manchester, 2012); The Society for Applied Philosophy Annual Conference (Oxford, 2012); The Association for Legal and Social Philosophy (Stirling, 2013); and a workshop in Rogoznica, Croatia (2013). I also presented material at The University of Toronto (2014); the Oxford War Discussion Group (2013); and at several seminars at The University of Sheffield. Many thanks to audience members at these events for stimulating and helpful discussions.

I also greatly benefitted from innumerable conversations, philosophical or otherwise, with my fellow graduate students at the University of Sheffield. In particular with Carl Fox, Tash McKeever, Jessica Begon, Angie Pepper, Peter Cavan, Kathy Puddifoot, Clara Sandelind, Richard Healey, Nicola Kemp, Jan Kandiyali, Angie Pepper and Stephen Wright. Special thanks to Steve, and to my parents Janette and Gareth, for proofreading portions of the thesis.

Work on this thesis was supported by grants from the Arts and Humanities Research Council, the Society for Applied Philosophy, and the Faculty of Arts and Humanities at the University of Sheffield. My thanks to these organisations.

For everything else, and much more besides, I thank my wife Kwun Ye.



## Table of Contents

<b>Introduction</b> .....	01
<b>Chapter 1 – The Role of the Authority Criterion</b> .....	11
1.1 Introduction .....	11
1.2 The Structure of Orthodox Just War Theory .....	12
1.3 The Standard Interpretation of the Authority Criterion .....	14
1.4 The Discontinuity Thesis .....	18
1.5 The Demarcation Requirement .....	23
1.6 The Revised Interpretation of the Authority Criterion .....	24
1.7 The (Real) Challenge of Irregular Belligerency .....	28
1.8 Classical Just War Theory and the Authority Criterion .....	35
1.9 Conclusion .....	39
<b>Chapter 2 –Scepticism Regarding the Authority Criterion</b> .....	40
2.1 Introduction .....	40
2.2 Reductivism: An Overview .....	42
2.2.1 The Continuity Thesis .....	42
2.2.2 The Content Thesis .....	44
2.3 Jettisoning the Restrictive Authority Criterion .....	47
2.3.1 The First Spectrum Argument .....	50
2.3.2 Deflating the Restrictive Authority Criterion .....	54
2.4 Jettisoning the Permissive Authority Criterion .....	56
2.4.1 The Continuity Thesis .....	57
2.4.2 Rejecting the Orthodox View .....	58
2.4.3 Rejecting the Classical View .....	60
2.4.4 The Second Spectrum Argument .....	62
2.5 Conclusion .....	64
<b>Chapter 3 – Non-Reductivist Approaches to Just War Theory</b> .....	66
3.1 Introduction .....	66
3.2 Motivating Non-Reductivism .....	67
3.2.1 The Permissibility of Participation in Unjust Wars .....	67
3.2.2 The Avoidance of Pacifism .....	68
3.3 Non-Reductivist Approaches .....	70
3.3.1 The Collectivist View .....	72
3.3.2 The Associative Duty View .....	75
3.4 The Relevance of Non-Reductivism .....	77

3.5 The War Crimes Objection .....	78
3.6 Alternative Groundings for Constraints on Targeting .....	84
3.6.1 Role-Based Obligations .....	85
3.6.2 The Avoidance of Cowardice.....	87
3.6.3 Modes of Agency.....	88
3.6.4 Risk.....	90
3.6.5 Vulnerability .....	91
3.7 Conclusion.....	97
<b>Chapter 4 – Authority and Harm .....</b>	<b>99</b>
4.1 Introduction .....	99
4.2 Authority: The Basics .....	100
4.3 Authority and the Ethics of Harming.....	102
4.4 The Ethics of Harm: <i>Completeness</i> and <i>Immunity</i> .....	104
4.5 Two Claims .....	108
4.6 Opposing the First Claim .....	111
4.6.1 <i>A Priori</i> Philosophical Anarchism .....	111
4.6.2 The Invalidation View .....	113
4.6.3 The <i>Pro Tanto</i> View.....	115
4.7 Service Accounts of Authority.....	116
4.8 The Authority View of Harm.....	124
4.9 Three Clarifications .....	128
4.9.1 The Irrelevance of Certain Services .....	128
4.9.2 The Option of Inaction .....	129
4.9.3 Cumulative Effects.....	130
4.10 The Moderate Objections Revisited .....	132
4.10.1 Against the <i>Pro Tanto</i> View.....	132
4.10.2 Against the Invalidation View.....	135
4.11 The ‘Irrelevant Reasons’ Objection .....	139
4.12 Defending the Second Claim.....	146
4.13 <i>Interaction</i> and Standard Justified Threat Cases .....	147
4.14 Authorised Threateners and <i>Immunity</i> .....	149
4.15 Authorised Threateners and Agent-Relativity .....	152
4.16 Authority and Constraints.....	156
4.17 Conclusion.....	159
<b>Chapter 5 – Authority and War .....</b>	<b>160</b>
5.1 Introduction .....	160
5.2 The Authority View of War: A First Cut.....	161
5.3 A Qualified Defence of the Permissive Authority Criterion.....	163
5.4 A Qualified Defence of the Restrictive Authority Criterion .....	164



5.5 The Authority View of War and Reductivism .....	168
5.6 Authority and Responsibility .....	173
5.7 The <i>A Posteriori</i> Objection.....	176
5.8 Response.....	180
5.8.1 Three Concessive Remarks.....	180
5.8.2 Fine-Grained Domains – <i>Jus ad Bellum</i> .....	182
5.8.3 Fine-Grained Domains – <i>Jus in Bello</i> .....	186
5.8.4 Authority and Coordination .....	187
5.9 The Authority View of War: A Second Cut .....	189
5.10 An Asymmetry .....	189
5.11 The War Crimes Objection Revisited.....	190
5.12 Response .....	191
5.12.1 Distinct Forms of Agent-Relativity .....	191
5.12.2 Accepting the Implication.....	194
5.13 Conclusion .....	196
<b>Conclusion.....</b>	<b>199</b>
<b>Bibliography.....</b>	<b>205</b>



## Introduction

The past decade has witnessed a remarkable resurgence in philosophical interest in the ethics of warfare. One of the key drivers behind this trend is a growing dissatisfaction with the common assumption – either explicit or implicit – that warfare is in some sense morally special, governed by its own *sui generis* moral rules. Instead, many theorists now believe that an acceptable theory of the morality of war must be consistent with the moral principles that we accept in other domains of action. In particular, with those principles that govern the permissibility of causing serious non-consensual physical harm to others.

This approach has led to the reassessment and rejection of many of the standard tenets of just war theory, on the ground that they are irreconcilable with our most stable views about the morality of harming more generally. This criticism has paved the way for the positive task of constructing a new theory of the ethics of war on more plausible foundations, drawing on the resources of practical philosophy more generally. It is to this project that my thesis aims to contribute.

To date, the leading contributions to the field have primarily drawn on the resources supplied by work in normative and applied ethics. In particular, many contemporary just war theorists aim to elucidate the morality of war in terms of sophisticated accounts of the moral principles that govern the use of defensive force between private individuals. By contrast, my thesis is concerned with the contribution that other branches of practical philosophy may make to uncovering the morality of war. In particular, it investigates the extent to which discussions that usually take place within the confines of political and legal philosophy can productively be brought to bear on debates within just war theory, and the morality of interpersonal harming more generally.

My thesis aims to shed new light on this question, by focussing on one specific component of just war theory – the traditional requirement that wars be waged by a ‘legitimate authority’. In the broadest possible terms, the inclusion of an authority criterion within the theory reflects the idea

that moral judgements regarding armed conflict are sensitive to facts about the *status* or *standing* of the belligerent parties engaging in that conflict.

I focus on the authority criterion for two main reasons. Firstly, unlike other standard components of the theory – such as those requiring wars to have a just cause, be the option of last resort, and prohibit targeting non-combatants – the authority criterion is an overtly *political* requirement, having no straightforward analogue in ordinary interpersonal morality. Furthermore, the topics of authority and legitimacy are of absolute centrality within political and legal philosophy. So, if we are interested in what (if anything) these branches of practical philosophy have to say about the morality of war, the authority criterion provides an obvious starting point for our inquiry.

Secondly, the authority criterion receives relatively little attention in contemporary discussions. For example, Michael Walzer's hugely influential *Just and Unjust Wars* contains no explicit discussion of the criterion.<sup>1</sup> Furthermore, of the few theorists to offer sustained discussions of the criterion, many are sceptical that considerations of authority have any significant role to play in the best theory of the morality of war. This lack of interest should be surprising, given a historical perspective on the just war tradition. For the founding fathers of the tradition, the authority criterion was considered the first and most important condition of just war. Yet, within current debates it remains the most under-theorised component.

With some sympathy for this venerable position, the overarching aim of the thesis is to deploy the resources of contemporary political and legal philosophy to show that considerations of legitimate authority are far more central to the morality of war than is commonly supposed, and to offer a novel (though importantly qualified) defence of the authority criterion against its critics.

---

<sup>1</sup> This omission is emphasised in Brian Orend, *Michael Walzer on War and Justice* (Cardiff: University of Wales Press, 2000), Ch.4.

<sup>2</sup> I leave aside here the more recently emphasised fields of justice in ending wars (*jus ex bello*) and justice following war (*jus post bellum*). For the former, see Darell Mollendorf, 'Jus ex Bello', *Journal of Political Philosophy* 16, No. 2 (2008), 123-136; Cecile Fabre, 'War Exit', *Ethics* (forthcoming). For the latter, see Gary Bass, 'Jus Post Bellum', *Philosophy and*

The thesis is divided into five chapters. The first chapter is largely conceptual in nature, providing a precise account of the role that the authority criterion plays within mainstream just war theory. I do so by distinguishing two questions we can ask about the authority criterion. Firstly, we can ask what normative consequences are meant to follow from the possession or non-possession of war-waging authority. I term this the question of *normative effect*. Second, we can ask what properties an entity must possess in order to acquire this authority. I term this the question of *relevant properties*.

In terms of its normative effect, the authority criterion is standardly interpreted as functioning solely as a condition of *jus ad bellum*, providing a necessary condition for justifying the resort to war. On this view, in order for a war to be justified, it must be initiated and waged by an entity that possesses the relevant properties. Wars fought by groups that lack these properties are therefore unjust. The criterion thus performs an essentially *restrictive* function. Its normative effect is to restrict the waging of just wars to a certain class of entities, standardly interpreted to mean those entities that possess the properties of statehood or, less restrictively, those that aspire to statehood.

In opposition to this interpretation, I argue that the authority criterion plays a much more expansive role within just war theory than standardly assumed. In addition to its restrictive function within *jus ad bellum*, the authority criterion also performs a crucial role within the *jus in bello* component of the theory, which is concerned with judgements about the permissibility of individuals' conduct in war, rather than with the justice of the war itself. On the interpretation that I argue for, the authority criterion serves to capture the idea that individuals who fight on behalf of certain kinds of entity enjoy additional moral permissions to cause harm by means of war. On this alternative interpretation, the normative effect that follows from the possession of the authority to wage war is essentially *permissive*. Acting on behalf of a legitimate war-waging authority serves to increase an agent's moral permissions to cause harm to others, beyond those that they would possess *qua* private actor.

With this conceptual background in place, the following four chapters turn to the central normative question raised by the preceding discussion: Can the authority criterion, in either its restrictive or permissive roles, be given a principled moral defence?

The second chapter begins this task by considering criticisms of the authority criterion, put forward by proponents of an increasingly influential 'reductivist' approach to the ethics of war. While reductivism encompasses a spectrum of possible views, it can be characterised in terms of a commitment to two basic theses. Firstly, it holds that war is morally continuous with all other activities and is governed by precisely the same moral principles that govern interpersonal harming in all other circumstances. Secondly, reductivism holds that the killing and injuring constitutive of warfare, when morally justified, is justified either because the individuals harmed have rendered themselves *liable* to that harm, and therefore suffer no violation of their rights, or because their right not to be harmed is overridden as the impartial *lesser-evil*.

A reductivist view has important revisionary implications, many of which have been discussed in great detail. For the purposes of my inquiry, I focus specifically on explaining how reductivism contains the resources for powerful objections to the authority criterion, in both its restrictive and permissive interpretation. In fact, as I explain, this is perhaps the most revisionary aspect of reductivism.

Reductivists reject the standard, restrictive interpretation of the authority criterion, because they hold that individuals' permissions to use force on grounds of liability and lesser-evil are *sufficient* to justify both the resort to war and its conduct. Given such a view, it cannot be a necessary condition of permissible warfare that it be waged by a particular type of entity, since liability and lesser-evil justifications are available to individuals in their private capacity. No authorisation is required in order to permissibly act on such justifications.

Reductivists also reject the permissive interpretation of the authority criterion, because they hold that individuals' permissions to use force on grounds of liability or lesser-evil are *necessary* for justifying both the resort

to war and its conduct. The fact that an agent participates in war on behalf of a certain kind of entity cannot serve to extend their permissions to cause harm to others. This seems highly plausible. If an individual retains their right not to be killed, and this right is not overridden by weightier considerations, how can it somehow become morally permissible for an agent to kill or maim that individual simply in virtue of that agent's membership or role in a certain kind of organisation?

For reductivists, then, the authority criterion is at best redundant or at worst deeply pernicious and ought to be jettisoned from just war theorising. With this critique in place, the following three chapters of the thesis investigate the prospects for rehabilitating the authority criterion. In particular, I argue for a qualified defence of the permissive, *in bello* interpretation of the criterion.

This requires showing that one of the claims associated with reductivism is mistaken. To this end, in the third chapter I consider a rival 'non-reductivist' approach to the ethics of war. On this view, reductivists are mistaken in holding that the morality of war is determined solely by considerations of liability and lesser-evil, on the ground that warfare differs from ordinary forms of interpersonal violence in some morally relevant respect. Proponents of non-reductivism typically locate this difference in the fact that war involves the use of force by groups of individuals engaged in morally significant relationships with one another. A central claim of non-reductivism is that these relationships are independently capable of generating permissions for killing and injuring in war, thus denying reductivism.

An assessment of non-reductivism is thus particularly relevant for the purposes of evaluating the permissive authority criterion, since it potentially provides the resources for a defence of the idea that participants in war enjoy additional permissions to cause harm in virtue of fighting on behalf of a certain kind of entity.

However, I argue that non-reductivism must be rejected. The argument works by pushing a dilemma onto non-reductivists: If non-reductivists are successful in showing that acting on behalf of a certain type of entity gives

rise to additional *permissions* to cause harm in war, they must also jettison the most intuitive *restrictions* on conduct in war – most saliently, the prohibition on intentionally killing morally innocent non-combatants. Since this conclusion is unacceptable, we must conclude that non-reductivism fails.

In the fourth chapter I provide the groundwork for an alternative defence of the authority criterion. I do so by exploring the connections between two central topics in moral and political philosophy: the morality of interpersonal harming on the one hand, and the moral legitimacy of authority on the other.

These topics are routinely discussed in isolation from one another. This is understandable, but unfortunate. Ultimately, theorists working in both these areas are engaged in the wider project of establishing what reasons moral agents have for action – with what individuals all-things-considered ought and ought not to do. We may characterise those working on the morality of harming as concerned with the basic question, ‘When, if ever, do individuals have undefeated moral reason to cause harm to others?’. Theorists of authority, by contrast, are concerned with a different question, ‘When, if ever, do individuals have undefeated moral reason to obey the commands of another?’ The former are concerned with the justification of harm, the latter with the justification of obedience.

Set up this way, we can see that there are a range of possible cases in which determining whether an agent is justified in causing a particular harm – with whether she has undefeated reason to do so – requires an answer to an important third question, which combines our response to each of the above, namely, ‘When, if ever, do individuals have undefeated moral reason to obey commands which require them to cause, or refrain from causing, harm to others?’

In response to this question, I defend two main claims. Firstly, I argue that, under certain conditions, the command of an authority can provide agents with an independent moral justification for causing serious physical harm to others. Controversially, I argue that this may be true even in cases where the harm would be straightforwardly unjustified in the absence of



the command. I imagine that many will find this claim highly unintuitive, even repugnant. Given that any plausible account authority must place limits on the obligation to obey, it is natural to conclude that wherever the precise limits lie, commands that require imposing serious unjust harm on others must surely exceed them, given the moral gravity of the wrongdoing involved.

I argue that this view, while intuitive, is mistaken. My argument draws on a service-based account of authority, defended most influentially by Joseph Raz. On this view, authorities possess the moral power to place their subjects under obligations to obey when, and to the extent that, their subjects better comply with reasons that apply to them by obeying the authority, compared to not obeying. This view thus offers a broadly instrumental justification of the duty to obey. Authorities are justified because they enable individuals to better do what they have reason to do independently of the authority.

I argue that the very same reasoning can be applied, *mutatis mutandis*, to the reasons that govern the distribution of harm. Put briefly, on the service-based view that I advance, if individuals will overall better distribute harms in accordance with right reason by obeying an authority, compared to not obeying, then obedience is morally required (and therefore justified). Importantly, this includes cases in which the authority issues commands that are mistaken and require distributing harm in ways not supported by the balance of authority-independent reasons. In such cases, I argue, the subject of the command may still have an undefeated reason – an instrumental reason – to obey the command. I term this argument the ‘Authority View of Harm.’ If successful, it reveals that the range of potential justifications for harming is broader than standardly assumed. In addition to standard justifications, such as liability and lesser-evil, the Authority View posits the existence of what I term *authority-based* justifications for harm.

With this argument in place I turn to a defence of a second claim, which concerns the permissibility of imposing defensive harm on individuals who threaten to impose unjust harm on others, but who possess an authority-

based justification for doing so. I argue that, in such cases, the fact that an agent is justified in threatening harm does not increase the justificatory burden for inflicting defensive harm on that agent. This argument provides a novel counter-example to a common view within the literature on the ethics of self-defence, which holds that there can be no justified defence against a justified threat. I argue that while this may be true of other forms of justification, it is not true of authority-based justifications. The reason for this, I contend, is that a subject's reasons to obey authorities are fully agent-relative, applying only to the subject of the command while leaving the normative situation of others untouched.

However, while I deny that an agent's possession of an authority-based justification *itself* counts against imposing defensive harm on them, this does not entail that the commands of authorities cannot raise the justificatory burden for acts of defensive harm. As I explain in the final section of the chapter, the same argument that supports the idea of authority-based justifications also supports the converse notion of an *authority-based constraint*.

The Authority View of Harm provides a fairly general and abstract account of how judgements about the morality of causing harm are sensitive to considerations of authority. In the fifth and final chapter I apply the Authority View to the specific case of warfare and draw out its implications for just war theorising.

According to the resulting 'Authority View of War', individuals are morally justified in obeying the commands of an authority on matters of warfare when, and to the extent that, by doing so the individual better conforms to the reasons that govern the distribution of harm *by means of war*, compared to not obeying that authority. When these conditions are met, authorities acquire the moral power to give their subjects decisive reasons for action within the domain of warfare. Importantly, this includes actions that involve causing serious harm to others.

As I will explain, this model of legitimate war-waging authority enables a qualified defence of the authority criterion. Firstly, the concept of an authority-based justification provides the resources for rehabilitating the

permissive interpretation of the authority criterion. On my view, individuals who fight on behalf of legitimate authorities may be morally justified in obeying commands to cause harm in war, even when doing so would not be justified on the basis of the authority-independent reasons, such as liability and lesser-evil. The most salient implication of this view is that individuals may be justified in participating in unjust wars in virtue of being commanded to do so by an authority.

Secondly, the notion of an authority-based constraint enables a partial defence of the restrictive interpretation of the authority criterion. On my view, legitimate authorities possess the moral power to give their subjects undefeated reasons to refrain from participating in wars, including in wars where participation would be morally justified on the basis of the authority-independent considerations.

With this defence in place I go on to compare and contrast the Authority View with a reductivist approach to the ethics of war. I argue that while the two positions come apart in important respects – over the permissibility of participation in unjust wars most obviously – the Authority View is broadly compatible with the basic commitments underlying reductivism, such as its commitment to the moral primacy of individuals and its rejection of the idea that war (or political action more generally) is morally *sui generis*.

In the remainder of the chapter I clarify and refine the Authority View in light of two objections that may be pressed against it. The first objection holds that the Authority View, even if correct, has very little practical relevance, since no real-world authorities will in fact satisfy the conditions it requires for legitimacy regarding warfare. In response, I argue that authority over warfare can be justified from much more modest assumptions about the character and competence of existing political authorities than one might think. This is because authority over warfare need not be all-or-nothing. Rather, an authority may be legitimate regarding certain aspects of warfare and not others, depending on the facts of the case.

A second objection argues that the Authority View, when applied to war, falls foul of the very same objection that I pressed against non-

reductivist views in Chapter 3 – that it morally licences the widespread intentional killing of non-combatants. The objection draws its force from the fact that both the Authority View and non-reductivist offer agent-relative forms of justification for killing in war. In response, I argue that there is a subtle, but important, difference between the forms of agent-relativity that underpin each view, which shows that that Authority View is not susceptible to the objection to nearly the extent that non-reductivism is. However, I concede that there will be special cases in which the Authority View does justify the targeting of non-labile non-combatants, but I argue that this conclusion is defensible in these cases.

# Chapter 1 – The Role of the Authority Criterion

## 1.1 Introduction

The dominant intellectual framework for morally evaluating warfare – the just war tradition – has, since its earliest incarnations, included a requirement that war be initiated and waged by an entity with the authority to do so. However, while recent years have witnessed a huge rise in philosophical interest in the ethics of war, the authority criterion is largely absent from contemporary discussions. In this opening chapter I aim to show that this is an oversight worth rectifying, by arguing that the authority criterion plays a much more significant role within just war theorising than is commonly supposed.

As standardly understood, the authority criterion provides a necessary condition for the justification of the resort to war, but has no bearing on questions of permissible conduct in war. By contrast, I argue for an alternative interpretation of the criterion, which attributes to it a fundamental role in assessing this latter question (Sections 1.2–1.6). This more expansive role is necessitated by the fact that orthodox accounts of permissible conduct in war are committed – either implicitly or explicitly – to the view that the norms of action in war are distinct from, and more permissive than, those that govern interpersonal harming in non-war contexts. I argue that, given this commitment, the authority criterion plays a crucial jurisdictional role within just war theory: it distinguishes the class of activities to which its war-specific norms apply from those to which they do not.

With this revised interpretation in place, I then demonstrate its advantages by applying it to the practical issue of armed conflicts that are initiated and fought by non-traditional belligerents (Section 1.7). While several theorists have recognised that this common feature of modern armed conflict poses a challenge to mainstream just war theory in general – and to the authority criterion in particular – I argue that existing discussions frequently misconstrue the nature of the challenge, since they

assume the standard interpretation of the authority requirement and its role within the theory. I then show that the revised interpretation provides a clearer account of both the challenge posed by non-traditional belligerency and the kind of response that it requires.

In the final part of the chapter (Section 1.8) I show that my arguments for the revised interpretation of the authority criterion generalise beyond contemporary orthodox accounts of just war theory. As I argue, the authority criterion plays a structurally similar role in classical just war theory – as espoused by Augustine and Aquinas – despite the fact that the classical view endorses a radically different conception of *jus in bello* compared to contemporary mainstream views. In doing so, I demonstrate an important continuity within the just war tradition that often goes unnoticed.

The arguments of this chapter are largely formal, in that they aim to show that given certain common views regarding permissible conduct in war, a more expansive account of the authority criterion and its role within just war theorising follows. I conclude (Section 1.9) by identifying the central normative question raised by my discussion of the authority criterion: Can the authority criterion, in either its standard or revised interpretations, be given a principled moral defence. This question will be addressed in the following four chapters of the thesis.

## **1.2 The Structure of Orthodox Just War Theory**

Just war theory ranges over two main fields of inquiry.<sup>2</sup> The first, known as *jus ad bellum*, specifies the conditions that must be satisfied for the resort to war to be morally justified. The second, known as *jus in bello*, addresses the moral permissibility of conduct in war by individual participants.

---

<sup>2</sup> I leave aside here the more recently emphasised fields of justice in ending wars (*jus ex bello*) and justice following war (*jus post bellum*). For the former, see Darell Mollendorf, 'Jus ex Bello', *Journal of Political Philosophy* 16, No. 2 (2008), 123-136; Cecile Fabre, 'War Exit', *Ethics* (forthcoming). For the latter, see Gary Bass, 'Jus Post Bellum', *Philosophy and Public Affairs* 32, No. 4 (2004), 384-412; Larry May, *After War Ends: A Philosophical Perspective*, (Cambridge: Cambridge University Press, 2012); Jeff McMahan, 'The Morality of Military Occupation', *Loyola International and Comparative Law Review* 31 (2009), 7-31; Brian Orend, 'Jus Post Bellum', *Journal of Social Philosophy* 31, No.1 (2000), 117-137.

Particular accounts of the content of *jus ad bellum* vary, but the following six criteria represent a reasonable consensus.<sup>3</sup> In order for the resort to war to be justified the following individually necessary and jointly sufficient conditions must be satisfied: the war must have a just cause; it must be fought with the right intentions; the harm caused by the war must be proportionate to the good achieved; it must be the last resort; it must have a reasonable prospect of success; lastly, it must be initiated and waged by a legitimate authority.

Discussion of *jus in bello* focuses on two main requirements that individual participants in war must satisfy in order to act permissibly. Firstly, they must discriminate between legitimate and illegitimate targets and attack only the former. The legitimate/illegitimate target distinction is typically held to track the distinction between combatants and non-combatants. This does not mean that the discrimination requirement prohibits causing harm to non-combatants in war, merely that deliberately targeting non-combatants is prohibited. Harming non-combatants collaterally, as a side-effect of attacking military targets, does not violate the discrimination requirement. Such harms are constrained by a second main requirement of *jus in bello*, which holds that harms caused in war must be proportionate to the military advantage gained by doing so.

The nature of the relationship between these two domains of just war theory is highly important. According to what has been termed the *orthodox* account of just war theory – as influentially defended by Michael Walzer and finding resonance in both the law of armed conflict and in folk judgements about war – *jus ad bellum* and *jus in bello* are “logically independent” of one another.<sup>4</sup> On this view, the question of whether a war satisfies the criteria of *jus ad bellum* has no bearing on whether individual participants in that war act permissibly or impermissibly when they fight in it.<sup>5</sup> This independence underpins a central aspect of orthodox accounts,

---

<sup>3</sup> For a useful survey and comparison of a range of different accounts of *jus ad bellum*, see Christopher Toner, ‘The Logical Structure of Just War Theory’, *The Journal of Ethics* 14, No.2 (2010), 81-102.

<sup>4</sup> Michael Walzer, *Just and Unjust Wars: A Moral Argument With Historical Illustrations*, 4<sup>th</sup> Edition (New York: Basic Books, 2006), p.21

<sup>5</sup> *Ibid.*

concerning how *in bello* permissions and prohibitions are distributed between opposing combatants in war. According to the orthodox view, *in bello* norms apply neutrally to all parties to a war and are equally satisfiable by all combatants, irrespective of the *ad bellum* status of the wars in which they fight. Term this the *Equality Thesis*.<sup>6</sup> According to the Equality Thesis, provided that combatants in war target only opposing combatants and proportion the harm they cause to military advantage, they do not act wrongly.

### 1.3 The Standard Interpretation of the Authority Criterion

This chapter, and thesis more generally, will centre on one of the above just war criteria – the requirement that a war be initiated and waged by a legitimate authority.<sup>7</sup> To begin, it is helpful to distinguish two questions we can ask about the authority criterion. Firstly, we can ask what normative consequences are meant to follow from the possession or non-possession of war-making authority. Term this the question of *normative effect*. Second, we can ask which properties an entity must possess in order to have war-making authority. Term this the question of *relevant properties*.

In terms of its normative effect, the authority criterion is standardly understood to function solely as an *ad bellum* requirement, providing a necessary condition for the resort to war to be morally justified, but having no bearing on the permissibility of conduct in war.<sup>8</sup> On this view, if a war is to be justified it must be initiated and waged by an entity that possesses a certain set of properties. If such properties are lacking, then wars fought by that entity are unjustified (though this *ad bellum* deficiency has no bearing on the permissibility of the actions of those fighting for that entity, given a commitment to the Equality Thesis).

---

<sup>6</sup> For an overview of the Equality Thesis and the debate surrounding it, see David Rodin and Henry Shue, 'Introduction' in David Rodin and Henry Shue (eds), *Just and Unjust Warriors: The Moral and Legal Status of Soldiers*, (Oxford: Oxford University Press, 2008), 1-18.

<sup>7</sup> The criterion is also variously referred to as the 'right', 'proper', 'competent', 'sovereign' or 'public' authority criterion in the literature.

<sup>8</sup> This interpretation is particularly reinforced on an orthodox approach to just war theory, on which these two components of the theory are independent of one another.



Christopher Toner usefully suggests that each component of *jus ad bellum* can be interpreted as tracking one of several distinct variables relevant to determining the permissibility of action. He points out that in order to morally evaluate any action, we need to know several things: what action has been done; why it was done; how it was done; and who did it. Applied to the particular act of resorting to war, Toner suggests that the *ad bellum* criteria of just cause, right intention, proportionality, success and last resort can be understood as pertaining to the 'what', 'why' and 'how' questions. The authority criterion, by contrast, addresses the distinct question of 'who' is initiating the war and directing its prosecution. The authority criterion thus stands apart from the other *ad bellum* criteria in an important respect. Whereas the others may be characterised as *external* conditions, pertaining to various morally relevant aspects of the interactions between belligerent parties, the authority criterion provides an *internal* constraint on the resort to war, focusing on the identity and character of a potential belligerent.<sup>9</sup>

The standard interpretation of the criterion holds that justified war-waging is an activity restricted to a certain kind of entity. But this still leaves open the question of what properties an entity must possess in order to produce this normative effect. Precise accounts of what war-waging authority consists in vary, but on a fairly standard view, a belligerent party must, in order to qualify, constitute a recognised state, or, less stringently, a non-state entity that shares the central features of statehood, such as being an organised political community with *de facto* control over territory and a population. A more precise view requires not only that a belligerent be of a certain state-like type, but also that the resort to war must be deliberated upon and declared through the official procedures of the state (or quasi-state entity) that are entrusted with this important task and which possess the legal authority to enact the various

---

<sup>9</sup> For discussion of the distinction between 'external' and 'internal' constraints on the resort to war, see Allen Buchanan, 'The Internal Legitimacy of Humanitarian Intervention', *Journal of Political Philosophy* 7, No.1 (1999), 71-87.

apparatus of war, such as raising the armed forces, instituting conscription, nationalising the means of production, etc.<sup>10</sup>

Restricting justified war-making to entities that possess the properties of statehood (or quasi-statehood) seems intuitively sensible and ethically advantageous in several respects. Most obviously, given that war is a terrible thing we have strong moral strong reasons to try to minimise its incidence. Limiting the class of entities that may permissibly engage in war seems an obvious way of contributing to this humanitarian goal by discouraging the initiation of war.<sup>11</sup> In addition, limiting justified war to entities with certain institutional procedures may have the advantages of making the resort to war less rash, providing a bulwark against private interests within a community gaining disproportionate influence, and may help ensure that the decision to go to war better accords with the will of the citizenry.<sup>12</sup>

However, perhaps as a result of its pragmatic appeal and legalistic flavour, the authority criterion receives much less attention in the contemporary just war literature compared to other elements of the theory.<sup>13</sup> It is far overshadowed by discussion of the *ad bellum* conditions

---

<sup>10</sup> See, for example, Nicholas Fotion, *War and Ethics: A New Just War Theory*, (London: Continuum, 2007), pp.18-20; Michael Quinlan and Charles Guthrie, *Just War*, (New York: Bloomsbury, 2007), p.13; Brian Orend, *Michael Walzer on War and Justice*, p.87; Richard J. Regan, *Just War: Principles and Cases*, (Washington DC: Catholic University of America Press, 1996), Ch.2; Henrik Syse and Helene Ingierd, 'What Constitutes a Legitimate Authority?', *Social Alternatives* 24, No.3 (2005), 11-16.

<sup>11</sup> Several theorists claim that this pragmatic humanitarian concern to reduce the incidence of war explains the inclusion of the authority criterion within the just war tradition. See, for example, Anthony Coates, *The Ethics of War*, (Manchester: Manchester University Press, 1997), p.125.

<sup>12</sup> See Syse and Helene, 'What Constitutes a Legitimate Authority?'

<sup>13</sup> This lacuna is emphasised most forcefully by Anthony Coates. Coates, *The Ethics of War*, p.123. It is also noted in: Cecile Fabre, 'Cosmopolitanism, Just War Theory and Legitimate Authority', *Ethics and International Affairs* 84, No. 5 (2008), 963-976; Anthony F. Lang Jr, Cian O'Driscoll and John Williams, 'Introduction' in *Just War: Authority, Tradition and Practice*, Anthony F. Lang Jr, Cian O'Driscoll and John Williams (eds), (Washington DC: Georgetown University Press, 2013), 1-19; Syse and Ingierd, 'What Constitutes a Legitimate Authority?', p.11; Anne Schwenkenbecher, 'Rethinking Legitimate Authority' in Fritz Alhoff, Nicholas G. Evans, and Adam Henschke (eds), *Routledge Handbook of Ethics and War* (Abingdon: Routledge), 161-170. Heather Wilson suggests the lack of discussion arises because contemporary just war theorists and international lawyers simply assume that only states may permissibly resort to war. "The idea that only a sovereign State may legitimately wage war seems a foregone conclusion in the twentieth century." Heather Wilson, *The Use of Force by National Liberation Movements*, (Oxford: Oxford University Press, 1988), p.15.

of just cause and proportionality, and the voluminous literature on practical questions of *jus in bello*. When the authority criterion is discussed, it is often treated as a rather bureaucratic addendum to the *jus ad bellum* 'checklist'.<sup>14</sup>

This contemporary neglect for the authority criterion should be surprising, given a historical perspective. For the founding fathers of the just war tradition, most notably Augustine and Aquinas, the authority criterion was of absolute central importance.<sup>15</sup> With some affinity with the classical view, it will be a central contention of this chapter that the authority criterion plays a much more important role within mainstream just war theory than is commonly assumed. More specifically, I will argue for a different account of the normative effects that follow from the possession and non-possession of war-making authority. Against the standard interpretation of the authority criterion – which holds that the authority criterion functions restrictively within *jus ad bellum* – I will show that the authority criterion also plays a crucial *permissive* role within the *jus in bello* component of the theory. An interesting implication of this interpretation is that it denies the strict independence of *jus in bello* from *jus ad bellum*.

---

<sup>14</sup> For example, several discussions the authority criterion focus on the legal question of the extent to which the President of the United States has the power to take the nation to war without congressional approval. Paul Christopher *The Ethics of War and Peace: An Introduction to Legal and Moral Issues*, 3<sup>rd</sup> Edition (New Jersey: Prentice Hall, 2004), pp.87-88; Orend, *Michael Walzer on War and Justice*, p. 97; Regan, *Just War*, Ch.2.

<sup>15</sup> This point is stressed in, among others, John Langan, 'The Elements of Augustine's Just War Theory', *Journal of Religious Ethics* 12, No. 1 (1984), 19-38; Stephen C. Neff, *War and the Law of Nations: A General History*, (Cambridge: Cambridge University Press, 2005), p.50; Frederick Russell, *The Just War in the Middle Ages*, (Cambridge: Cambridge University Press 1975), pp. 68-71; James Turner Johnson, 'Aquinas and Luther on Sovereign Authority', *Journal of Religious Ethics*, 31, No. 1 (2003), 3-20; The authority criterion, or something closely analogous to it, was also a preeminent component of Confucian approaches to the morality of warfare. For discussion, see Sumner B. Twiss and Jonathan Chan, 'The Classical Confucian Position on the Legitimate Use of Military Force', *Journal of Religious Ethics* 40, No. 3 (2007), 447-472.

## 1.4 The Discontinuity Thesis

The first step in this argument is generated by comparing an orthodox conception of *jus in bello* with our settled views regarding the permissibility of harming and killing in ordinary, non-war contexts.

Despite the pervasiveness of an orthodox conception of *jus in bello* in common-sense thought about war and in the law of armed conflict, there is an obvious and important question here concerning whether these norms can be given a principled moral grounding. After all, morality contains a strong presumption against the kind of actions typical of conduct in war – namely, intentional killing and maiming. The normal explanation of this presumption is that individuals possess basic rights against such harm, which impose stringent correlative duties on others to refrain from doing so. Yet orthodox just war theory permits homicide on a large scale.

An attractive strategy for grounding an orthodox conception of *jus in bello* norms would be to show that it can be derived from the same justifications for harming that we accept in cases of interpersonal harming outside of warfare.

However, as several theorists have recently pointed out, despite an orthodox view representing the default position on the morality of conduct in war, it is very difficult to ground these norms in our more general views about harming and killing in non-war contexts. In particular, the orthodox view's neutral distribution of the permission it grants to combatants to intentionally kill their opponents, independently of considerations of *jus ad bellum*, has proven stubbornly immune to such an explanation.<sup>16</sup> I lack the space here to provide a full overview of this important debate. Instead, by way of illustration, I will show how two popular attempts to support the orthodox conception of *jus in bello* in this way are unsuccessful.

A standard way of justifying intentional killing is by showing that the subject of harm lacks their normal right against being killed, thus removing

---

<sup>16</sup> Jeff McMahan has done most to argue for this view. Jeff McMahan, 'Innocence, Self-Defense and Killing in War', *Journal of Political Philosophy* 2, No. 3 (1994), 193-221; Jeff McMahan, 'The Ethics of Killing War', *Ethics* 114, No.4 (2004), 693-733; Jeff McMahan, *Killing in War*, (Oxford: Oxford University Press, 2009), Chs. 1-2.

the constraint against doing so. The notion of rights loss forms the standard justification for killing in self- and other-defence. The basic idea here is that an individual can lose her right against being harmed if harming her is instrumental to averting a threat for which that individual is appropriately responsible. Under these conditions, that individual is *liable* to defensive harm.

Michael Walzer influentially appeals to a liability-based justification to ground a neutrally distributed permission to kill in war.<sup>17</sup> On this view, by threatening their opponents with lethal harm, combatants thereby render themselves liable to their enemy's lethal defensive force. In virtue of their liability, combatants may be killed without their rights being violated or a corresponding duty breached. As Walzer puts it, combatants in war are all "dangerous men".<sup>18</sup> Since all combatants pose this danger irrespective of whether their wars are just or unjust, all forfeit their right not to be killed by their opponents. Hence, all combatants possess an "equal right to kill" their opponents in self-defence.<sup>19</sup>

However, the key problem with this neat defence of an orthodox permission to kill in war is that it relies on a conception of liability to defensive force that we would not accept in any circumstance other than war. Theories of self-defence typically treat mere threat-posing as insufficient for liability. This is for good reason. Potential murder victims, for example, do not render themselves liable to defensive force if they defend themselves against their attacker with proportionate and necessary harm. Yet a threat-based account of liability, which supports a neutral permission to kill in war, would seem to have precisely this implication. A plausible account of liability requires, at the very least, the imposition of a threat of *unjustified* harm. This revised criterion of liability yields the correct result in the murderer/victim case, since only the attempted murderer poses an unjustified threat.

However, this more plausible account of liability, when applied to the case of war, cannot yield the view that the permission to kill applies

---

<sup>17</sup> Walzer, *Just and Unjust Wars*, pp.144-145.

<sup>18</sup> *Ibid.* p.145.

<sup>19</sup> *Ibid.* p.41.

neutrally to all combatants. The reasoning here parallels that in the simple murderer/victim case, the thought being that any plausible account of liability, when applied to combatants in war, must be sensitive to whether the overall aims of the wars in which they fight are justified or unjustified. So, while combatants who fight in wars that lack a just cause may satisfy the revised criterion of liability to attack and therefore be killed without being wronged, those who fight in justified wars do not thereby become liable.

A different attempt to show that a neutral conception of permissible killing in war is rights-consistent appeals to consent-based justifications for harming. On this view, by voluntarily joining the military and participating in wars combatants thereby consent to the risk of being killed by their opponents and waive their right against such treatment, regardless of whether they fight in just or unjust wars. Given this mutual consent, combatants may kill their opponents without violating their rights, independently of the justice of their wars.<sup>20</sup>

There are, however, several problems with this argument. Firstly, it has limited scope, since it is unlikely that all, or even many, combatants join the military under the stringent conditions of voluntariness that are presumably required in order for one's act of waiving of one's right not to be killed to be morally efficacious. The most obvious cases are combatants who are conscripted or otherwise coerced into joining the military, but cases in which combatants enlist through a lack of better alternatives or through ignorance are also problematic.

Secondly, and more seriously, it is doubtful that the act of freely entering the military and accepting the associated risks entails waiving one's right not to be killed by one's opponents in war.<sup>21</sup> For example, imagine that Benny is due to testify in court against the Mafia in a week's time and, understandably, fears for his life. He explains his situation to Julie, who then consents to be Benny's bodyguard in full knowledge of the

---

<sup>20</sup> Versions of this argument have been put forward, independently, by Michael Walzer and Thomas Hurka. Walzer, *Just and Unjust Wars*, p.37; Thomas Hurka, 'Liability and Just Cause', *Ethics and International Affairs* 20 (2007), 199-218.

<sup>21</sup> This problem is raised by Jeff McMahan, *Killing in War*, p.52.

risks. If Benny is then attacked by a Mafia assassin and Julie bravely takes a bullet for him, it seems wrong to say that Julie is not wronged *by the assassin* when he shoots her, even though she consented to the risk. While Julie's act of consent may successfully alter her rights and duties *vis-à-vis* Benny, it does not affect the normative situation between her and the assassin.

Thirdly, even if we grant that a combatant's consenting to joining the military does successfully result in the loss of his right not to be killed, this still seems insufficient to render it all-things-considered permissible for his opponent to kill him. For if his opponent is fighting in a war which has unjust aims, she will still be promoting those aims by killing her just opponent, even if she does not violate his rights by doing so.<sup>22</sup> For these reasons, the appeal to consent also seems unable to ground a neutral distribution of *in bello* permissions.

This brief comparison between an orthodox account of *jus in bello* and our views about the justification of interpersonal harming in non-war contexts reveals a serious tension. Ordinarily, we think that the moral assessment of an activity's ultimate aims are directly relevant to determining the moral permissibility of its component actions, particularly so when these acts involve inflicting serious harms. Yet the orthodox view holds that this is not the case in war. The norms governing causing harm in war are more permissive than those that apply in other circumstances.

My aim here is not to take a stand on whether this result should be taken to count against an orthodox view.<sup>23</sup> Instead I am interested in what follows from a commitment to an orthodox view of *jus in bello*, once this tension is made overt, in terms of theorising about the authority criterion and its role within just war theory.

Given this aim, the important point here is that in order to avoid an obvious problem of inconsistency, orthodox just war theory requires a commitment to what I term the *Discontinuity Thesis*. According to the Discontinuity Thesis, the moral norms that govern killing in war are not

---

<sup>22</sup> Ibid, pp.56-57.

<sup>23</sup> This is discussed in Chapter 2.

grounded in the moral principles that determine the permissibility of causing harm in ordinary, non-war contexts.<sup>24</sup> On this ethically dualist view, there need be no conflict between the moral norms that govern each domain, since war and non-war fundamentally differ in some morally relevant respect. War is to some degree morally *sui generis*, irreducible to the principles that govern interpersonal harming in other areas of life.<sup>25</sup> This view of war and its relationship to other kinds of activity thus enables one to hold that the norms of conduct in war are more permissive than those that govern other forms of violence.

The idea that war is morally distinct from other walks of life is central to folk judgments regarding war and has a long historical pedigree. It is often claimed that war is governed by a 'public' morality, separate from the 'private' morality that governs interactions between individuals.<sup>26</sup> In his historical survey of the law of war, Stephen C. Neff points out that many of the ancient civilisations worshipped separate deities for war and mere violence, indicating the perceived distinctness of the two spheres. Neff also emphasises the pervasiveness of the distinction in language, noting the difference words used to describe interpersonal violent conflicts compared to the kind of enmity encountered in war, a trend which holds up well cross-culturally.<sup>27</sup>

---

<sup>24</sup> The Discontinuity Thesis is similar to a view Seth Lazar has labeled 'exceptionalism'. Seth Lazar, 'National Defence, Self-Defence, and the Problem of Political Aggression', in Seth Lazar and Cecile Fabre (eds), *The Morality of Defensive War* (Oxford: Oxford University Press, 2014), 11-39.

<sup>25</sup> Henry Shue and Michael Walzer each endorse a version of the discontinuity thesis. Henry Shue, 'Do We Need a 'Morality of War' in Rodin and Shue (eds), *Just and Unjust Warriors*, 87-111; Michael Walzer, 'Response to McMahan's Paper', *Philosophia* 34, No.1 (2006), 43-45. Both Judith Jarvis Thomson and Frances Kamm have also suggested versions of it, albeit without explicit endorsement. See Frances Kamm *The Moral Target*, (Oxford: Oxford University Press, 2012), p.27; Judith Jarvis Thomson, 'Self-Defense', *Philosophy and Public Affairs* 20, No.4 (1991), 283-310 at p.297.

<sup>26</sup> For a book length discussion of this contrast, see Andrew Fiala, *Public War, Private Conscience: The Ethics of Political Violence*, (London: Continuum, 2010).

<sup>27</sup> Neff, *War and the Law of Nations*, pp.13-20. Neff does note that a small minority of cultures, such as the Inuit, do seem to lack a strong distinction between war and other forms of violence.



## 1.5 The Demarcation Requirement

The next step the argument for a more expansive account of the role of the authority criterion proceeds by noting an important problem for orthodox just war theory, entailed by its commitment to the Discontinuity Thesis. If it is true that conduct in war and non-war contexts are governed by different and incompatible sets of norms, then it is vitally important that just war theory is able to clearly distinguish between the activities to which its norms apply and those to which it doesn't.<sup>28</sup> As George Fletcher puts it, "not every shootout at the OK Corral qualifies."<sup>29</sup> For example, just war theory needs to be able to distinguish between genuine wars and so-called 'gang wars'. While these two phenomena may share some important empirical features – large-scale, protracted and organised violence most obviously – orthodox just war theorists presumably do not want to conclude that participants in such conflicts possess the same permissions to cause harm that are granted to regular combatants under *jus in bello*, which apply independently of the justice of the overall aims of the conflict in which they participate. In cases of violence-other-than-war, just war theorists need to be able to argue that the standard, peacetime norms of interpersonal harming apply.<sup>30</sup> To put the point another way – since orthodox just war theorists are committed to the view that war is morally distinct from all other violent activities, they need to be able to know a war when they see one. Term this the *Demarcation Requirement*.<sup>31</sup>

Without a means of clearly resolving the Demarcation Requirement, acts of violence will have an indeterminate moral status under an orthodox approach to just war theory, since whether such acts are morally

---

<sup>28</sup> Jeff McMahan also raises this requirement for orthodox conceptions of just war theory in *Killing in War*, p.31.

<sup>29</sup> George Fletcher, *Romantics at War: Glory and Guilt in the Age of Terrorism*, (Princeton: Princeton University Press, 2002), p.3.

<sup>30</sup> This worry motivates Robert Goodin's argument that just war principles are inapplicable to the action of terrorist organisations. Robert E. Goodin, *What's Wrong With Terrorism?* (Malden, MA: Polity Press, 2006), Ch.1.

<sup>31</sup> David Luban raises an analogous problem regarding demarcating the boundary between the application of human rights law and the humanitarian law of armed conflict. David Luban 'War Crimes: The Laws of Hell' in Larry May (ed), *War: Essays in Political Philosophy* (Cambridge: Cambridge University Press, 2008), 266-288 at p.276-280.

permissible or impermissible will depend on whether or not they are committed within a state of war. In order to determine whether a token act of intentional killing is morally justified or unjustified, the orthodox just war theorist cannot appeal to *jus in bello* criteria, since these criteria are specific to war. The most that could be said is that ‘this act of killing would be permissible under the norms of *jus in bello*, if it is in fact the case that these norms apply to it’. But it is precisely the antecedent of this conditional that needs to be established in order to satisfy the Demarcation Requirement.

### **1.6 The Revised Interpretation of the Authority Criterion**

I have argued that orthodox approaches to just war theory rest on a commitment to the Discontinuity Thesis and must therefore resolve the Demarcation Requirement. These two points provide the basis for an alternative and more expansive account of the role of the authority criterion in orthodox just war theory.

Returning to the six standard *ad bellum* requirements set out at the beginning of this chapter we can see that only one of these is capable of providing orthodox just war theory with the resources to satisfy the Demarcation Requirement – the authority criterion.

Consider, first, the *ad bellum* criteria of just cause, right intention, proportionality, last resort and prospect of success. Each of these requirements plays a necessary role in making the central distinction within just war theory between wars that are morally justified and those that are not. However, these criteria cannot be appealed to in order to help just war theory make the more fundamental distinction between those activities that constitute war, subject to *jus in bello* norms, and those activities which do not. Unjustified wars are importantly still *wars* on an orthodox view, subject to precisely the same war-specific norms of conduct. Hence, to invoke one of these five criteria in order to make the distinction between wars and non-wars would lead to a very odd definition of war, in which an unjust war would no longer counts as a war in the relevant sense.

More importantly, doing so would dramatically undermine the orthodox commitment to the Equality Thesis, since it would make *jus in bello* permissions conditional on some degree of *jus ad bellum* adherence.

Having excluded these *ad bellum* criteria from being invoked in order to resolve the Demarcation Requirement, only the authority requirement remains. Recall the earlier characterisation of the authority criterion as an ‘internal’ constraint on the resort to war. Rather than focussing on the rights and wrongs of the interactions between belligerent parties, as the other *ad bellum* criteria do, the authority criterion inquires into whether a war is waged by the appropriate type of entity. It is this difference in evaluative focus that enables a more expansive interpretation of the authority criterion, which attributes to it the role of resolving the Demarcation Requirement.

According to the standard interpretation, the authority criterion provides a necessary condition for determining whether a pre-established war is justified or unjustified. On this view, the possession or non-possession of war-making authority has precisely the same normative effect that the satisfaction or non-satisfaction of any other *ad bellum* criterion does.

By contrast, on the revised interpretation that I propose, the authority criterion also fulfils the more fundamental role of determining which class of activities counts as war for the purpose of evaluation under the war-specific norms of just war theory. On this interpretation, the criterion serves not simply to identify the entities who are capable of waging *justified* wars, but to identify the entities which are capable of waging war *simpliciter*.<sup>32</sup> The idea here is that war – understood as the class of activities to which the norms of just war theory apply – is identified by “determining

---

<sup>32</sup> Helen Frowe makes a similar, though brief, observation about the possibility of this role for the authority condition in defining war. Helen Frowe, *The Ethics of War and Peace* (Abingdon: Routledge, 2011), p.59. However, Frowe does not provide an account of *why* it matters that war be defined in such terms. It is this missing element that I aim to have provided here.

the entities that are allowed to begin and engage in war.”<sup>33</sup> In a recent work, Mark Evans has suggested that we may understand satisfying the authority criterion as requiring the possession of “the right to wage war.”<sup>34</sup> We may adopt Evans’ terminology here and flesh out the idea of a right to wage war as consisting in an entity’s entitlement to have its organised violence, whether *ad bellum* justified or unjustified, evaluated under the norms of just war theory rather than the moral norms that govern other uses of force.<sup>35</sup>

This revised interpretation of the authority criterion allows orthodox just war theory to resolve the Demarcation Requirement. By defining war as an activity of a limited class of entities, orthodox theorists are able draw the necessary normative boundary between the violent activities that are covered by its norms and those that are not. War and its attendant norms are distinguished from other activities by distinguishing between the entities which are able to create a state of war from those who are not. The authority criterion thus has a much more expansive and fundamental role within orthodox just war theory than is commonly assumed under the standard interpretation. Rather than simply contributing to determining whether a candidate war is *ad bellum* justified, the criterion determines the logically prior question of whether the activity in question constitutes a war in the first place.<sup>36</sup>

On this revised reading, the authority criterion also importantly functions as an *in bello* criterion on an orthodox view. At first glance, it may be unclear why this is so. After all, one might think, *jus in bello* pertains to quite specific norms of conduct, such as the prohibition on targeting non-combatants. It is not obvious what the authority criterion has to do with

---

<sup>33</sup> Alexander Moseley, ‘The Philosophy of War’, *Internet Encyclopedia of Philosophy*, available at <http://www.iep.utm.edu/war/>. Moseley does not offer this statement as an interpretation of the role of the authority criterion.

<sup>34</sup> Mark Evans, ‘Moral Theory and the Idea of a Just War’, in Mark Evans (ed), *Just War Theory: A Reappraisal*, (Edinburgh: Edinburgh University Press, 2005), p.13.

<sup>35</sup> A similar point is also made, independently, by Cecile Fabre and Christopher Finlay. Cecile Fabre *Cosmopolitan War* (Oxford: Oxford University Press, 2012), p.160. Christopher Finlay ‘Legitimacy and Non-State Political Violence’, *Journal of Political Philosophy*, 18, No.3 (2010), 287-312.

<sup>36</sup> The logical priority of the authority criterion is also pointed out by Anthony Coates, who advances a structurally similar account of the criterion to my own. Coates, *The Ethics of War*, p.124.

these. We can appreciate the *in bello* function of the authority criterion by distinguishing between the *content* of *jus in bello* norms and their *jurisdiction*. By the *content* of these norms, I mean the substantive, first-order rules of conduct in war, such as the prohibition on targeting non-combatants. This is what most people are talking about when they refer to ‘the rules of war’. The *jurisdiction* of *jus in bello* norms, by contrast, refers to the question of the range of activities and persons to whom these norms are applicable. Given this distinction, we can see that the authority criterion plays the role of setting the jurisdictional scope of *jus in bello* norms. On this view, in order for acts of harming to fall within the realm of *jus in bello*, they must be performed by an individual fighting on behalf of an entity which satisfies the authority criterion. The key normative effect of the possession of war-making authority is that those individuals who act on behalf of such entities are subject to a more extensive range of permissions to cause harm than if they fought on behalf of an entity which lacks that authority. On the revised interpretation, the authority criterion thus plays an essentially *permissive* role.

An interesting implication of this argument is that the independence of *jus in bello* from *just ad bellum* cannot strictly be true, by the orthodox just war theorist’s own lights. This is because it is precisely a commitment to an orthodox view which necessitates an increased role for the authority criterion, to bridge the *ad bellum/in bello* divide in order to determine the jurisdiction of *jus in bello*.<sup>37</sup>

With the revised interpretation of the authority criterion in place, we can appreciate the moral significance of philosophical definitions of war, such as Brian Orend’s, which make explicit reference to the character of the entities engaged in violence, in comparison to the definitions employed by empirical social-scientists, which focus instead on the quantitative intensity of violence.<sup>38</sup> According to Orend, “War is a phenomenon which occurs

---

<sup>37</sup> As Christopher Finlay nicely puts it, the authority criterion “silently” links the two fields. Finlay, ‘Legitimacy and Non-State Political Violence’, p.300.

<sup>38</sup> For example, the Correlates of War Project, which provides a major source of data for social scientific research on war and violence, classifies wars as sustained combat, involving organised forces, resulting in at least one thousand combat deaths per twelve-month period. See, [http://www.correlatesofwar.org/COW2%20Data/WarData\\_NEW/](http://www.correlatesofwar.org/COW2%20Data/WarData_NEW/)

only between political communities, defined as those which either are states or intend to become states.”<sup>39</sup> In light of the preceding argument, we can interpret such definitions as offering a fairly standard account of the relevant properties that an entity must possess in order to have war-making authority, understanding the normative effects that follow from the possession of such authority in the way that I have claimed. The idea being that the more extensive permissions to cause harm contained within *jus in bello* apply only to individuals who fight on behalf of states or entities which aspire to statehood.

### 1.7 The (Real) Challenge of Irregular Belligerency

My arguments in support of the revised interpretation have so far proceeded at a fairly abstract level. In this section I demonstrate its strengths by considering a common feature of contemporary armed conflicts – that they are often fought by ‘irregular’ belligerents. This term refers to groups engaging in armed conflict that are not recognised states. Examples include ethnic or religious groups, revolutionary and secessionist movements, factions within a civil war, ‘warlords’, terrorist groups, and armed criminal gangs.

While several theorists have pointed out that this empirical phenomenon poses an important challenge to just war theory<sup>40</sup> – and to the

---

COW%20Website%20-%20Typology%20of%20war.pdf. For an accessible discussion of the different methods of quantitatively analysing armed conflict, see Stephen Pinker, *The Better Angels of Our Nature: A History of Violence and Humanity*, (London: Penguin Books, 2011), pp.241-267 and pp.358-361.

<sup>39</sup> Brian Orend, ‘War’ in *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), Edward N. Zalta (ed), available at: <http://plato.stanford.edu/archives/fall2008/entries/war/>

<sup>40</sup> See, for example, Endre Begby, Gregory Reichberg and Henrik Syse, ‘The Ethics of War. Part II: Contemporary Authors and Issues’, *Philosophy Compass* 7, No.5 (2012), 328-347 at p.331. More broadly, proponents of the ‘new wars’ thesis, such as Mary Kaldor, argue that the diversity of actors involved in modern armed conflict also poses deep challenges to normative international political theory more generally. Mary Kaldor, *New and Old Wars: Organised Violence in a Global Era*, 2<sup>nd</sup> Edition, (Cambridge: Polity Press, 2007). For arguments disputing the ‘newness’ of the new wars thesis, see Mats Berdal, ‘The ‘New Wars’ Thesis Revisited’ in Huw Strachan and Sibylle Scheipers (eds), *The Changing Character of War* (Oxford: Oxford University Press, 2011), 109-133.

authority criterion in particular<sup>41</sup> – I will argue that contemporary discussions frequently misconstrue the precise nature of this challenge, because they assume the standard interpretation of the criterion. I then demonstrate that the revised interpretation of the criterion provides a clearer account of the challenge posed by irregular belligerency and the kind of response that it requires.

While my arguments are intended to apply to all accounts of just war theory that endorse both: (i) an orthodox conception of *jus in bello* norms and (ii) the standard interpretation of the authority criterion, my discussion here will focus on Nicholas Fotion's recent work.<sup>42</sup> This is worthy of particular attention, since Fotion endorses both these views, while addressing the issue of irregular belligerency head on.<sup>43</sup>

Fotion rightly points out that irregular belligerency poses a particular problem for the requirement that a war be initiated and directed by an appropriate authority, which Fotion interprets to mean that a war must be initiated through the appropriate legal institutions of the state. The challenge, as Fotion sets it up, is that this requirement seems to heavily bias just war theory against irregular belligerents, since such groups will often lack the relevant properties. As he puts it, such groups "may have leaders, even charismatic ones, but these leaders do not have legitimate authority."<sup>44</sup> Taking a standard view of the normative effect that follows from non-possession of war-making authority, Fotion's worry is that just war theory will render the vast majority of potential wars fought by

---

<sup>41</sup> Eric A. Heinze and Brent J. Steele, 'Introduction: Non-State Actors and the Just War Tradition', in Eric A. Heinze and Brent J. Steele (eds), *Ethics, Authority and War* (Basingstoke: Palgrave MacMillan, 2009), 1-20; Coates, *The Ethics of War*, pp.125-126; Frowe, *The Ethics of War and Peace*, p.191.

<sup>42</sup> Nicholas Fotion, *War and Ethics*, Ch.9. See also, Nicholas Fotion, "Two Theories of Just War", *Philosophia* 34 (2006), 53-64.

<sup>43</sup> Fotion does not explicitly endorse the orthodox conception of *jus in bello*. However, a commitment to an orthodox view can be inferred from what he does say on the topic. For example, he writes "it is easy to separate that part of a theory of exception having to do with starting a war (justice of the war) and the protracted period of time that follows once the war is started (justice in the war)." *War and Ethics*, p.21. Fotion also endorses conceptions of discrimination and proportionality in war that are equally satisfiable by combatants who participate in just and unjust wars. Fotion, *War and Ethics*, pp.21-22. Thus, taken together, it seems fair to attribute to Fotion a commitment to an orthodox conception of *jus in bello*.

<sup>44</sup> Fotion, *War and Ethics*, p.121.

irregular belligerents as unjustified *a priori*. This statist bias is unattractive, since presumably *some* wars fought by irregular belligerents could in principle be justified.<sup>45</sup> We may term this the *Standard Challenge* of irregular belligerency – since it corresponds to the standard interpretation of the authority criterion.

In light of this challenge, Fotion offers a novel revision to just war theory, by dividing it into two parts – ‘Regular Just War Theory’ (JWT-R), which applies to traditional state belligerents, and ‘Irregular Just War Theory’ (JWT-I), which applies to non-state actors engaging in organised violence. While JWT-R retains all the standard criteria of *jus ad bellum*, Fotion boldly jettisons the authority criterion from JWT-I.<sup>46</sup> On this revised theory of just war, the authority criterion functions only as a conditional requirement, in that the criterion need only be satisfied *if* it can in fact be satisfied. As Fotion puts it, “Since they [irregular belligerents] are unable to satisfy the principle, even in theory, JWT-I does not ask them to do so.”<sup>47</sup> This ingenious twin version of just war theory is designed to bring the theory up-to-date, removing its unpalatable statist bias.

However, there is a deep problem with this response to the challenge of irregular belligerency, arising from Fotion’s jettisoning of the authority criterion while maintaining a commitment to an orthodox conception of *jus in bello*. As argued above, such a conception entails a commitment to the Discontinuity Thesis, which in turn gives rise to the Demarcation Requirement. The problem, for Fotion, is that by jettisoning the authority criterion he denies himself the resources to satisfy the Demarcation Requirement, which generates some highly implausible results.

The problem can be drawn out by considering the following example.

---

<sup>45</sup> For a classic critique of just war theory on the grounds of its statism, see David Luban, ‘Just War and Human Rights’, *Philosophy and Public Affairs* 9, No.2 (1980), 160-181.

<sup>46</sup> Fotion also relaxes the criterion of reasonable prospect of success in JWT-I. *War and Ethics*, p.120

<sup>47</sup> Fotion, *War and Ethics*, p.121.



**The Zepos:** The Zepos are a large and well-organised criminal organisation. In order to increase their criminal empires, the Zepos declare themselves to be at war with the state in which they operate and which threatens to hamper their activities. The Zepos then conduct a campaign of lethal violence in pursuit of this aim. The Zepos armed fighters target only military personal and do not use any force beyond that required for achieving their goals.

Imagine that when confronted with the accusation that they are acting seriously morally wrongly by fighting and killing in pursuit of such aims, individual Zepo fighters offer the following reply – “It may well be true that our aims are unjust. However, this injustice is a matter of *jus ad bellum*. Given that the norms of *jus in bello* obtain independently of considerations of *jus ad bellum*, we cannot be deemed to be acting wrongly by reference to such aims. Our action can only be evaluated by the standards of *jus in bello*, which we followed to the letter.”

Clearly, something has gone amiss with this argument, since it is – I take it – uncontroversial that Zepo fighters *are* acting morally wrongly by targeting and killing their victims, even if they are members of the military. The obvious rejoinder to their claims is to point out that the norms of *jus in bello* are not applicable to the violent activity which the Zepos are engaged in. The underlying thought here is that groups such as the Zepos lack the ability to create a genuine state of war, and hence their violence falls outside the jurisdiction of orthodox just war theory and its war-specific norms of conduct. Once this argument is made, the claims of individual Zepo members to moral and legal immunity for their acts of killing can be rejected. Their claims are not mistaken because they have misunderstood what the norms of *jus in bello* are, but because they have misunderstood what a state of war is and who is able to bring one into existence.

As argued above, this crucial idea that war and non-war are to be distinguished by reference to facts about the parties engaging in violence is captured by adopting a more expansive conception of the role of the authority criterion. The revised interpretation of the criterion allows

orthodox just war theory to make the argument that violence such as the Zepos' falls outside the jurisdiction of just war theory, thus resolving the Demarcation Requirement. The thought is that Zepos fighters are not morally permitted to kill opposing military personnel because they do not fight on behalf of an entity that possesses war-making authority.

The problem for Fotion's twin just war theory is that it seems unable to make this necessary argument, since it does not include an authority criterion within JWT-I, the part of his theory which applies to armed conflicts waged by irregular belligerents, such as the Zepos. The troubling implication of this is that JWT-I lacks the resources to explain why Zepo fighters act morally wrongly. The most that can be said under JWT-I is that the Zepos clearly lack a just cause, but all that can be concluded from this is that the Zepos are fighting an unjust *war*, which is precisely the concession that the orthodox just war theorist needs to reject.

Fotion does consider the possibility of a demarcation problem arising for his twin theory. Fotion's concern is that the theory may suffer from vagueness when applied to particular cases, so that it may be difficult to distinguish between groups that are capable of satisfying the authority criterion and those which are not. The result of this is that we may not know whether JWT-R or JWT-I applies to the particular case, the worry being that just wars may be wrongly classified as unjustified, and *vice versa*. After considering the problem, Fotion concludes that, "In the end the dual-theory approach to Just War Theory seems viable. It helps keep us clear about what kind of war we are engaged in, it doesn't suffer from so much vagueness as it seemed it did at the outset and although it perhaps triggers more wars than the classic theory, those additional wars are few in number."<sup>48</sup>

However, this misses the deeper problem of demarcation facing Fotion's twin theory. What should be of concern here is not that it may struggle to tell us "what kind of war" we are dealing with, and hence which version of just war theory applies. Rather, it is that it lacks the resources to explain why, in the vast majority of cases of violent activity, *no* version of

---

<sup>48</sup> Fotion, 'Two Theories of Just War', pp.60-61.

just war theory applies. This is because, as mentioned, JWT-I lacks the component of the theory capable of performing the vital role of distinguishing the class of activities that fall within its jurisdiction from those that do not.

The root of the problem for Fotion's twin theory is that it misconstrues the nature of the challenge posed by irregular belligerency, because it assumes the standard interpretation of the authority criterion and the normative effects that follow from its possession and non-possession. If my arguments motivating a revised interpretation of the authority criterion are successful, we can see that the real challenge posed by irregular belligerency is not – *contra* the standard challenge – that it makes it more difficult to distinguish between *ad bellum* just and unjust wars. Rather, irregular belligerency makes it far more difficult for just war theory to cleanly distinguish those activities to which its norms apply from those to which they do not.<sup>49</sup> Term this the *Revised Challenge*. Importantly, the revised challenge is essentially a problem of *jus in bello*, not *jus ad bellum*. It is a problem about how to establish the precise scope of *jus in bello* norms, given that reflection on irregular belligerency suggests that legitimate war making is not solely the preserve of easily identifiable states. Fotion's novel solution to the standard challenge, which jettisons the problematic *ad bellum* authority criterion, thus undercuts the ability of his twin-theory to respond to the deeper revised challenge. The revised challenge reveals that orthodox just war theory cannot jettison the authority criterion, since the theory relies on the authority criterion for its coherence.

As mentioned at the outset of this section, the problem that I have pressed against Fotion's twin theory is intended to have wider significance. The very same demarcation problem arises on any orthodox account that assumes the standard interpretation of the authority criterion. The problem is just more easily diagnosable in Fotion's account. On any theory which includes an authority requirement, but treats it as solely an *ad*

---

<sup>49</sup> For a discussion of this problem in terms of applying the law of war, see Rosa Ehrenreich Brooks, 'War Everywhere: Rights, National Security Law, and the Law of Armed Conflict in an Age of Terror', *University of Pennsylvania Law Review* 153, No.2 (2004), 675-761.

*bellum* criterion, the fact that a particular group initiating a campaign of large-scale violence fails to satisfy it tells us nothing about whether or not that group's agents fall within the jurisdiction of the norms of *jus in bello* or not.

To demonstrate, let us return, one last time, to the case of the Zepos. The example was designed to show that an account of just war theory such as Fotion's JWT-I, which endorses an orthodox conception of *jus in bello* but does not contain an authority criterion, is incapable of explaining why individual Zepo fighters act morally wrongly. The very same problem arises for any orthodox account that assumes the standard interpretation of the authority criterion. When applied to cases such as the Zepos, these accounts are (unlike Fotion's) able to conclude that the Zepos fail to possess legitimate war-making authority. However, given that such accounts treat the authority criterion as solely an *ad bellum* requirement, all that follows from this failure is that a belligerent's armed conflict is *ad bellum* unjustified. Making this point does not provide the theory with any resources for denying the Zepo's claim to evaluation under the norms of just war theory. In order to do so, orthodox just war theory must endorse the revised interpretation.

In this section I have sought to reveal the theoretical significance of the practical issue of irregular belligerency. Once the challenge of irregular belligerency is recognised a challenge regarding the precise jurisdictional scope of *jus in bello* norms, we can see that the authority requirement cannot be restricted to the *ad bellum* domain, nor jettisoned in the difficult cases raised by irregular belligerents. To address the challenge of irregular belligerency, orthodox just theorists must fulfil two important tasks.

Firstly, a much more precise account of the relevant properties that give rise to the possession of war-making authority (in its revised role) is required, which tells us more than the vague intuition that agents who use force on behalf of established states fall within it, while agents who fight on behalf of criminal gangs such as the Zepos clearly do not. The challenge of irregular belligerency requires orthodox just war theorists to venture into the murky spectrum of cases between these two poles and provide criteria

for clearly demarcating war from other violent activities, and to do so by reference to properties of the of entities involved.

Secondly, and most importantly, in order to avoid a charge of arbitrariness, such an account of war-making authority not only has to provide an intuitively acceptable account of the scope of just war theory and its war-specific norms of conduct, it also has to explain *why* members of certain types of entity are entitled to evaluation under those distinct, and more permissive, moral norms. If the norms of conduct in war apply only to those who fight on behalf of entities that possess a certain set of properties, it needs to be explained what it is about those properties which give rise to additional permissions to cause harm.

We can see then that the practical issue of irregular belligerency has great significance for theoretical debates within just war theory regarding the relationship between war and other spheres of action. The plausibility of the view that war is morally discontinuous will depend on whether its proponents can provide a satisfactory response to the challenge of irregular belligerency. A failure to do so will provide strong support for critics of the orthodox view, who argue that the correct account of the morality of war must be consistent with our settled moral views about the morality of killing and injuring in all other contexts.

### **1.8 Classical Just War Theory and the Authority Criterion**

I have argued that endorsing an orthodox conception of *jus in bello* entails a commitment to a more expansive conception of the role of authority within just war theorising and the normative effects that follow from its possession or non-possession. In this final section, I demonstrate that the argument I have offered for this conclusion generalises to other approaches to just war theory which do not endorse this specific conception of *jus in bello*. In particular, I show that the authority criterion also plays the same expansive role within classical just war theory, as influential formulated in the writings of Catholic theologians such as Augustine and Aquinas, despite the important differences between the classical and orthodox views.

As I have argued, the importance of the authority criterion arises from a commitment to the view that the norms governing conduct in war are distinct from and irreducible to the norms that govern the use of violence in other contexts. As explained above, within orthodox just war theory, it is the commitment to the Equality Thesis that necessitates such a view, since the notion of moral equality is generally absent in cases of justified harming outside the context of war.

Given the ubiquity of the orthodox view, it is tempting to assume that this conception of just conduct in war has a long and distinguished pedigree. However, this is not the case. The Equality Thesis is in fact a relatively recent development, only becoming established in the eighteenth and nineteenth centuries with the rise of international law.<sup>50</sup> By contrast, classical just war theorists did not endorse the Equality Thesis. For these thinkers, only combatants who fought in wars with a just cause were permitted to attack and kill their opponents. Those who fought in unjust wars committed a serious moral wrong by doing so.<sup>51</sup>

Interestingly, despite rejecting the Equality Thesis, classical just war theorists did not attempt to ground the permission to fight and kill in just wars in the ordinary moral permissions to use force available to individuals, such as self-defence. As Gregory Reichberg explains, on a classical view,

It was...understood that public war should be waged and its morality judged by reference to a set of norms that are not directly reducible to those governing private self (and other)-defense.<sup>52</sup>

---

<sup>50</sup> Gregory Reichberg, 'The Moral Equality of Combatants – A Doctrine in Classical Just War Theory? A Response to Graham Parsons', *Journal of Military Ethics* 12, No.2 (2013), 181-194.

<sup>51</sup>Ibid, at 181-185. This point is also stressed in, among others, McMahan, 'War'; Neff, 'War and the Law of Nations', p.63; Gregory Reichberg, 'Just War and Regular War: Competing Paradigms' in Rodin and Shue (eds), *Just and Unjust Warriors*, 193-214; Uwe Steinhoff, 'Rights, Liability and the Moral Equality of Combatants', *Journal of Ethics* 16, No.4 (2012), 339-366.

<sup>52</sup> Reichberg, 'The Moral Equality of Combatants – A Doctrine in Classical Just War Theory? A Response to Graham Parsons', p.182.

In the terminology introduced earlier, we can say that classical just war theory was also committed to the Discontinuity Thesis. The reason for this is the classical theorists endorsed a much more restrictive account of permissible homicide than that generally accepted today. In particular, they took a much more restrictive view of permissible self-defensive killing.<sup>53</sup> On the most extreme view, such killing was straightforwardly morally wrong.<sup>54</sup> On a more moderate view, defensive killing may have been permissible in a weak sense, but only under extreme and unusual circumstances, where the agent's 'back was against the wall', so that the act of killing could be construed as an unintended and instinctive lashing out.<sup>55</sup> I use the term 'permissible in a weak sense' because, even in these cases, self-defence was viewed more as an excuse than a justification, in contrast to contemporary practice.<sup>56</sup> Given such a conception of permissible homicide, it is clear why the classical just war theorists did not attempt to ground the permission to participate in just wars in the ordinary norms of interpersonal morality. Such an attempt could only yield pacifism as a result, since these norms would not permit individuals to engage in the premeditated and organised acts of killing typical of war.

So, despite their very different substantive accounts of permissible conduct in war, both classical and orthodox just war theory share a common commitment to the Discontinuity Thesis. Once this is made overt, we can see the central importance of the authority criterion for classical

---

<sup>53</sup> See Oliver O'Donovan, *The Just War Revisited* (Cambridge: Cambridge University Press, 2003), p.22.

<sup>54</sup> This appears to have been Augustine's view, as well as that of St. Ambrose. See Lackey, *The Ethics of War and Peace*, p.17; Langan, 'The Elements of St. Augustine's Just War Theory', p.27; Louis J. Swift, 'St. Ambrose on Violence and War', *Transactions and Proceedings of the American Philological Society* 101 (1970), 533-543 at p.537.

<sup>55</sup> See Neff, *War and the Law of Nations*, pp.59-62; Reichberg, 'The Moral Equality of Combatants - A Doctrine in Classical Just War Theory? A Response to Graham Parsons' p.189; Michael Thompson, 'Aquinas and Locke on Self-Defense', *University of Pittsburgh Law Review* 57 (1996), 677-684.

<sup>56</sup> Thompson, 'Aquinas and Locke on Self-Defense', at pp.677-679. George Fletcher points out that a successful invocation of the medieval common law defence of *se defendendo* resulted only in the defendant's avoidance of punishment by execution, while conceding the wrongfulness of the act. The successful defendant was still required to give up property to the crown by way of recompense for the homicide. Fletcher notes that, "Until the Statute of Henry VIII, passed in 1532...there was no theory of self-defense that rendered a killing fully lawful, justifiable and therefore free of the taint that affected excusable homicide." George Fletcher, 'Defensive Force as an Act of Rescue', *Social Philosophy and Policy* 7, No. 2 (1990), 170-179 at p.171.

just war theorists, for the same reasons that I have argued regarding orthodox just war theory. For both the classical and orthodox just war theorist, the domains of war and non-war require demarcation, and this role is performed by the authority criterion. On each view, the normative effects that follow from the possession and non-possession of war-making authority are the same – those who use force on behalf of war-making authorities are subject to a more extensive set of permissions to inflict lethal harm, whereas those who use force privately (or on behalf of entities which also lack the relevant authority) are covered only by the more minimal permission (or perhaps only excuse) of self-defence.<sup>57</sup> As Reichberg notes, for the classical theorists,

Chief among the factors separating the two sorts of conflict was the role assigned to legitimate authority – war requiring it, and self-defense not...Those called to war (i.e. *bellum* in the proper, public denotation of the term) and serve in it, are bound by a norm that has little applicability within the sphere of private self-defense.<sup>58</sup>

This coheres nicely with the arguments that I have offered for the necessity of a more expansive conception of the authority criterion within orthodox just war theory. For both orthodox and classical just war theory, the authority criterion plays an important *in bello* role, determining which individuals are covered by war specific permissions to cause harm and

---

<sup>57</sup> Michael Thomson makes a similar point, arguing that “premodern practice,...restricted justified killing to that legitimately performed by agents of the state.” Thomson, ‘Aquinas and Locke on Self-Defense’, p.678. This thought is endorsed by Elizabeth Anscombe, the most influential recent proponent of a classical conception of just war. “The right to attack with a view to killing is something that belongs only to rulers *and those they command to do it*”. G.E.M. Anscombe, ‘War and Murder’, in *The Collected Papers of G.E.M. Anscombe. Volume Three: Ethics Religion and Politics* (Oxford: Blackwell, 1981), 51-61 at p.53 (emphasis added).

<sup>58</sup> Reichberg, ‘The Moral Equality of Combatants – A Doctrine in Classical Just War Theory? A Response to Graham Parsons’, p.189. James Turner Johnson makes a similar point, noting that within classical just war theory, “There is a fundamental moral difference between the use of the sword by one in sovereign authority or on his behalf and the use of the sword by a private individual. The former may wage *bellum*, which is the use of the sword on behalf of the common good; the latter may not.” James Turner Johnson, ‘Aquinas and Luther on Sovereign Authority’, pp.9-10.



which individuals are not, by identifying the entities which are morally entitled to bring a state of war into existence from those which are not.<sup>59</sup> Despite their many differences, viewing classical and orthodox approaches from the perspective of the authority criterion reveals an important continuity in the just war tradition that often goes unnoticed.

This brief discussion of classical just war theory is intended to demonstrate the wider point that the argument I have offered in favour of the revised interpretation of the authority criterion is generalisable. The importance of the authority criterion arises from a general commitment to a particular conception of war and its relationship to other activities, and not on a particular substantive account of just conduct in war.

## 1.9 Conclusion

I have argued that the traditional requirement that a war be waged by a legitimate authority has much greater significance within just war theorising than is usually recognised. The normative effects that follow from the possession or non-possession of war-making authority are broader than commonly supposed. Whereas standard views treat the criterion as functioning restrictively – imposing a constraint on the justification of the resort to war – I aim to have shown that the criterion can also be understood as functioning permissively – expressing the core intuition that individuals who use force on behalf of certain kinds of entities are subject to a more extensive range of permissions to cause harm than if they acted privately.

The arguments offered here have are largely formal and interpretive. I have aimed to show that given certain common views about the morality of

---

<sup>59</sup> It is often argued that classical just war theory does not contain a particularly developed account of *jus in bello*, compared to modern practice. Most saliently, classical views did not contain a clear prohibition on attacking certain classes of person in war akin to the contemporary principle of discrimination. (See, for example, Langan, 'The Elements of St. Augustine's Just War Theory', p.31). While this is true, the arguments presented here suggest that classical theories do have an account of *jus in bello* in an important sense, in that they view conduct in war as governed by different norms to those that govern private action, and ground this locate in the fact that those who permissibly participate in wars do so on behalf of an entity with the authority to engage in the business of war-making.

participation in war, a more expansive conception of the role of the authority criterion follows. But I have not taken any stand on the question of whether these commitments are plausible. As we will see in the following chapter, many of the standard tenets of just war theory have recently come under sustained criticism, including important objections to the authority criterion, in both its standard restrictive role and the permissive *in bello* role that I have argued for.

## Chapter 2 – Scepticism Regarding the Authority Criterion

### 2.1 Introduction

In the previous chapter I argued that the authority criterion plays a much more expansive role within the just war tradition than is commonly supposed. According to the standard interpretation, the normative effect that follows from an entity's possessing the relevant authority-conferring properties is that it has the ability to fight justified wars. Entities lacking those properties may still fight wars, but those wars will be unjust by *ad bellum* standards. By contrast, on the revised interpretation, the normative effect that follows from an entity's possession of the relevant properties is that individuals who fight on behalf of that entity are subject to additional permissions to cause harm. The criterion thus performs both a restrictive and permissive function.

With this conceptual groundwork in place, in this chapter I turn to the normative question of whether the authority criterion is morally defensible, in either its restrictive or permissive roles. I begin exploring this question by setting out a recent and increasingly influential 'reductivist'<sup>60</sup> approach to the ethics of war, defended most prominently by Jeff McMahan, Cecile Fabre and Helen Frowe (Sections 2.2).<sup>61</sup> Reductivism has been

---

<sup>60</sup> The view I term 'reductivism' is also sometimes referred to as 'individualism'. See, for example, Helen Frowe, *The Ethics of War and Peace*, Ch.2. As I use the term, reductivism can be understood as synonymous with individualism. To my knowledge, the term was first coined in David Rodin, *War and Self-Defense* (Oxford: Oxford University Press, 2002), Ch.6.

<sup>61</sup> McMahan, 'Innocence, Self-Defense and Killing in War'; McMahan, 'The Ethics of Killing in War'; Jeff McMahan, 'War as Self-Defense', *Ethics and International Affairs* 18, No.1 (2004), 13-18; Jeff McMahan, *Killing in War*; Cecile Fabre, 'Guns, Food and Liability to Attack in War', *Ethics* 120, No.1 (2009), 36-63; Cecile Fabre, *Cosmopolitan War* (Oxford: Oxford University Press, 2012); Cecile Fabre, 'Interneccine War Killings', *Utilitas* 24, No.2 (2012), 214-236; Cecile Fabre, 'Cosmopolitanism and Wars of Self Defence' in Cecile Fabre and Seth Lazar (eds), *The Morality of Defensive War* (Oxford: Oxford University Press, 2014), 90-114; Helen Frowe, 'Self-Defence and the Principle of Non-Combatant Immunity', *Journal of Moral Philosophy* 8, No.4 (2011), 530-546; Helen Frowe, *Defensive Killing: An Essay on War and Self-Defence* (Oxford: Oxford University Press, forthcoming). For other statements of a reductivist view, see, Richard J. Arneson, 'Just Warfare Theory and Non-Combatant Immunity', *Cornell International Law Journal* 39 (2006), 663-668; David Rodin, 'The Moral Inequality of Combatants: Why *jus in bello* Asymmetry is Half-Right' in Rodin and Shue (eds), *Just and Unjust Warriors*, 44-68; C.A.J. Coady, 'The Status of Combatants' in Rodin and Shue (eds), *Just and Unjust Warriors*, 153-175; Andrew Altman and Christopher

subject to much discussion, but here I aim to highlight an important aspect of this view, which is that it provides the resources for powerful objections to the moral significance of the authority criterion in both its permissive and restrictive roles (Sections 2.3-2.4). In fact, as we shall see, reductivists argue that the criterion should be jettisoned from just war theorising altogether. Section 2.5 concludes.

## 2.2 Reductivism: An Overview

I begin by providing an overview of the reductivist approach to the ethics of war. While there are important differences between reductivist thinkers and theories, the view can be characterised by a shared commitment to two theses.<sup>62</sup>

### 2.2.1 *The Continuity Thesis*

The fundamental commitment of reductivism is that war is morally continuous with all other domains of action. The moral reasons that govern harm by means of war are exactly the same as those that govern all other cases of interpersonal harming. According to reductivism, “there is no special morality of war.”<sup>63</sup> There is one single moral domain and all activities fall within its jurisdiction. On this view, both the resort to war and conduct within war, when justified, are justified precisely because, and to the extent that, they can be justified under the same moral principles that

---

Heath Wellman, ‘From Humanitarian Intervention to Assassination: Human Rights and Political Violence’, *Ethics* 118, No.2 (2008), 228-257; Lionel McPherson, ‘Innocence and Responsibility in War’, *Canadian Journal of Philosophy* 34, No.4 (2004), 485-506; Gerhard Overland, ‘Killing Soldiers’, *Ethics and International Affairs* 20, No.4 (2006), 455-475; James Pattison, ‘When Is it Right to Fight? Just War Theory and the Individual-Centric Approach’, *Ethical Theory and Moral Practice* 16, No.1 (2013), 34-54; Uwe Steinhoff, *The Ethics of War and Terrorism* (Oxford: Oxford University Press, 2007); Uwe Steinhoff, ‘Jeff McMahan on the Moral Equality of Combatants’, *Journal of Political Philosophy* 16, No. 2 (2008), 220-226.

<sup>62</sup> In formulating the two theses discussed in this section I have drawn on Seth Lazar’s helpful discussion. Lazar, ‘National Defence, Self-Defence, and the Problem of Political Aggression’ in Fabre and Lazar (eds), *The Morality of Defensive War*, 11-39.

<sup>63</sup> Jeff McMahan, ‘War’ in David Estlund (ed), *The Oxford Handbook of Political Philosophy*, (Oxford: Oxford University Press, 2012), 298-318 at p.309.

justify interpersonal harming in all other circumstances. Term this the *Continuity Thesis*.

Jonathan Glover provides an early statement of this view:

It is widely held that killing in war is quite different. It is not, and we need to think about the implications of this...apart from important special side-effects, killing in war is morally on a par with other killing. Declarations of war, military uniforms and solemn utterances by national leaders in no way reduce the burden of justification for an act of killing.<sup>64</sup>

The Continuity Thesis is not only endorsed by reductivist just war theorists. It also finds support among theorists of political and legal ethics more generally. For example, discussing the use of force by public officials in general, John Gardner also endorses a view analogous to the Continuity Thesis (which he terms “the unity view”<sup>65</sup>):

the differences between the duties of police officers and of other people, in other roles, are ordinary moral differences. Although police officers as such are indeed in special moral positions, there is no distinct ‘political’ morality applicable to them that displaces ordinary moral judgement. Morality is just morality, and it applies to all people. It applies to public officials (judges, soldiers, parliamentarians, police officers, local authority librarians, etc.) because they are people. They do not stop being people and hence do not stop being bound by morality when they put on their uniforms, or otherwise go on duty.<sup>66</sup>

---

<sup>64</sup> Jonathan Glover, *Causing Death and Saving Lives* (London: Penguin, 1977), pp.251-252

<sup>65</sup> John Gardner, ‘Criminals in Uniform’, in Anthony Duff, Lindsey Farmer, Sandra Marshall, Massimo Renzo and Victor Tadros (eds), *The Constitution of Criminal Law* (Oxford, Oxford University Press, 2012), 97-118 at p.118

<sup>66</sup> Gardner, ‘Criminals in Uniform’, p.116. A similar view is endorsed by Kimberley Brownlee, ‘Responsibilities of Criminal Justice Officials’, *Journal of Applied Philosophy* 27,

Two further features of reductivism are worth highlighting. Firstly, it is *individualist* in that it takes individuals to be the fundamental unit of moral concern and evaluation. The morality of war should be wholly assessed in terms of the principles of interpersonal morality that govern the use of force between agents *qua* individuals. Secondly, and relatedly, reductivism is *aggregative*, in that the war is to be morally analysed as a composite of all the individual actions that make it up. On this view, the only morally relevant difference between war and ordinary cases of interpersonal violence is that war takes place on a larger and more complex scale.

### 2.2.2 *The Content Thesis*

While central, the Continuity Thesis is a purely formal component of reductivism, providing an adequacy condition that any acceptable account of the morality of war must satisfy. It simply holds that whatever the moral reasons are that determine the permissibility of individuals killing and injuring others in war, they are precisely the same reasons that determine the permissibility of killing and injuring in all other circumstances. But this leaves it entirely open what these justifications consist in. As Seth Lazar points out, one can perfectly coherently combine any substantive account of the morality of interpersonal harming, such as an act-consequentialist account, with a commitment to the Continuity Thesis.<sup>67</sup> Indeed, Glover frames his endorsement of the thesis from within a broader consequentialist outlook.<sup>68</sup>

However, reductivism is not a consequentialist view. Instead, reductivists generally endorse a broadly 'threshold deontological' position within normative ethics, and of the morality of interpersonal harming

---

No.2 (2010), 123-139 at p.136. Brownlee also notes the parallel between this view and reductivist approaches to just war theory.

<sup>67</sup> Lazar, 'National Defence, Self-Defence, and the Problem of Political Aggression', p.14

<sup>68</sup> For other consequentialist perspectives on the ethics of war, see R.B. Brandt, 'Utilitarianism and the Rules of War', *Philosophy and Public Affairs* 1. No.2 (1972), 145-165; R.M Hare, 'Rules of War and Moral Reasoning', *Philosophy and Public Affairs* 1. No.2 (1972), 166-181; William Shaw, 'Utilitarianism and Recourse to War', *Utilitas* 23, No. 4 (2011), 380-401.

specifically.<sup>69</sup> This view is committed to the idea that there are weighty non-consequentialist constraints on harming others, which is typically expressed in terms of individual's possession of important rights against being killed and maimed. But it also holds that these constraints are not absolute and may be permissibly overridden when doing so brings about a considerably more valuable state-of-affairs compared to that which would result from refraining from transgressing the right. Despite their focus on individuals' rights, reductivists accept that numbers also matter morally.<sup>70</sup>

Furthermore, reductivists typically assess the permissibility of interpersonal harming not only in terms of trade-offs between respect for rights and the promotion of value, but also in terms of the modes of agency by which harm is brought about. For example, reductivists generally endorse the standard deontological distinctions between doing and allowing, intending and foreseeing, and between manipulative and eliminative agency.<sup>71</sup> In each of these pairs, causing harm via the first mode of agency is considered subject to a higher justificatory burden than the second, other things being equal.

According to reductivism, the morality of war is constituted by two basic forms of justification for causing harm drawn from ordinary interpersonal morality. Each form of justification provides a different

---

<sup>69</sup> For explicit endorsement, see Fabre, *Cosmopolitan War*, pp.13-14.

<sup>70</sup> As Helen Frowe puts it, "There's nothing in reductivism that prohibits caring about numbers. The central claim of reductivism is that the moral rules of war are the moral rules of ordinary life, and aggregation is certainly a feature of morality in ordinary life." *Defensive Killing*, Ch.5.

<sup>71</sup> On the distinction between doing and allowing, see, for example, Jeff McMahan, 'Killing, Letting Die, and Withdrawing Aid', *Ethics* 103, No.2 (1993), 250-279; Warren Quinn, 'Actions, Intentions and Consequences: The Doctrine of Doing and Allowing', *The Philosophical Review* 98, No.3 (1989), 287-312; Philippa Foot, *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford: Oxford University Press, 1978), Ch.2; Samuel Scheffler, 'Doing and Allowing', *Ethics* 114, No.2 (2004), 215-239. On the distinction between intending and foreseeing, see Warren Quinn, 'Actions, Intentions and Consequences: The Doctrine of Double Effect', *Philosophy and Public Affairs* 18, No.4 (1989), 334-251; Foot, *Virtues and Vices and Other Essays in Moral Philosophy*, Ch.2; Jeff McMahan, 'Intention, Permissibility, Terrorism and War', *Philosophical Perspectives* 23 (2009), 345-372; Victor Tadros. *The Ends of Harm: The Moral Foundations of Criminal Law* (Oxford: Oxford University Press, 2011), Ch.7. On the distinction between eliminative and manipulative harm, see Quinn, 'Actions, Intentions and Consequences: The Doctrine of Double Effect'; Jonathan Quong, 'Killing in Self-Defense', *Ethics* 119, No.2 (2009), 334-351; Tadros, *The Ends of Harm*, Ch.6.

account of how the usual constraint on killing and injuring can be defeated by reasons in favour of causing harm.

Firstly, killing in war can be justified by appeal to principles of permissible killing in self- and other-defence. As standardly understood, causing defensive harm is permissible because the subject of harm has lost or forfeited their normal right against being harmed. This loss arises when harming that individual serves to avert a threat of harm to which they are appropriately connected. In the relevant terminology, the subject of harm is *liable* to a certain amount of harm, and suffers no wrong by being subjected to it. The key debate within the literature on self-defence concerns the correct rendering of the ‘appropriately connected’ clause for grounding liability.<sup>72</sup> For reductivists, the intentional killing of combatants in war, when permissible, is primarily justified because the combatants have acted in such a way as to render themselves liable to defensive killing. They may thus be killed by their opponents without being wronged.<sup>73</sup> The appeal to individual defensive rights is a defining component of a reductivist view. As McMahan puts it:

the morality of defense in war is continuous with the morality of individual self-defense. Indeed, justified warfare just *is* the collective exercise of individual rights of self- and other-defense in a coordinated manner against a common threat.<sup>74</sup>

Of course, liability justifications alone will not suffice to justify warfare, since wars typically involve causing serious harm to those who are not

---

<sup>72</sup> For a representative sample, see Thomson, ‘Self-Defense’; Kimberly Ferzan, ‘Justifying Self-Defense’, *Law and Philosophy* 24, No.6 (2005), 711-749; Jeff McMahan, ‘The Basis of Moral Liability to Defensive Killing’, *Philosophical Issues* 15 (2005), 386-405; Jonathan Quong, ‘Liability to Defensive Harm’, *Philosophy and Public Affairs* 40, No.1 (2012), 45-77; David Rodin, *War and Self-Defence*, Chs.1-4; Victor Tadros, ‘Duty and Liability’, *Utilitas* 24, No.2 (2012), 259-277.

<sup>73</sup> Jeff McMahan has recently argued that the intentional killing of combatants may also be justified by appeal to a *combination* of liability and lesser-evil justifications. See Jeff McMahan ‘What Rights May We Defend By Means of War’ in Fabre and Lazar (eds), *The Morality of Defensive War*, 115-158 at pp.138-140.

<sup>74</sup> McMahan, ‘The Ethics of Killing in War’, p.717. See also, McMahan, *Killing in War*, pp.155-158



liable. Most obviously, conducting a war will invariably involve collaterally killing non-liable non-combatants. For reductivists, such killing, when justified, is justified by appeal to considerations of *lesser-evil*. According to this form of justification, an individual's right not to be harmed or killed may, under certain circumstances, be justifiably overridden by weightier moral reasons in favour of harming or killing them. Applied to the collateral killing of non-liable persons in war, reductivists hold that such acts can be justified by appeal to the impartial good brought about by doing so, in conjunction with discounting the moral gravity of the rights violation in virtue of it being an unintended side-effect of producing the good effect.

According to reductivists, the moral permissibility of both the resort to war and of conduct in war is determined by these two forms of justification. As Seth Lazar summarises, on this view "Justified warfare...is no more than the coextension of multiple acts justified under these two principles."<sup>75</sup> Term this the *Content Thesis*.<sup>76</sup> The thesis can be understood as essentially a claim about the range of reasons that are capable of justifying causing serious harm to others. According to this thesis, for any act of violence – in war and in any other context – if the recipient of harm retains their normal right not to be harmed, and their right is not overridden by considerations of impartial good, then agents have insufficient moral reason to cause that harm and are unjustified in doing so.

### **2.3 Jettisoning the Restrictive Authority Criterion**

With this characterisation of reductivism in place, I now move on to explain why a commitment to reductivism generates considerable scepticism regarding the authority criterion. I begin with the standard, restrictive interpretation of the criterion and the normative effects that follow from its possession and non-possession. This interpretation can be summarised as follows:

---

<sup>75</sup> Lazar, 'National Defence, Self-Defence and the Problem of Political Aggression', p.12

<sup>76</sup> In addition to endorsing a version of the Continuity Thesis, John Gardner also endorses a view that is analogous to the Content Thesis. 'Criminals in Uniform', pp.112-113.

***The Restrictive Authority Criterion:*** In order for a war to be justified, it must be initiated and fought by an entity that possesses a certain set of properties.

As explained in the previous chapter, the relevant properties for constituting a war-waging authority are standardly specified in terms that restrict the authority to wage war to states. As we saw, one problem with this specification is that it implies that no rebellion or civil war could be justified on the part of the insurgents, since these belligerents lack the relevant statist properties. This seems implausible, since we can generate counter-examples – either real or imagined – in which a belligerent entity lacks those properties, but the normative conclusion that its war is unjustified intuitively does not follow. The possibility of justified wars fought by non-state actors cases is often taken to refute the idea that fighting on behalf of a certain kind of entity is in fact a necessary condition for justifying the resort to war.

However, such cases are not sufficient to warrant jettisoning the restrictive authority criterion from just war theorising. This is because an alternative conclusion can be drawn from the intuitive possibility of justified wars fought by non-states. It can be argued that these cases only provide counter-examples to the statist conception of the relevant properties that constitute being an authority, and not to the normative effect that follows from the possession or non-possession of the relevant properties. In other words, they only show that mainstream just war theory is mistaken regarding who has war-making authority, and that the hallmarks of statehood provide too narrow an account of what the authority to wage war consists in.<sup>77</sup> They do not licence the stronger conclusion that the possession of such authority is not necessary for justifying the resort to war.

---

<sup>77</sup> This is the conclusion drawn, independently, by Ian Holliday and Anne Schwenkenbecher. Ian Holliday, 'When is a Cause Just?', *Review of International Studies* 28, No.3 (2002), 557-575 at pp.567-568; Schwenkenbecher, 'Rethinking Legitimate Authority', p.167.

This response is supported by the fact that the just war tradition predates the modern state by over a thousand years, yet has retained a central place for the authority criterion throughout its development. Within medieval just war theory, formulated in an environment of diverse, overlapping and hierarchical forms of political organisation, one of the key topics of debate, if not *the* key topic, was over precisely which kind of entities possessed the relevant properties for being war-making authorities.<sup>78</sup>

With this perspective in mind, one may reply to the putative counter-examples that they arise from a myopic fixation on statehood, out of keeping with the just war tradition as a whole, and that a more nuanced account of war-making authority can resolve the objections to the authority criterion that they give rise to. Several theorists have aimed to provide such an account, which de-couples the authority to wage war from the explicit notion of statehood. On one view, the possession of war-waging authority is understood as requiring that belligerent entities meet certain political or organisational conditions, but which do not presuppose the formal characteristics of statehood.<sup>79</sup> On another view, authority requires only that a belligerent entity be sufficiently representative of those on whose behalf they fight.<sup>80</sup> Given these possible reformulations of the relevant authority-conferring properties, the intuitive plausibility of justified non-state war need not pose as great a challenge to the restrictive authority criterion as one might suppose.

---

<sup>78</sup> For a detailed account of these debates, see Frederick Russell, *The Just War in the Middle Ages* (Cambridge: Cambridge University Press, 1977). See also, Heather Wilson, *International Law and the Use of Force By National Liberation Movements*, Ch.1.

<sup>79</sup> Anthony Coates, *The Ethics of War* (Manchester: Manchester University Press, 1997), Ch.1; Janna Thompson, 'Terrorism and the Right to Wage War' in C.A.J. Coady and Michael O'Keefe (eds), *Terrorism and Justice: Moral Argument in a Threatened World* (Melbourne: Melbourne University Press, 2002), 87-96; Janna Thompson, 'Terrorism, Morality and Right Authority' in Georg Meggle (ed), *Ethics of Terrorism and Counter-Terrorism* (Frankfurt: Ontos, 2005), 51-60. For criticism of Coates' and Thompson's formulations of the criterion, see Steinhoff, *The Ethics of War and Terrorism*, Ch.1

<sup>80</sup> See, for example, Finlay, 'Legitimacy and Non-State Political Violence'; Lionel McPherson, 'Is Terrorism Distinctively Wrong?', *Ethics* 117, No.3 (2007), 524-546; Schwenkenbecher, 'Rethinking Legitimate Authority'.

### 2.3.1 *The First Spectrum Argument*

A reductivist view provides the resources for a deeper and more powerful objection to the restrictive authority criterion. Rather than simply appealing to counter-examples, which challenge only the standard conception of the relevant authority-generating properties, reductivism denies the normative effect directly, arguing that the possession or absence of *any* such properties are irrelevant to justifying the resort to war. If successful, such an argument will explain *why* the intuitive counter-examples are genuine objections to the inclusion of an authority criterion within *jus ad bellum*.

The argument follows from the foundational reductivist claim that war is to be morally analysed solely in terms of aggregations of individuals' permissions to use defensive force. On this view, war lies on a continuum with ordinary acts of interpersonal harming and killing. Jeff McMahan provides a concise statement of this conception of war and its ethical basis:

First imagine a case in which a person uses violence in self-defense; then imagine a case in which two people engage in self-defense against a threat they jointly face. Continue to imagine further cases in which increasing numbers of people act together with increasing coordination to defend both themselves and others against a common threat, or a range of threats they face together. What you are imagining is a spectrum of cases that begins with acts of individual self-defence and, as the threats become more complex and extensive, the threatened individuals more numerous, and their defensive action more integrated, eventually reaches cases involving a scale of violence that is constitutive of war.<sup>81</sup>

Such a view enables what I term the *First Spectrum Argument* against the authority criterion. The main claim of the argument is that ordinary

---

<sup>81</sup> McMahan, 'War as Self-Defense', p.75.

principles of interpersonal morality are *sufficient* to justify both the resort to war and the individual acts of killing and injuring necessary for its prosecution. We can move from individual acts of justified defence to a state of justified war by a process of aggregating acts that are each independently justified on the basis of principles of self- and other-defence (more specifically – liability and lesser-evil justifications).

The key point is that these principles, at least as usually understood, contain no requirement that defence be authorised or carried out by a particular kind of entity. If a liability or lesser-evil justification exists, any appropriately motivated individual may act on this justification. From this, reductivists are able to draw the conclusion that “it is not a necessary condition of just or justified war that it be initiated only by persons who are properly authorized to do so.”<sup>82</sup> On this view, all the normative materials required to justify war and its conduct can be derived from ordinary norms of interpersonal morality. The authority requirement is therefore redundant.

The First Spectrum Argument can be put more forcefully by drawing out the implications of the opposing view – that authority *is* a necessary condition of justified war. If true, this implies that as we move across the spectrum of cases of defensive action, at some point the addition of additional acts would change the situation from a state of non-war to one of war, thus activating an additional moral constraint – the requirement of authority. Past this point, the sum of defensive actions would be rendered morally unjustified if carried out in the absence of a war-waging authority. But this seems implausible. How can the mere addition of justified actions to other justified actions render the sum of actions morally unjustified? Yet

---

<sup>82</sup> Jeff McMahan ‘Just War’ in Robert E. Goodin, Philip Pettit and Thomas Pogge (eds), *A Companion to Political Philosophy* 2<sup>nd</sup> Edition, (Oxford: Blackwell, 2007), 669-677 at p.671. For further reductivist rejections of the authority criterion as a requirement of *jus ad bellum*, see Cecile Fabre, ‘Cosmopolitanism, Just War Theory and Legitimate Authority’, *Ethics and International Affairs* 84, No. 5 (2008), 963-976; Fabre, *Cosmopolitan War*, Chs. 3-4; Jeff McMahan, ‘Just Cause for War’, *Ethics and International Affairs* 19, No.3 (2005), 1-21 at p.4; Steinhoff, *The Ethics of War and Terrorism*, Ch.1. The criterion is also rejected as an *ad bellum* requirement in David K. Chan, *Beyond Just War: A Virtue Ethics Approach* (Basingstoke: Palgrave MacMillan, 2012), pp.46-48; Gabriel Palmer-Fernandez, ‘Cosmopolitan Revisions to Just War Theory’, in Michael Boylan (ed), *The Morality and Global Justice Reader* (Boulder: Westview Press, 2011), 325-342; Darrel Moellendorf, *Cosmopolitan Justice* (New York: Basic Books, 2001), p.121.

the defender of the authority criterion, in its restrictive role, seems committed to this conclusion. Term this the *aggregation problem*.<sup>83</sup>

One way to defend the authority criterion while avoiding this implication would be to argue that authority *is* in fact a necessary condition of permissible self- and other-defence. On this view, acts of individual defence are only morally permissible if carried out with the endorsement of a relevant authority. Defensive harming is thus unjustified across the whole spectrum of cases if not appropriately authorised. There is then no point at which the addition of individual acts of defence inverts the deontic status of the sum of defensive acts.

However, this move, while successfully avoiding the aggregation problem, seems equally implausible. One doesn't have to be a Hobbesian to find it powerfully intuitive that individuals possess a pre-institutional right of self-defence.<sup>84</sup> To demonstrate, imagine that a generally legitimate state issues a blanket prohibition on its citizens using force in self-defence. Is it plausible to think that a citizen of that state would be acting morally wrongly if, following the prohibition, they use lethal force against another individual who would otherwise kill two innocent persons?<sup>85</sup> Additionally, consider the case of individuals who exist outside the jurisdiction of any political or institutional body, in a completely failed state for example. Would such individuals necessarily act wrongly if they inflict defensive harm on culpable aggressors?<sup>86</sup>

---

<sup>83</sup> For discussion of related aggregation issues in the context of defensive harm, and endorsement of the claim that mere aggregation cannot affect the deontic status of individual acts of harming, see Ron Aboodi, Adi Borer and David Enoch, 'Deontology, Individualism, and Uncertainty: A Reply to Jackson and Smith', *Journal of Philosophy* 105, No.5 (2008), 259-272.

<sup>84</sup> For defence of this view, see Kimberly Ferzan, 'Self-Defense and the State', *Ohio State Journal of Criminal Law* 5, No.2 (2008), 449-504.

<sup>85</sup> A sentiment forcefully echoed in a United States Court of Appeals opinion, "It is difficult to the point of impossibility to imagine a right in any state to abolish self-defense altogether." *Griffin v. Martin*, 785 F.2d 1172, 1187 n.37 (4th Cir. 1986). Cited in Ferzan, 'Self-Defense and the State', n.111.

<sup>86</sup> Those of a Kantian persuasion may be more sympathetic to the view that self-defence requires authority. For example, Anna Stilz argues that private defence is morally sub-optimal and regrettable compared to distributing harms under a legitimate legal order. Anna Stilz, 'Authority, Self-Determination and Community in Cosmopolitan War', *Law and Philosophy*, (forthcoming). However, I am not aware of any contemporary theorist who defends the much stronger view that self-defence is morally *impermissible* in the absence of authority.

Making a similar point on more principled grounds, Cecile Fabre argues that if we accept the uncontroversial premise that individuals possess fundamental rights against certain forms of treatment *qua* individuals, then we must also accept that they have a right to defend those rights *qua* individuals. She writes,

it would be incoherent, on the one hand, to claim that individuals' fundamental interests in life-saving resources and basic freedoms are important enough to be protected by rights, and on the other hand, to deny that individuals' interest in the protection of those rights is important enough to be protected by a right to defend those rights themselves.<sup>87</sup>

For Fabre, and for reductivists more generally, if one accepts a fairly minimal view of individuals' pre-institutional moral rights, one cannot avoid the conclusion that acting on behalf of an authority is not a necessary condition for justified resort to war. On this view, the right to wage war is ultimately an individual right.<sup>88</sup>

According to the First Spectrum Argument, then, a defender of the restrictive authority criterion faces a dilemma between confronting the aggregation problem, or denying that individuals possess basic, pre-institutional rights of self-defence. Since neither of these options is plausible, the traditional just war requirement that a justified war must be fought by a certain kinds of entity has no principled moral basis and ought to be jettisoned from *jus ad bellum*. The argument also shows how counter-examples to the restrictive authority criterion can be generated, however the relevant properties are formulated. Since, for reductivists, justified war and individual self-defence lie on a continuum, any formulation of the criterion will necessarily be under-inclusive, failing to capture all the

---

<sup>87</sup> Fabre, *Cosmopolitan War*, p.115. For an additional defence of the view that the possession of basic rights entails the possession of 'remedial' defensive rights, see Tony Honore, 'The Right to Rebel', *Oxford Journal of Legal Studies* 8, No.1 (1988), 34-54.

<sup>88</sup> Fabre, *Cosmopolitan War*, p.144. See also, Uwe Steinhoff, "If, under certain conditions, a *right to war* comes into operation, it is, or is based upon, an *individual right*." *The Ethics of War and Terrorism*, p.20.

possible cases of justified warfare.<sup>89</sup> Since anyone may potentially use force in self-defence, permissible warfare cannot be non-arbitrarily restricted to a certain class of entities. On this view, “groups of actors which act in unstructured, disorganized way, as well as individuals themselves, can have the right to wage war.”<sup>90</sup>

### 2.3.2 Deflating the Restrictive Authority Criterion

Although reductivists reject the idea that a belligerent entity must be of a certain type in order to permissibly resort to war, this should not be interpreted to mean that reductivists are committed to the view that that certain properties of belligerent entities – such as their size, political organisation, possession of territory, etc. – are irrelevant to whether that entity is all-things-considered justified in resorting to war. This is because, even if the authority criterion is not an independent requirement of *jus ad bellum*, a war must still satisfy the other *ad bellum* requirements in order to be justified. Whether these requirements are satisfied will often be determined by the possession of certain properties, including those that are standardly associated with the possession of war-making authority. Given this, reductivists can support their rejection of the restrictive authority criterion by providing a deflationary account of the intuition that only certain kinds of groups are justified in resorting to war.<sup>91</sup>

---

<sup>89</sup> Unless, of course, one formulates the requirement so broadly as to include private individuals. But this is practically equivalent to rejecting the criterion altogether, since no actors will fail to possess it. Indeed, Fabre sometimes phrases her rejection of the authority criterion as the claim that all individuals potentially have the authority to wage war.

<sup>90</sup> Fabre, *Cosmopolitan War* pp.144-145. For further endorsements of the possibility of justified wars fought by private individuals, see Uwe Steinhoff, ‘The Ethics of War and Terrorism’, p.20; Uwe Steinhoff ‘What is War? And Can a Lone Individual Wage One?’, *International Journal of Applied Philosophy* 29, No.1 (2009), 133-150; Pattison, ‘When Is It Right to Fight’, p.53. The right of private war is also endorsed by some philosophical anarchists, who reject the view that political authorities are in a morally special position vis-à-vis any activity. See, for example, Michael Huemer, *The Problem of Political Authority: An Examination of the Duty to Obey and the Right to Coerce* (Basingstoke: Palgrave Macmillan, 2013), p.144.

<sup>91</sup> The following two paragraphs draw on Fabre, *Cosmopolitan War*, pp.116-118 and p.147. See also, Schwenkenbecher, ‘Rethinking Legitimate Authority’, pp.167-168.



Most obviously, whether a potential war satisfies the requirement of reasonable prospect of success will depend heavily on whether the belligerent waging it has certain capacities, such as having sufficient combatants, military hardware, strategic competence, economic resources, etc. Given this, it will invariably be the case that certain kinds of entity lack the ability to successfully prosecute a war, or to prosecute it at a proportionate moral cost, and will therefore not be justified in resorting to war.

In addition, the requirement of last resort will often rule out certain groups from permissibly resorting to war, even if they have the capacity to successfully prosecute it. For example, if the members of a sub-state ethnic group are threatened with rights violations of sufficient seriousness to provide a just cause for war, it may still be impermissible for them to resort to war themselves if their state, or some other institution, is able and willing to prevent those rights violations and can do so at a lesser moral cost. Under these conditions, it would be impermissible for the prospective victims to exercise their rights of self-defence themselves, since there is a morally superior alternative option available.

Given the additional requirements of *jus ad bellum*, it will often be the case that non-state belligerents – disorganised groups of individuals more obviously – will not be justified in waging war, even if they possess a just cause. Moreover, these entities will lack justification in virtue of their failure to possess certain properties standardly associated with the possession of war-making authority. But, importantly, reductivists can explain the absence of justification without invoking an independent authority requirement. The moral significance of these properties can be attributed to the contribution that they make to the satisfaction of other *ad bellum* requirements, which can be accounted for in wholly reductive terms.

## 2.4 Jettisoning the Permissive Authority Criterion

The reductivist critique of the authority criterion has primarily focussed on its traditional restrictive role within *jus ad bellum*. This makes perfect sense, for this is how most people have interpreted the criterion. Furthermore, I believe that reductivists have succeeded in showing that it is not a necessary condition of justified war that it be fought by a certain type of entity. However, as argued in the previous chapter, this restrictive function is not the only role, or perhaps even the main role, played by the authority criterion within the just war tradition. The criterion also has an important permissive function within *jus in bello*, which can be characterised as follows:

***The Permissive Authority Criterion:*** If an agent participates in war on behalf of an entity that possesses a certain set of properties, then that agent is subject to a more extensive range of permissions to cause harm than if they fought privately, or on behalf of an entity which lacks that set of properties.

Under the permissive interpretation, a different normative effect follows from the possession of war-making authority than under the standard, restrictive interpretation. Moreover, these two effects are conceptually independent of one another, in that one can coherently reject the restrictive, while endorsing the permissive. On this conjunction of views, it is not a necessary condition of justified resort to war that it be waged by a certain kind of entity, but when certain kinds of entities wage war, their agents are subject to more extensive permissions to cause harm *in bello*, that they would lack if they fought in the absence of an authority.

Given this possible combination of positions, arguments against the restrictive interpretation of the authority criterion, if successful, do not settle the question of whether the inclusion of an authority criterion within just war theory is morally defensible. Further argument is required in order to support jettisoning the criterion completely. Critics of the criterion

often overlook this since they assume, along with the majority of just war theorists, that the authority criterion functions solely in the standard, restrictive manner.<sup>92</sup>

#### 2.4.1 *The Continuity Thesis*

Reductivism also contains the resources for rejecting the permissive authority criterion. As argued in Chapter 1, a major source of motivation for the inclusion of an authority criterion within just war theory lies in a commitment – either explicit or implicit – to the idea that war is morally exceptional, so that conduct in war is governed by different moral principles to those that govern conduct in other circumstances. Given this commitment it is necessary to have a means of distinguishing the class of activities governed by war-specific moral principles. As I argued, it is the authority criterion that provides the resources for fulfilling this crucial demarcating function. What distinguishes combatants – those subject to additional war-specific permissions to cause harm – from other violence-users is that the former use force on behalf of a certain kind of entity.

Given their central commitment to the Continuity Thesis, reductivists wholeheartedly reject the view that war is morally exceptional. As McMahan puts it, “a state of war...does not call forth a different set of principles, but merely complicates the application of moral principles that are of universal application.”<sup>93</sup> War and all other activities inhabit a single moral domain. Given such a view, the pressure to include an authority criterion completely disappears. There is no need to morally demarcate war and non-war, since action in both domains is governed by exactly the same moral principles. The problem that the authority criterion is required to resolve simply does not arise on a view that endorses the Continuity Thesis. War can be defined in entirely non-normative, quantitative terms

---

<sup>92</sup> Cecile Fabre being an exception (see, *Cosmopolitan War*, p.160). However, despite this, Fabre still devotes the majority of her discussion to rejecting the criterion in its standard role.

<sup>93</sup> Jeff McMahan, ‘Précis of *The Morality and Law of War*’, *Israel Law Review* 40, No. 3 (2007), 310-323 at p.316.

familiar from the social scientific literature on warfare, since nothing morally significant turns on whether or not a particular act of violence takes place within the context of a state of war.<sup>94</sup>

#### 2.4.2 *Rejecting the Orthodox View*

Given their commitment to the Continuity Thesis, reductivists reject theories of just war that presuppose the view that war is morally exceptional. Most saliently, reductivists typically reject an orthodox conception of just war theory, which endorses the Equality Thesis. As explained, it is extremely difficult to ground the Equality Thesis in standard justifications for interpersonal harming, such as self-defence. In non-war contexts moral evaluation of the ends pursued is directly relevant to determining the moral permissibility of causing harm in furtherance of those ends. In order to avoid a conflict with the ordinary norms of interpersonal morality, it is necessary to combine a commitment to orthodox just war theory with a rejection of the view that war is morally continuous with all other spheres of activity.

Reductivists resolve the conflict differently, by rejecting the orthodox conception of permissible conduct in war that generates the conflict. As Jeff McMahan puts it,

“the reductive strategy is incompatible with the traditional just war theory—in particular the latter’s assumption that the rules of *jus in bello* are independent of the rules of *jus ad bellum*. But the conclusion we should draw is not that the reductive strategy is false but that the traditional theory is.”<sup>95</sup>

For reductivists, whether an agent acts permissibly in war cannot be determined independently of whether their war is *ad bellum* justified. Combatants who participate in unjust wars act seriously morally wrongly

---

<sup>94</sup> David Rodin makes a similar observation in ‘The Myth of National Defence’, in Fabre and Lazar (eds), *The Morality of Defensive War*, 69-89 at n.21.

<sup>95</sup> McMahan, ‘War as Self-Defense’, p.79.

by doing so, because they typically lack a liability or lesser-evil justification for intentionally killing their opponents and collateral killing non-combatants.<sup>96</sup> A commitment to reductivism thus generally entails a revisionary view of permissible conduct in war, in which there is a fundamental moral inequality between combatants who fight in just wars and those who fight in unjust wars.<sup>97</sup> While killing in a just war can be justified on the basis of individual rights of self-and other-defence, those who kill in unjust war commit a species of unjustified homicide.

It is worth pointing out that despite this revisionary implication, reductivists do not argue that the current legal regime governing war – which grants equal legal permissions to combatants who participate in both just and unjust wars – should be abandoned. What reductivists do argue is that the justification of the law’s permission to kill in war lies not in deep moral principle, but in largely pragmatic considerations concerning the difficulties and undesirable consequences of reforming the current legal regime.<sup>98</sup>

Much of the recent discussion and controversy over reductivism has focused on this particular substantive conclusion, to the extent that a rejection of the Equality Thesis is often taken to be the defining feature of reductivism. Indeed, many writers, including the main proponents of reductivism, use the label ‘revisionism’ to refer to the reductivist view. However, it is important to keep in mind that the rejection of the Equality Thesis is simply an important implication of a broader theoretical

---

<sup>96</sup> The ‘typically’ caveat is important, since there are possible cases in which a combatant is permitted to kill, on reductivist grounds, despite fighting in an unjust war. For example, a combatant on the unjust side may be permitted to kill a combatant on the just side who would otherwise commit war crimes. McMahan, *Killing in War*, pp.16-17. In addition, Saba Bazargan has argued that it may be permissible on reductivist grounds for combatants to participate in justified phases of wars that are overall unjustified. Saba Bazargan, ‘The Permissibility of Aiding and Abetting Unjust Wars’, *Journal of Moral Philosophy* 8, No.4 (2011), 513-529.

<sup>97</sup> For the most detailed and sustained argument for the moral inequality of combatants, see McMahan, *Killing in War*, Chs 1-2.

<sup>98</sup> See McMahan, ‘The Ethics of Killing in War’, at pp.729-733; Jeff McMahan, ‘The Morality of War and the Law of War’, in Rodin and Shue (eds), *Just and Unjust Warriors*, 19-43. For discussion and criticism, see David Rodin, ‘The Morality and Law of War’, in Strachan and Scheipers (eds), *The Changing Character of War*, 446-463; Henry Shue, ‘Do We Need a ‘Morality of War’ in Rodin and Shue (eds), *Just and Unjust Warriors*, pp.87-111.

commitment.<sup>99</sup> For our purposes, what is noteworthy about the reductionist rejection of the Equality Thesis is that it removes an important motivation for the claim that individuals acting on behalf of certain kinds of entity enjoy additional permissions to cause harm in war.

### *2.4.3 Rejecting the Classical View*

This last point can be emphasised by noting that reductionism equally rejects a classical conception of just war theory, despite the fact that both views are united in rejecting the Equality Thesis. This is because, as explained in the previous chapter, the classical view is also committed to idea that war is morally exceptional. It shares this feature with the orthodox view, despite their opposing positions on the Equality Thesis.

On a classical view, the key challenge for just war theory is to square the non-pacifist position that participation in war can be morally permissible with a fairly restrictive account of permissible interpersonal harming. It does so by endorsing the view that war is morally exceptional, subject to more permissive norms of conduct. Since the morality of conduct in war is not grounded in norms of interpersonal morality, there is no conflict. It is this commitment that necessitates a role for the permissive authority criterion within a classical view. On both an orthodox and classical view there is a justificatory gap between permissible conduct in war and the norms of private interpersonal morality, which is bridged by the claim that agents who act on behalf of certain kinds of entities enjoy additional permissions to cause harm.

---

<sup>99</sup> More generally, the conflation of reductionism and revisionism is unhelpful, since it runs together two dimensions along which we can distinguish positions within the ethics of war that ought to be kept separate: their theoretical commitments and their practical implications. It is often assumed that reductionism at the theoretical level entails revisionism in practice, and, conversely, that non-reductionism entails a more traditional view. While there is a general correlation between these pairs of view, the connection is not necessary. Positions can be practically coextensive while diverging at the theoretical level. For an example of a non-reductionist view that is revisionist in rejecting the Equality Thesis, see Noam Zohar, 'Collective War and Individualist Ethics: Against the Constriction of Self-Defense', *Political Theory* 21, No.4 (1993), 606-622. For an example of a minority reductionist view that defends the Equality Thesis, see Steinhoff, 'Jeff McMahan on the Moral Equality of Combatants'.

In response to each of these views, reductivists deny that there is in fact a justificatory gap. In the case of the orthodox view, reductivists close the gap by rejecting the conception of right conduct which conflicts with norms of interpersonal morality. In the case of the classical view, by contrast, reductivists close the gap by rejecting the conception of interpersonal morality that leads to it. As explained in the previous section, reductivists hold that the acts of killing and injuring required to wage justified wars can be accounted for solely in terms of ordinary interpersonal morality.<sup>100</sup> Whereas a reductivist view rejects the orthodox view as overly permissive, it equally rejects the classical view for operating with an overly restrictive conception of permissible interpersonal harming, and of permissible self-defence in particular. Since, for reductivists, the permissibility of participation in justified wars can be accounted for solely in terms of justifications of causing harm contained within ordinary interpersonal morality, a role for the authority criterion in grounding a more expansive set of permissions to cause harm in war is removed.

This comparison of the classical and reductivist views in terms of their positions on the importance of the authority criterion thus reveals the deep differences between them, which is obscured by the fact that they give equivalent verdicts on certain substantive questions of permissible conduct in war, such as the truth of the Equality Thesis.<sup>101</sup> On both a classical and orthodox view the authority criterion is indispensable, whereas for reductivists it is a spare wheel. The source of this difference lies in their divergence on the question of the relationship between the morality of war and the moral norms that govern interpersonal harming in other contexts.

---

<sup>100</sup> See McMahan, 'War as Self-Defense'.

<sup>101</sup> Given their mutual rejection of the Equality Thesis, some reductivists refer to their view as the 'neo-classical' view (for example, Fabre, 'Guns, Food and Liability to Attack in War'). As the discussion in this section shows, reductivists should be wary of over-emphasising the historical pedigree of their view. Despite their practical convergence on this issue, they remain very different in terms of their deeper theoretical commitments.

#### 2.4.4 The Second Spectrum Argument

As we have seen, the reductivist commitment to the Continuity Thesis removes a powerful source of motivation for the permissive authority criterion. However, the Continuity Thesis alone is not sufficient to jettison the criterion. As explained above, the Continuity Thesis is purely formal, holding that the morality of conduct in war is wholly determined by the same moral principles that determine the permissibility of interpersonal harming in all other circumstances.

Given this, there remains conceptual space for the view that that interpersonal morality assigns additional permissions to cause harm to those who do so on behalf of certain kinds of entities or authority. This is not a claim about there being any discontinuity between war and ordinary life, or that individuals “stop being bound by morality when they put on their uniforms, or otherwise go on duty.”<sup>102</sup> Rather it is a claim about the precise content of a single, unified set of moral principles that govern all agents and all activities.

The reductivist Content Thesis closes off the space for such a view, thus enabling the complete jettisoning of the authority criterion. The thesis holds that liability and lesser-evil justifications *exhaust* the range of justifications for non-consensual harming. As McMahan puts it, “On [a reductivist] view, the *limits* of individual self- and other-defense are also the *limits* of national defense.”<sup>103</sup> In other words, if an act of killing or injuring – in war or any other context – is not independently justified on the basis of liability or lesser-evil, then that act is morally impermissible. This is true regardless of whether the agent performing the act does so on behalf of a certain type of entity.<sup>104</sup>

To demonstrate this position, the reductivist can offer what I term the *Second Spectrum Argument* against the authority criterion. Recall, in McMahan’s spectrum example, we consider an escalating range of cases,

---

<sup>102</sup> Gardner, ‘Criminals in Uniform’, p.116.

<sup>103</sup> McMahan, ‘War’, p.310 (my emphasis).

<sup>104</sup> In a talk given at the University of Stockholm in May 2014, McMahan explicitly labeled this view the ‘No Extensions Principle’ and affirmed his commitment to it.



starting with a small-scale case of interpersonal harming involving a handful of individuals using force against others. We then, by a process of aggregating such cases, eventually reach a level of violence constitutive of a state of war. Whereas the First Spectrum Argument is designed to show that liability and lesser-evil justifications are *sufficient* to justify the kind of harming and killing constitutive of war (so that the authority criterion does not impose an additional constraint), the Second Spectrum Argument holds that these two forms of justification are also *necessary* for justifying such actions (so that the authority criterion does not give rise to additional permissions or justifications.)

The argument holds that if the range of justifications for harming is exhausted by considerations of liability and lesser-evil in the small-scale cases, then the same is true of harming in a state of war. All that changes as we move across the spectrum is that the cases become more complex, organised and protracted, but the basic moral principles remain the same throughout. The conditional claim seems highly plausible, as does the antecedent. In the small-scale cases, it is uncontroversial that it would be impermissible for an individual to non-consensually kill another in the absence of a liability or lesser-evil justification.

While the First Spectrum Argument aims to deny that additional constraints on causing harm emerge as we move across the spectrum, the Second Spectrum Argument holds that no additional forms of permission emerge either. The limits of permissible conduct in war are set by the justifications for harming that apply to individuals pre-institutionally. The fact that an agent fights on behalf of a certain kind of entity cannot increase their permissions to cause harm above this baseline.<sup>105</sup> For reductivists, then, the claim encapsulated in the permissive authority criterion is false. There are no relevant properties that give rise to this particular normative effect.

It is worth pointing out that reductivists are not (necessarily) committed to the much stronger view that the whole range of reasons for action that apply to agents in their institutional roles are exhausted by

---

<sup>105</sup> See Gardner, 'Criminals in Uniform', pp.112-113.

those that apply to individuals in their private capacity. They need only make the more specific and weaker claim that the limits of permissible harming are set by the limits of private morality. This view accepts that agents may be subject to additional moral reasons in favour of causing harm in virtue of occupying a role or position within an institution. What it denies is that these reasons are ever sufficient to justify causing serious non-consensual harming in the absence of a liability or lesser-evil justification.

## **2.5 Conclusion**

According to an increasingly influential reductivist approaches to the ethics of war, the standard forms of justification for causing harm contained within private interpersonal morality are both necessary and sufficient for justifying the resort to war and the prosecution of wars. Given this dual-claim, reductivism endorses jettisoning the authority criterion from just war theorising, in both its restrictive and permissive roles. Justified warfare cannot be restricted to a certain class of entities and acting on behalf of a particular kind of entity cannot generate any additional permissions to kill and injure beyond those that obtain between private individuals. At best, the authority criterion is redundant, at worst it is deeply pernicious. Given the centrality of considerations of authority within the just war tradition throughout its development, it is not an overstatement to say that the most revisionary aspect of a reductivist view is its rejection of the authority criterion.

As mentioned above, I accept the reductivist critique of the restrictive authority criterion. It is not a necessary condition of a war's being justified that it be waged by a certain kinds of entity. In the following chapters I examine the prospects for a qualified defence of the authority criterion in its alternative, permissive role (though I will briefly return to the restrictive criterion in Chapter 5). Given that the truth of a reductivist view entails rejecting the permissive criterion, a rehabilitation of the criterion

will need to show that one of the claims commonly associated with reductivism is in fact mistaken.

## Chapter 3 – Non-Reductivist Approaches to Just War Theory

### 3.1 Introduction

As explained in the previous chapter, a reductivist approach generates powerful objections to including an authority criterion within just war theory. On this view, the fact that wars are waged by particular kinds of entity has no non-derivative moral significance for justifying either the resort to war or individuals' conduct within wars. These questions are wholly settled by the application of the moral principles that govern the use of force between private individuals.

Any defence of the authority criterion must therefore show that there is something wrong with the reductivist picture. To this end, in this chapter, I consider a rival 'non-reductivist' approach to the ethics of war. On this alternative view, reductivists are mistaken in holding that the morality of war is determined solely by considerations of liability and lesser-evil, on the ground that warfare differs from ordinary forms of interpersonal violence in some morally relevant respect. Proponents of non-reductivism typically locate this difference in the fact that war involves the use of force by groups of individuals engaged in morally significant relationships with one another. A central claim of non-reductivism is that these relationships are independently capable of generating permissions for killing and injuring in war, thus denying reductivism.

An assessment of a non-reductivist approach is particularly relevant for the purposes of investigating the prospects for a defence of the authority criterion. This is because non-reductivism potentially provides the resources for justifying the central thought that I have argued underpins the inclusion of the criterion within just war theory – that combatants in war enjoy additional permissions to cause harm in virtue of fighting on behalf of a certain kind of entity.<sup>106</sup>

---

<sup>106</sup> Jeff McMahan also suggests, albeit briefly, a possible connection between non-reductivism and the authority criterion, though he does so in terms of the standard restrictive interpretation of the criterion. McMahan, 'Just War', p.672.

However, I will argue that non-reductivism must be rejected. The argument works by pushing a dilemma onto non-reductivists: If non-reductivists are successful in showing that acting on behalf of a certain type of entity gives rise to additional *permissions* to cause harm in war, they must also jettison the most intuitive *restrictions* on conduct in war – most saliently, the prohibition on intentionally killing morally innocent non-combatants. Since this conclusion is unacceptable, we must conclude that non-reductivism fails.

### **3.2 Motivating Non-Reductivism**

One way of motivating a non-reductivist view is by arguing that a certain class of killing in war is intuitively permissible yet cannot be justified on grounds of liability or impartial lesser-evil, which are the sole units of currency for morally evaluating warfare on a reductivist view.

#### *3.2.1 The Permissibility of Participation in Unjust Wars*

As we have seen, one important aspect of a reductivist approach is that it rejects the view that the permission to kill in war applies equally to all combatants, independently of the justice of the wars in which they fight. The Equality Thesis is a central component of orthodox just war theory and the laws of war, yet has no analogue within ordinary interpersonal morality. More specifically, the permission to kill in an unjust war cannot be squared with liability-based justifications for killing. Reductivists hold that there is a genuine incompatibility here, which must be resolved by rejecting the permissibility of fighting in unjust wars. Doing so involves intentionally killing morally innocent individuals and is straightforwardly morally impermissible for that reason.

Many find the conclusion that mainstream just war theory and the laws of war lack principled moral foundations too revisionary to be accepted and have sought to defend the permissibility of fighting in unjust wars. One common strategy for doing so is to deny the premise underpinning the

reductivist critique – that warfare is morally continuous with all other forms of violence and justified solely in terms of aggregated liability and lesser-evil justifications. On this view, there need be no tension between the permissibility of fighting in unjust wars and standard forms of justification for interpersonal harming, since the morality of conduct in war is grounded in some further moral considerations, which reductivism fails to accommodate.

### 3.2.2 *The Avoidance of Pacifism*

A more general objection holds that reductivism cannot justify killing even in wars that are intuitively just and thus entails a form of pacifism. This objection is more troubling than the previous challenge, since it argues that reductivism cannot justify a class of killing which reductivists are not willing to accept as unjustified. The most forceful version of this objection has recently been put forward by Seth Lazar, who argues that the widespread intentional killing of combatants that is necessary to wage practically any war cannot be justified on the ground that those combatants are liable to defensive killing.<sup>107</sup>

Lazar derives this conclusion by arguing that reductivism faces a dilemma, resulting from the primacy it assigns to considerations of individual liability in justifying intentional killing in war. According to the dilemma, reductivists cannot consistently hold that the intentional killing of combatants in war is justified on the basis that those combatants have rendered themselves liable to defensive killing, while also maintaining a general prohibition on the targeting of non-combatants. One of these commitments has to go.

Reductivists typically accept that *some* non-combatants on the unjust side of a war will be liable to being killed, and so do not endorse an absolute prohibition on targeting non-combatants. For example, those who

---

<sup>107</sup> Seth Lazar, 'The Responsibility Dilemma for Killing in War: A Review Essay', *Philosophy and Public Affairs* 38, No.2 (2010), 180-213. For different arguments that a reductivist approach entails a form of pacifism, see Richard Norman, *Ethics, Killing and War* (Cambridge: Cambridge University Press, 1995); David Rodin, *War and Self-Defense*, Ch.6.

have made significant and culpable contributions to the initiation or continuation of an unjust war, such as war-mongering politicians, propagandists, weapons scientists, etc. But reductivists are generally committed to the view that the vast majority of non-combatants on an unjust side of a war may not be permissibly targeted, and explain this prohibition in terms of non-combatants retaining full possession of their normal right not to be killed. For example, in virtue of having not made sufficiently large causal contributions to the unjust threats posed by combatants, by lacking sufficient moral responsibility for their contributions, or because of the causal remoteness of their contributions.<sup>108</sup>

Lazar argues that this position cannot be squared with the reductivist position that intentionally killing combatants is primarily justified on the grounds that they have rendered themselves liable to defensive killing by participating in an unjust war. In order to argue that a sufficiently large proportion of enemy combatants are liable to be killed, reductivists must set the threshold for liability fairly low. This is required in order to render combatants liable who make relatively small contributions to unjust threats and who bear little culpability for doing so. Lazar claims that this will include a significant proportion of the members of militaries engaged in unjust wars.

However, the consequence of this necessarily low bar for liability, argues Lazar, is that it will also render a large proportion of non-combatants on the unjust side liable to defensive killing. There are insufficient differences between combatants and non-combatants in terms of their contributions to the war effort and their degree of responsibility for contributing that could ground a general difference in liability. Lazar argues that this implication is unacceptable and that in order to avoid it reductivists must instead endorse a higher threshold of contributory and moral responsibility for liability to defensive killing. The problem, however, is that this now leads to a highly pacifistic conclusion, since this more

---

<sup>108</sup> McMahan, *Killing in War*, Ch.5; Fabre, 'Guns, Food and Liability to Attack in War'; David Rodin, 'The Moral Inequality of Combatants: Why *jus in bello* Asymmetry is Half-Right' in Rodin and Shue (eds), *Just and Unjust Warriors*, 44-68.

demanding account of liability will now render large portions of unjust combatants non-liable. Herein lies the 'responsibility dilemma' for reductivism.

Lazar argues that since pacifism is also unacceptable, we must reject reductivism as the correct account of the morality of war. What the dilemma reveals is that the vast majority of wars, whether just or unjust, will involve the widespread intentional killing of non-liable individuals. Hence, if warfare is to be justified, it cannot be because the combatants that would have to be killed have rendered themselves liable to defensive harm. The correct theory of the morality of war must appeal to other moral considerations.

The jury is still out on the success of this objection. One line of resistance maintains that the dilemma can be denied, by arguing there are in fact morally relevant asymmetries between combatants and non-combatants in terms of the factors that ground liability.<sup>109</sup> Another response aims to show that modifying the relevant conception of liability can enable reductivists to draw the desired distinction between combatants and non-combatants.<sup>110</sup> A more radical view accepts the existence of the dilemma, but denies that the correct resolution requires rejecting reductivism. On this view, the correct conclusion is that many non-combatants on the unjust side of a war simply are liable to defensive killing.<sup>111</sup> The dilemma shows not that reductivism is false, but that the standard principle of non-combatant immunity is.

### 3.3 Non-Reductivist Approaches

The important point to note is that both lines of objection to reductivism surveyed in the previous section, though motivated by different concerns, are united in holding that reductivism cannot justify acts of intentional

---

<sup>109</sup> Jeff McMahan, 'Who is Morally Liable to be Killed in War', *Analysis* 71, No.3 (2011), 544-559. See also, Bradley Jay Strawser, 'Walking the Tightrope of Just War', *Analysis* 71, No.3 (2011), 58-66.

<sup>110</sup> Saba Bazargan, 'Complicitous Liability in War', *Philosophical Studies* 165 (2013), 177-195.

<sup>111</sup> Frowe, 'Defensive Killing', Chs.6-8.



killing that are, by hypothesis, morally permissible. For both, this conclusion shows that reductivists are mistaken to hold that the permissibility of killing in war is determined by the same considerations that determine the permissibility of killing in ordinary cases of interpersonal violence. More specifically, that reductivists are wrong to hold that liability is a near-necessary condition of justified intentional killing in war.<sup>112</sup>

This response expresses an alternative, non-reductive view of the morality of war. Non-reductivism encompasses a spectrum of possible positions, but the unifying idea is that warfare differs from standard cases of interpersonal violence in some morally relevant respect, so that our conclusions about the permissibility of killing in the latter cases cannot be straightforwardly imported into the former in the way that reductivists claim. Michael Walzer captures the spirit of this view nicely in his pithy objection to McMahan's reductivism: "What Jeff McMahan provides is a detailed account of how killing in war would be permissible if war were a peacetime activity."<sup>113</sup> While many are intuitively attracted to a non-reductivist view, the crucial task for non-reductivists is to identify a relevant differentiating feature of warfare that is both morally significant and supportive of a more extensive account of permissible killing in war.

Proponents of non-reductivism typically locate the relevant moral difference in war's collective and political character. They point out that war does not simply involve otherwise-unconnected private individuals using force against one another, as in the simplified self-defence cases that reductivists often draw on in deriving conclusions about the morality of conduct in war. Rather, war involves the use of force by groups or associations of individuals engaged in morally significant relationships with one another. War, on this view, is essentially an activity that people do

---

<sup>112</sup> The 'near' qualification is necessary, since reductivists are generally not absolutists regarding the constraint on intentionally killing non-labile individuals. Reductivists typically hold that the constraint on intentionally killing the non-labile is much weightier than the constraint on unintentionally but foreseeably killing the non-labile, so that intentional killing could be justified as the lesser-evil provided that the evil averted were sufficiently great to outweigh the more stringent constraint on intentional killing.

<sup>113</sup> Michael Walzer, 'Response to McMahan's Paper', p.43.

together.<sup>114</sup> Reductivists, it is charged, are blinded to this important feature of war by their exclusive focus on individuals.<sup>115</sup>

### 3.3.1 *The Collectivist View*

This focus on collectives rather than individuals continues a tradition of thinking about war with origins in Rousseau's conception of political community and war. On Rousseau's view, a legitimate polity is constituted not by any external markers but by an internal relation of its citizens' wills. When the polity acts, through the coordinated action of its citizens in accordance with the general will, this act has a normative personality over and above the sum of its component actions. War is a paradigmatic example of this form of action. It is this internal relation of wills which distinguishes war from merely private violence such as duels, riots, feuds and brigandry. War, properly understood, is an activity that individuals simply cannot perform, "War...is not a relationship between one man and another, but between one State and another."<sup>116</sup>

Christopher Kutz has developed a version of non-reductivism based on this collectivist understanding of war, which aims to defend a more orthodox conception of permissible conduct in war, which grants a permission to kill to combatants who participate in both just and unjust wars.<sup>117</sup> For Kutz, the norms governing conduct in war must reflect the fact that when soldiers confront each other on the battlefield they do so not merely as private individuals, but as citizen-representatives of their

---

<sup>114</sup> For examples, see Fletcher, *Romantics at War*; Christopher Kutz, 'The Difference Uniforms Make: Collective Violence in Criminal Law and War', *Philosophy and Public Affairs* 33, No.2 (2005), 148-180; Seth Lazar, 'Associative Duties and the Ethics of Killing in War', *Journal of Practical Ethics* 1, No.1 (2013), 6-51; Seth Lazar, 'Necessity in Self-Defense and War', *Philosophy and Public Affairs* 40, No.1 (2012), 3-44; Tamar Meisels, 'In Defence of the Defenceless: The Morality of the Laws of War', *Political Studies* 60, No 4 (2012), 919-935; Robert Sparrow, 'Hands Up Who Wants to Die: Primoratz on Responsibility and Civilian Immunity in War', *Ethical Theory and Moral Practice* 8, No.3 (2005), 299-319; Zohar, 'Collective War and Individualist Ethics: Against the Conscriptio of Self-Defense'.

<sup>115</sup> This criticism of reductivism bears an interesting resemblance to communitarian objections to liberalism.

<sup>116</sup> Jean Jacques Rousseau, 'The Social Contract' cited in Gregory Reichberg, Henrik Syse and Endre Begby (eds), *The Ethics of War: Classic and Contemporary Readings* (Oxford: Blackwell, 2006), p.482.

<sup>117</sup> Kutz, 'The Difference Uniforms Make'.

communities engaging in collective political action. On this view, the permission to kill in war is grounded in the collective and political character of the activity and not, as reductivists argue, in the liability of combatants *qua* individuals. Combatants in war possess what Kutz terms an “essentially political permission” to harm their opponents.<sup>118</sup>

This collectivist strategy allows Kutz to reassert the idea that the permission to kill in war obtains independently of considerations of *jus ad bellum*. Since combatants on all sides of a conflict fight on behalf of their political collective and this fact obtains independently of the justice of their cause, each possesses a permission to fight and kill in an unjust war, even though such actions would be “infamous crimes” if committed privately.<sup>119</sup> On this view, the clash that reductivists perceive between the traditional norms of permissible conduct in war and considerations of individual liability can be explained away as resting on an under-appreciation of the normative significance of collective political action.

It is worth pointing out that although Kutz’s collectivist view is primarily intended to provide a defence of one specific class of intentional killing in war that is not justified on the basis of liability – the killing of just combatants by unjust combatants – it also provides the resources for a defence of killing in war more generally. If, as Lazar claims, participation in both unjust *and* just wars involves the widespread intentional killing of non-liable combatants, the collectivist permission to kill can also be invoked to cover the latter class of cases as well.

Reductivists are, naturally, rather suspicious of this line of argument. For McMahan, it is straightforwardly bizarre, trading on “a form of moral alchemy.”<sup>120</sup> For how can it be, he argues, that simply by forming a particular type of bond among themselves and together pursuing certain kinds of political goals, that members of a community can exempt themselves from the duties which would otherwise be imposed by the basic rights of non-members?

---

<sup>118</sup> Ibid. p.153.

<sup>119</sup> Ibid.

<sup>120</sup> Jeff McMahan, ‘Collectivist Defences of the Moral Equality of Combatants’, *Journal of Military Ethics* 6, No.1 (2007), 50-59 at p.53.

A version of non-reductivism as strong as this certainly does seem implausible. However, the view can be refined in at least two directions which may render this objection less than decisive. Crucially, the most plausible versions of non-reductivism will be those which do not attempt to deny the reductivist insight that fighting in an unjust war involves violating the rights of one's opponents. Rather, they will seek to show that the collective character of warfare is sufficiently morally important to support a norm permitting the intentional killing of non-labile combatants *in spite of* the wrongdoing that this involves. In fact, Kutz quite clearly indicates that his view is restricted in this manner, acknowledging that "the unjust combatant morally wrongs who he kills."<sup>121</sup>

The first refinement limits the collectivist view to offering a principled (rather than pragmatic) defence of the legal norms governing conduct in war. In particular, by providing support for the immunity from sanction the law grants to combatants who participate in unjust wars. There are many ways in which an agent can be legitimately exempted from the law's sanction without it needing to be shown that the agent acted with justification. For example, the broad legal defence of excuse functions to defeat liability to sanction in precisely this way.

If the tricky notion of a collectivist 'political permission' is interpreted as justifying a special type of legal defence, grounded in the normative significance of collective political activity, then the traditional legal permission to kill in war may be defended against the reductivist challenge without having to make the difficult claim that the mere existence of political relationships can somehow render permissible otherwise wrongful conduct. This view can accept the reductivist claim that the moral status of acts of intentional killing in war are primarily determined by considerations of liability, but maintains that the collective character of war renders it morally inappropriate to respond to the killing of non-labile individuals in war with the same reactive practices and attitudes that are appropriate in cases of private violence.

---

<sup>121</sup> Kutz, 'The Difference Uniforms Make', p.173.

One way of making this argument, which Kutz suggests in several passages, is to understand the normative significance of collective political action as taking place at the level of how acts committed as part of a collective political action are attributable to specific individual agents for the purposes of determining their liability to blame and sanction. The thought is that, in contrast to private wrongdoers, those who inflict equivalent wrongs as participants in a collective political project bear a different normative relationship to those acts, in such a way that renders the usual attribution of responsibility for specific wrongs to particular agents inappropriate.<sup>122</sup> In these cases, the wrongs are attributable to the collective as a whole, and perhaps its members *qua* members, but not to individual combatants *qua* private individuals.<sup>123</sup> On this view, the fact that agents do wrong as part of a collective political activity provides an important exculpatory consideration, which justifies the legal practice of not criminalising killing in war, including the intentional killing of non-labile combatants.

### 3.3.2 *The Associative Duty View*

A non-reductive view may also be refined in a second direction, which aims to preserve the stronger claim that the collective nature of warfare is able to render it morally permissible for combatants to intentionally kill their non-labile combatants. This view accepts that such killing involves a serious injustice, but denies that this is sufficient to render so doing impermissible. On one way of arguing for this position, the negative duty not to intentionally kill the non-labile, while certainly stringent, may be overridden by weightier moral considerations, rendering the rights violation all-things-considered justified.

In a very recent article, Seth Lazar argues that certain features of collectives can generate these weighty reasons.<sup>124</sup> More specifically, such

---

<sup>122</sup> Ibid, p.179.

<sup>123</sup> On this point, see Judith Lichtenberg, 'How to Judge Soldiers Whose Cause is Unjust', in Rodin and Shue (eds), *Just and Unjust Warriors*, 112-129 at pp.125-129.

<sup>124</sup> Lazar, 'Associative Duties and the Ethics of Killing in War'.

reasons arise from the value of a range of significant relationships that combatants share with many of their fellow group members, such as those that obtain between family members, loved ones, friends, colleagues, comrades-in-arms and co-citizens. Lazar emphasises that participants in these relationships incur extremely stringent moral duties to protect their associates from harm.

Crucially, for Lazar, these associative duties to protect may be weightier than the negative duty not to harm the non-labile, so that when these duties come into conflict, harming the non-labile may be justified as the lesser breach of duty. Lazar's argument for this conclusion is sophisticated and lengthy, but the underlying thought appeals to our intuitions about cases in which the duty to protect appears to trump the duty not to kill the non-labile.<sup>125</sup> For example, in a scenario in which a parent is able to divert a falling boulder that will otherwise crush her child onto a non-labile bystander, many find it intuitive that the parent would be justified in doing so, in virtue of their special relationship with their child. Extrapolating from such cases, Lazar argues that killing non-labile combatants in war can be justified on the same basis, since such killing may be necessary to protect one's associates from serious harm. As he puts it,

sometimes we can protect our associates only if we fight and kill. We have duties to protect our associates, grounded in the value of these special relationships. Our armed forces are the executors of those duties. When they fight, those duties may clash with the rights that they must violate to win the war. In some cases, the associative duty to protect can override those rights, thus rendering some acts of killing all things considered justified."<sup>126</sup>

The 'associative-duty view' is primarily intended to provide an account of how intentional killing in *ad bellum* justified wars can be justified, since, for

---

<sup>125</sup> Lazar supports his view with an argument from transitivity. Ibid. pp.19-30.

<sup>126</sup> Ibid. p.9.

Lazar, such killing cannot be justified in terms of liability. However, as Lazar recognises, the associative-duty view also enables a qualified defence of the permissibility of participating in unjust wars, since fighting and killing may be necessary to protect one's associates whether or not one's war is justified or unjustified.<sup>127</sup> On the associative-duty view, the facts that ground the permission to kill in war may obtain independently of considerations of *jus ad bellum*.

The associative-duty view can be contrasted with the collectivist view in terms of the precise challenge that each poses to reductivism. A collectivist view may be characterised as denying the fundamental reductivist commitment to the Continuity Thesis, which holds that the morality of war is exhausted by the same principles of interpersonal morality that govern the use of force between private individuals. For collectivists, war, as a manifestation of collective political agency, is genuinely morally *sui generis*.

The associative-duty view, on the other hand, can accept the Continuity Thesis. Instead, it challenges the reductivist Content Thesis, which holds that considerations of liability and impartial lesser-evil exhaust the range of justifications for non-consensual harming and killing contained within interpersonal morality. For Lazar, interpersonal morality includes an additional class of justification, arising from the comparative stringency of one's duties to protect one's associates. On this view, war is not morally special in any strong sense, it is simply a case in which these duties most strikingly come into conflict with the weaker duty not to intentionally kill those who are non-liable.

### **3.4 The Relevance of Non-Reductivism**

As argued in Chapter 1, the inclusion of the authority criterion within just war theory serves to capture the intuition that those who use force on behalf of certain kinds of entities are subject to additional permissions to cause harm. More specifically:

---

<sup>127</sup> Ibid. pp.45-46.

***The Permissive Authority Criterion:*** If an agent participates in war on behalf of an entity that possesses a certain set of properties, then that agent is subject to a more extensive range of permissions to cause harm than if they fought privately, or on behalf of an entity which lacks that set of properties.

Non-reductivist approaches to the ethics of war are particularly interesting for the purposes of an investigation into the defensibility of the authority criterion, since these views can be interpreted as providing a precise account of both what these additional permissions consist in and the relevant properties that give rise to them.

On both variants of non-reductivism canvassed above, the fact that an agent uses force as a member of a certain type of group or association gives rise to a permission or justification to intentionally kill opposing combatants in war in the absence of a liability or impartial lesser-evil justification for doing so. On a collectivist view, the relevant properties that ground this permission (or exculpation) are the group's ability to exercise collective political agency. On the associative-duty view, the relevant properties are the weighty duties to protect that obtain between participants in certain forms of valuable relationship.

The success of a non-reductionist view thus provides the resources for a qualified defence of the authority-criterion, in its permissive role, against the sceptical challenges surveyed in Chapter 2. However, in the following two sections I will argue that such views are unsuccessful.

### **3.5 The War Crimes Objection**

I offer a different objection to non-reductivism than that standardly put forward by reductivists. I argue that by its own lights and granting its major assumptions about the moral significance of collectives or associative relationships, a non-reductive approach fails to provide an acceptable account of permissible conduct in war. In particular, non-reductivists cannot provide a convincing account of the fundamental idea that certain



classes of person are immune from attack in war. The objection is intended to apply, *mutatis mutandis*, to each of the variants of non-reductivism canvassed above.

To date, the debate between reductive and non-reductive approaches to war has largely centred on the justification of intentionally killing opposing combatants. This focus makes perfect sense. As we have seen, it is common intuitions about this permission that have proved most problematic under a reductive view. However, the norms of *jus in bello* are far from exhausted by the permission to kill. The majority of these rules are prohibitive in character, most notably the restrictive component of the requirement of discrimination, which prohibits attacks on certain classes of person in war. This is such a central and uncontroversial part of the rules of *jus in bello* that it is understandable that it does not receive as much attention as the debate over the permissibility of killing combatants. To paraphrase Walzer, the hard part in thinking ethically about war is not to explain why certain people ought not to be deliberately killed. It is to show how anyone can be.<sup>128</sup>

Despite the consensus that attacks on certain persons are prohibited, there is considerable debate among just war theorists as to precisely which persons are immune from attack under the correct principle of discrimination. While virtually all accounts of discrimination start from the truism that clearly innocent people ought not to be deliberately targeted in war, there is dispute as to how the notion of innocence is to be understood and, most importantly, how innocence can be lost.

Disagreements about these two issues yield significant differences in how the distinction between legitimate and illegitimate targets in war is drawn. For example, the disagreement between Walzer's and McMahan's positions on how innocence is lost via liability to defensive force leads each to endorse a very different conception of the discrimination requirement. Walzer's threat-based account of liability yields an endorsement of the standard view of discrimination in war as drawn, roughly, between combatants and non-combatants. By contrast McMahan's account, which

---

<sup>128</sup> Walzer, *Just and Unjust Wars*, p.41.

focuses on moral responsibility for the existence of threats of unjustified harm as the basis of liability, yields a more complex and revisionary principle of discrimination. It takes the correct distinction to be that combatants who lack a just cause are legitimate targets, whereas just combatants and non-combatants on both sides are illegitimate targets.<sup>129</sup>

Despite the practical disagreement between these two influential views, it is important to note the similarities in the approach each takes to the question of discrimination. For both, it is impermissible to target persons who have done nothing to lose their normal immunity from attack. The dispute arises from competing conceptions of innocence and liability, not from disagreement as to the relevance of these concepts to the permissibility of targeting. In cases where innocence is clear and uncontroversial, such as young children, all the main views are in agreement that targeting these individuals is prohibited under the principle of discrimination and that the reason for this is that such individuals have done nothing to lose their normal right against being attacked. We may term this straightforward approach to discrimination a *target-centred* account. It explains why attacks on certain persons are prohibited or permitted by reference to certain facts about those persons. This account also underpins a commonsense understanding of why certain acts in war, such as the deliberate killing of non-combatants, are appropriately criminalised as war crimes. The offence of the war crime is grounded in the rights violation suffered by the innocent victim.

However, this straightforward and highly intuitive explanation of the restrictions on intentional killing in war is not available to non-reductivists. On the non-reductivist views considered earlier, the permissibility (or at the least non-criminality) of a certain kind of killing in war – the killing of

---

<sup>129</sup> This is a convenient oversimplification, in two respects. Firstly, as noted above and in fn.108, reductivists accept that *some* non-combatants on an unjust side may render themselves liable to defensive force by making certain significant kinds of contribution to their side's war effort (though, as we have seen, some dispute whether a general distinction between the combatants and non-combatants on grounds of liability can be maintained). Secondly, as noted in fn.96, reductivists also accept that combatants on a just side can render themselves liable, by threatening harms that are not justified by the aims of their war.

combatants by other combatants – is grounded not in the non-innocence or liability of the combatants who are harmed, but in certain facts about the combatants who inflict the harms. For Kutz, it is the fact that combatants act as agents of their collectives which explains their permission to kill opposing combatants. On Lazar’s view, it is combatants’ associative duties which do the normative heavy-lifting.

Since it is this fact about those who *inflict* harm in war, rather than facts about those who *suffer* harm, which determine whether such harming is permissible, we may understand non-reductivists as offering an *agent-centred* account of permissible killing in war. On such an account, the question of the liability of the target of harm does not play the same decisive role in determining whether inflicting that harm is permissible that it does on a standard, target-centred account of discrimination.<sup>130</sup>

As explained above, the attractiveness of an agent-centred view is clear. It enables a defence of the permission to fight in wars, including unjust wars, without having to argue that such action can be permitted on the basis of a target-centred, liability-based justification. However, the major difficulty facing an agent-centred view is that if it succeeds in supporting the permission kill non-liaible combatants, it seems unable to account for the foundational idea that there are restrictions on who may be targeted.<sup>131</sup> For if the permission enjoyed by combatants to kill their opponents, who have done nothing to lose their normal right not to be killed, rests on certain facts about the unjust combatants, then there seems no reason why targeting and killing other non-liaible persons in war, such as non-combatants, are not also rendered permissible on the basis of the very same facts. Consider the following pair of examples:

---

<sup>130</sup> I borrow the terms ‘target-centred’ and ‘agent-centred’ from McMahan, who employs them to distinguish different conceptions of the justification of self-defence. Jeff McMahan, ‘Self-Defense and the Problem of the Innocent Attacker’, *Ethics* 104, No.2 (1994), 252-290 at p.268.

<sup>131</sup> The novelty of this objection should not be overstated. McMahan briefly raises its possibility, but does not pursue it at any length. Jeff McMahan ‘Précis of *The Morality and Law of War*’, p.674. McMahan also pursues an analogous line of objection to agent-centred accounts of self-defence in his ‘Self-Defense and the Problem of the Innocent Attacker’, pp.270-271.

***Siege 1:*** Nation A, a fully-functioning political collective, wages an unjust war against Nation B with the aim of annexing an area of Nation B's territory. Nation B militarily opposes the invasion and a conventional war ensues, endangering the lives of combatants and non-combatants on both sides. At a pivotal stage in the war Nation A has one of Nation B's strategically vital cities surrounded. In order to win the war it is vital that the city be taken. Nation A's combatants can only do so by bombarding the city's armed defenders, killing 10,000 of Nation B's combatants.

***Siege 2:*** The circumstances are exactly the same as *Siege 1*, except that Nation A does not have the option of attacking the city's armed defenders. Instead, in order to take the city, Nation A's combatants can only do so by bombarding residential areas of the city, killing 3,000 of Nation B's non-combatants.

According to both the law of war and orthodox accounts of just war theory, combatants who engage in *Siege 1* would not act impermissibly nor be legitimately punished for doing so, since they act in accordance with the rules of *jus in bello*. In particular, their actions respect the principle of discrimination as standardly interpreted. Combatants who engage in *Siege 2*, by contrast, would be judged to act impermissibly. They clearly violate any plausible requirement of discrimination and may appropriately be held liable to prosecution for war crimes. I take it as uncontroversial that any adequate theory of the ethics of war must judge participation in *Siege 2* to be straightforwardly impermissible.

The problem non-reductivists face is that by explaining why participation in cases like *Siege 1* can be permissible in terms of certain agent-centred facts about the combatants who participate, they are unable to account for the impermissibility of participation in cases like *Siege 2*.

Non-reductivists such as Kutz and Lazar accept that Nation B's combatants are in the same situation as Nation B's non-combatants in terms of liability. The members of neither group have acted in such a way as to lose their rights not to be killed. Furthermore, all the relevant agent-centred

facts that non-reductivists may cite as grounding the permission to intentionally kill non-liaible individuals in *Siege 1* also obtain in *Siege 2*. In each case, Nation A's combatants will be acting as participants in a collective political project, and in each case participation is a necessary means for Nation A's combatants to discharge their associative duties to protect.

The problem is that if these agent-centred facts are capable of rendering participation in cases like *Siege 1* permissible, and obtain to the same degree in *Siege 2*, it seems that killing must also be permissible in *Siege 2*. Nothing in these agent-centred considerations provides a means for morally distinguishing between targeting non-liaible combatants and non-liaible non-combatants. In fact, *Siege 2* is morally preferable in one respect, since it involves significantly less killing of non-liaible persons than *Siege 1*. Yet, presumably, non-reductivists would want to resist the conclusion that participation in *Siege 2* is permissible.

It is important to remember that non-reductivists cannot make the obvious appeal to the non-liability of the targets as accounting for the impermissibility of targeting innocent non-combatants in war. To do so would be to concede that the target-centred fact of non-liability is determinative of the permissibility of intentional killing in war, and not the agent-centred facts. For if it is conceded that considerations of liability are determinative of the permissibility of harming in cases like *Siege 2*, there is no reason why this should not also be true in cases like *Siege 1*. This is precisely the view that non-reductivists set out to deny.

So, while often invoked in defence of certain aspects of a traditional understanding of just conduct in war, a non-reductive approach in fact seems to have deeply revisionary implications of its own that warrant rejecting the view. If successful in providing a defence of the permissive component of the *in bello* principle of discrimination the view lacks the resources to explain the more fundamental restrictive component of the principle. Term this the 'war-crimes objection' to non-reductivism. The problem flows from the fact that prohibitions on targeting are most easily and naturally explained in terms of the non-liability of the target of harm.

Yet a non-reductive view is committed to rejecting the idea that this fact is decisive in determining the permissibility of intentional killing in war.

The war crimes objection is a particular problem for non-reductivists such as Lazar, who seek to motivate the view by arguing that a reductivist approach cannot justify warfare without also jettisoning the idea of non-combatant immunity. If, as Lazar claims, a failure to accommodate this principle is sufficient for rejecting an approach to the morality of war, then non-reductivism must be abandoned.

### **3.6 Alternative Groundings for Constraints on Targeting**

What the war-crimes objection highlights is that the acceptability of a non-reductive approach rests not only on whether it can provide an account of the permission to kill in war, but also, more importantly, on whether it can provide a similarly non-reductive account of the most intuitive *in bello* restrictions. To avoid the objection, a plausible distinction between legitimate and illegitimate targeting in war needs to be drawn on the basis of facts that non-reductivists can accept as determinative of the permissibility of killing. This requires a revisionary account of the basis of the prohibition on targeting non-liable non-combatants.

As it stands, the war-crimes objection holds that it would be arbitrary for non-reductivists to appeal to non-liability in order to explain the impermissibility of participation in cases like *Siege 2*, while arguing that the agent-centred facts trump considerations of non-liability in cases like *Siege 1*. In response, non-reductivists may object that this poses a false dilemma, resting on the assumption that the target-centred fact of non-liability is the sole consideration capable of accounting for the intuitive prohibitions on targeting in war. Perhaps these prohibitions can be grounded in considerations other than non-liability. If such an argument can be made, then non-reductivists may be able to avoid the war-crimes objection.

On this view, the agent-centred considerations discussed above should be understood as providing necessary, but not sufficient, conditions for the permissibility of targeting non-liable persons in war. The explanation of the

intuitive prohibitions on targeting will arise through an elucidation of the further conditions required for sufficiency. This response accepts that both the combatants and non-combatants in *Siege 1* and *Siege 2* are not liable and that the agent-centred facts that ground the non-reductive permission obtain equally in both cases. But it aims to resist the conclusion that it is therefore permissible to participate in both cases by arguing that there are *additional* considerations – either agent-centred or target-centred – that may render participation in *Siege 2* impermissible, but which do not oppose participation in *Siege 1*.

To be successful, such a response must meet two requirements. It must identify an additional consideration which gives rise to a genuine and sufficiently weighty moral constraint on harming. It must then be able to draw the desired distinction between legitimate and illegitimate targets on the basis of this consideration. This section considers several possible responses of this type and argues they fail to provide a plausible account of the prohibition on targeting non-combatants in war.

### *3.6.1 Appealing to additional agent-centred facts I: Role-based obligations*

One plausible, agent-centred, candidate for the prohibition on targeting non-combatants appeals to a professional role-based obligation on the part of combatants to refrain from targeting non-combatants.<sup>132</sup> Members of many professions are considered subject to stringent constraints on their behaviour as a result of taking up their role. The military profession may plausibly be considered a paradigmatic source of such obligations. Coupled with the assumption that a prohibition on targeting non-combatants is a central role-based obligation for military professionals, a response to the war crimes objection becomes available. On this view, the permission to kill combatants and the prohibition on targeting non-combatants are explained by the fact that soldiers have a stringent professional obligation to refrain from the latter but not the former.

---

<sup>132</sup> Thanks to Jeff McMahan for suggesting this line of response.

However, this response is susceptible to a pair of related objections. The first is that it makes the prohibition on targeting non-combatants entirely contingent on whether the military organisation in question actually endorses a professional code that prohibits such action. For example, imagine a society that endorses an alternative (but not wildly implausible) professional military ethic that does not include a strict prohibition on targeting non-combatants. Instead, it endorses a professional code which requires that targeting decisions be made so as to minimise the overall number of non-liable people who are harmed. Now imagine a variation on the *Siege* examples in which Nation A has the option of waging their war by engaging in either *Siege 1* or *Siege 2*. This alternative military code would place combatants under a professional obligation to engage in the latter rather than the former. The role-based response to the war crimes objection seems committed to accepting that members of such a military organisation would not possess a role-based obligation to refrain from targeting non-combatants and would therefore not be prohibited from participating in *Siege 2*. But this is deeply unpalatable. Few, I take it, would want to defend the view that it is impermissible for soldiers to target non-liable non-combatants solely on the ground that those soldiers just happen to be members of an organisation whose conventions forbid such action.

If this implication is to be avoided, the role-based obligation to refrain from attacks on non-combatants needs to be grounded in some relevant moral constraint that is not itself conventional. However, while this move may help alleviate the charge of contingency, it does so at the cost of a second objection, which is that the appeal to role obligations becomes redundant. For now, on this revised view, the actual source of the obligation not to target non-combatants lies in a moral constraint that is external to the professional role. The role itself no longer does any independent moral work.<sup>133</sup> The deeper problem highlighted by this redundancy arises when we consider what external considerations could

---

<sup>133</sup> See also, A. John Simmons, 'External Justifications and Institutional Roles', *Journal of Philosophy* 93, No.1 (1996), 28-36.



be capable of explaining the prohibition on participation in Option 2. One candidate seems obvious: the target-centred fact that the non-combatants are not liable to attack. But, as explained above, non-reductivists are excluded from this intuitive explanation.

The role-based response to the war crimes objection is then, at best, unsatisfactorily contingent, or, at worst, no response at all.

### *3.6.2 Appealing to additional agent-centred facts II: The avoidance of cowardice*

Lazar suggests an alternative agent-centred constraint on targeting non-combatants, which is not susceptible to the charge of contingency. On this view, the prohibition on targeting non-liable non-combatants lies in the fact that such action manifests the vice of cowardice, whereas the targeting of non-liable combatants does not. It is the avoidance of this vice which imposes the additional moral constraint required to avoid the war-crimes objection.<sup>134</sup>

There are, however, three problems with this response. The first is that it seems to get the order of explanation backwards. It seems much more natural to say that actions are cowardly because they are wrong, rather than wrong because cowardly. To accuse someone of cowardice is to provide an evaluation of their character – to claim that they culpably failed to respond appropriately to the reasons they had for action. It is not to claim that there existed an additional moral reason not to act as the agent did, which the agent failed to heed.

Secondly, a problem for the argument-from-cowardice lies in the *inwardness* of the moral constraint to which it appeals. On this view, combatants ought not to target non-combatants because of the effect that this would have on the combatants themselves. But the fact that acting in a certain way would be bad *for the agent* does not usually impose a moral obligation on that agent not to act. Common-sense morality permits agents to make self-sacrifices in pursuing their aims, provided that the aims are

---

<sup>134</sup> Lazar, 'Associative Duties and the Ethics of Killing in War', p.39

not morally prohibited on independent grounds. So, even if targeting non-combatants is cowardly and combatants incur a moral cost by doing so, this cost should not, in itself, morally prohibit the agent from acting.

Thirdly, if we grant that the avoidance of cowardice provides a weighty reason for action (weighty enough to justify the killing of an additional 7,000 non-labile persons in the *Siege* cases), then we get some very counter-intuitive results in other cases. For example, the argument implies if a group of combatants have a choice between attacking a small number of enemy combatants who they greatly outnumber, or attacking a much greater number of combatants in order to achieve the same military goal, they have a very strong cowardice-based reason to attack the larger number, despite the fact that many more combatants will be killed as a result of doing so. This seems highly implausible.<sup>135</sup>

### *3.6.3 Appealing to additional target-centred facts I: Modes of agency*

Appeals to additional agent-centred considerations have proved insufficient to ground the required prohibition on targeting non-combatants. A more promising approach aims to ground the prohibition in the effect that such harming has on its targets, over and above the serious rights violation suffered. This move appeals to additional target-centred considerations, other than that of non-liability.

One response of this type appeals to the moral distinction between acts of intentional harming that are *manipulative* – that involve using their victim as a means – and those that are *eliminative* – that are inflicted in order to remove a threat that their victim poses or contributes to. Many believe that acts of manipulative harming are subject to a higher justificatory burden than acts of eliminative harming, since they involve a more objectionable *mode of agency*. Other things being equal, one suffers a graver wrong if one is unjustly harmed manipulatively rather than eliminatively.

---

<sup>135</sup> Thanks to Daniel Viehoff for this example.

If we grant that the distinction between eliminative and manipulative harm has moral significance, a response to the war crimes objection becomes available. One may argue that the agent-centred facts that ground the non-reductive permission are sufficiently weighty to permit eliminatively killing the non-liable, but not manipulatively.<sup>136</sup> Combined with the descriptive claim that targeting non-combatants in war functions manipulatively, whereas targeting combatants involves only eliminative killing, one can non-arbitrarily limit the scope of the non-reductive permission to the killing of non-liable combatants only.<sup>137</sup>

However, there are two problems with this response, which have been recognised by non-reductivists.<sup>138</sup> The first is that targeting civilians may often function eliminatively. Imagine that the civilian areas to be targeted in *Siege 2* are populated by industrial and financial workers and that killing them will contribute to Nation A's victory by removing the contribution that they make to Nation B's war effort. In such cases, then the constraint imposed on the non-reductive permission by the eliminative/manipulative distinction would not rule out the targeting of non-combatants.

This is a particular problem for those such as Lazar, who aim to motivate a non-reductivist view by arguing that reductivism renders too many non-combatants liable, on the ground that non-combatants are causally and morally implicated in the existence of unjust threats to the same degree as many combatants. If this is true, the intentional killing of many non-combatants will function eliminatively, in the same way that killing combatants functions to prevent them from contributing to future threats. If this is false, then reductivists do not face the dilemma that Lazar claims that they do.

Secondly, and conversely, it may often be the case that targeting combatants functions manipulatively. The deaths of these combatants may be intended to shock and coerce their side's political leaders or civilian population into ending the war. It is also plausible that targeting

---

<sup>136</sup> Jonathan Quong employs an analogous argument in order to constrain his agent-centred account of permissible self-defensive killing. See Quong, 'Killing in Self-Defense'.

<sup>137</sup> Lazar, 'Associative Duties and the Ethics of Killing in War', p.39

<sup>138</sup> Ibid.

combatants will routinely function manipulatively, in order to persuade military commanders to surrender or adopt an alternative strategy more amenable to the opposing party. Given such cases, the proposed constraint on the non-reductive permission may often prohibit the targeting of combatants, undermining the ability of the non-reductive view to support the permission to kill non-liable combatants in war.

#### *3.6.4 Appealing to additional target-centred facts II: Risk*

A different target-centred approach aims to locate a response to the war-crimes objection in the moral significance of the uncertainty and moral risk that accompanies killing in war.<sup>139</sup>

On this view, intentionally killing a non-liable person is certainly a grave moral wrong, but the wrong is greater the higher the *ex ante* probability that that individual targeted is non-liable. Other things being equal, killing an individual when one is certain that they are not liable constitutes a more serious disrespect for their moral status than killing an individual when one is unsure of their non-liability.

Conjoining this idea the descriptive claim that non-combatants are much more likely to be non-liable than combatants generates a possible means of generating the prohibition on targeting non-combatants in war. It may be argued that the agent-centred facts that give rise to the non-reductive permission are sufficiently morally weighty to permit the killing of non-liable combatants – because the probability that such targets are non-liable falls below some threshold – but not weighty enough to render the targeting of non-liable non-combatants permissible – because the probability of these targets being non-liable falls above that threshold.

However, even granting the normative claim, this response does not seem to provide the desired result in the *Siege* cases, in two respects. Firstly, the descriptive claim that combatants “are undoubtedly more likely to be liable than non-combatants” is not true when the combatants in

---

<sup>139</sup> Ibid.

question are participating in a war that is *ad bellum* justified.<sup>140</sup> In *Siege* neither Nation B's combatants or non-combatant members have behaved in a way that could form the basis of liability. Hence, considerations of moral risk will not favour the targeting of one of these sub-populations over the other.

Secondly, when numbers are taken in consideration, the argument-from-risk may not in fact weigh against the targeting of non-combatants. To demonstrate, consider a variation on the *Siege* cases in which Nation B's combatants are more likely to be liable than Nation B's non-combatants. Let's also assume, generously, that if Nation A's combatants participate in *Siege 1*, every act of intentional killing runs a 0.3 chance of killing a non-liable person, whereas if they participate in *Siege 2* every act of killing runs a 0.95 chance of killing a non-liable person.

The argument-from-risk aims to show that these facts can explain why the non-reductive permission does not permit participation in *Siege 2* while permitting participation in *Siege 1*, on the grounds that each act of killing in *Siege 2* is more morally objectionable than each act of killing in *Siege 1*.

However, this ignores the important fact that *Siege 1* involves more acts of killing than *Siege 2*. Taking this into consideration, it is true of each of Nation A's combatants that they will be taking part in the killing of 3,000 people who are near-certainly non-liable if they participate in Option 1 ( $10,000 \times 0.3$ ), whereas they will be taking part in the killing of 2,850 people who are near-certainly non-liable if they participate in Option 2 ( $3000 \times 0.95$ ). Assessed from this perspective, considerations of moral risk would not seem to favour targeting combatants over non-combatants in the revised *Siege* cases.

### 3.6.5 *Appealing to additional target-centred facts III: Vulnerability*

One further proposal, endorsed by several authors, appeals to the fact that non-combatants in war are especially defenceless or vulnerable to harm as

---

<sup>140</sup> Ibid.

grounding the prohibition on targeting them.<sup>141</sup> The basic idea is that there is something significantly and independently morally objectionable about harming a person who is vulnerable, which gives rise to a general humanitarian duty to refrain from so acting.

If defensible, this may provide a means of avoiding the war crimes objection. Non-reductivists may argue that while the relevant agent-centred considerations have priority over considerations of non-liability in determining the permissibility of targeting in war, they remain insufficiently weighty to trump the duty not to harm the vulnerable. This duty then limits the range of actions that a non-reductive view can permit in war. Since, by hypothesis, non-combatants typically exhibit the duty-generating property of vulnerability in wartime whereas combatants generally do not, only attacks on combatants are rendered permissible on the basis of agent-centred considerations, whereas attacks on non-combatants are not. Interestingly, the appeal to vulnerability also provides a means of rehabilitating the Walzerian thought that the permission to target combatants is grounded in the fact that combatants pose threats, since presumably the ability to pose a threat negates an individual's vulnerability, even if does not vitiate their right not to be killed. The invocation of vulnerability also provides a way of substantiating the argument-from-cowardice discussed above – the thought being that the constraint on acts of harming that manifest cowardice is explained in terms of the constraint on attacking the vulnerable.

This response can be challenged on two fronts. Firstly, we can question whether the property of vulnerability does generate an additional and significant constraint on harming. One difficulty is that there seem to be clear cases in which the fact that the target of harm is vulnerable seems to provide no moral reason against inflicting harm, much less rendering it morally impermissible. Consider the following example:

---

<sup>141</sup> For versions of this view, see Lazar, 'Associative Duties and the Ethics of Killing in War', pp.40-41; Larry May, *War Crimes and Just War* (Cambridge: Cambridge University Press, 2007), pp.67-117; Meisels, 'In Defence of the Defenceless: The Morality of the Laws of War'; Henry Shue, 'Torture', *Philosophy and Public Affairs* 7, No.2 (1978), 124-143.

**Crime Boss:** A crime boss, who is paralysed from the neck down, offers a cash reward to anyone who assassinates a crusading government prosecutor. The prosecutor's bodyguards learn of the plot and can prevent the assassination in one of two ways. One involves confronting and killing each assassin. The other involves killing the crime boss in his hospital bed, thus preventing the assassination by removing the possibility of reward.

In this case it seems clear that the crime boss would be highly vulnerable to the harm inflicted by the bodyguards if they take the latter option, yet there is no reason to think that this renders killing him impermissible, or even counts against killing him. Such cases provide counter-examples to the claim that the property of vulnerability has sufficient moral significance to ground a prohibition on certain types of killing in war.

However, there is a plausible response available here. It may be argued that the moral force of the property of vulnerability is *conditional* on the target also being non-labile.<sup>142</sup> On this revised view, it is the target-centred properties of vulnerability and non-liability *in conjunction* that constrain the agent-centred permission to kill in war. This would explain why the fact of vulnerability does not render it impermissible to kill the crime boss, but could decisively count against participation in Option 2 in *Siege*. However, there are cases which tell against the moral significance of vulnerability even when the targets of harm are also non-labile. Consider the following example:<sup>143</sup>

**Bandits:** A gang of bandits plan to steal the possessions of a group of innocent villagers, thereby reducing the villagers to poverty. The bandits can achieve their aim in one of two ways. The first involves lethally targeting 10 villagers at random. This option gives the victims a 5% chance of survival, since the villagers may be able to hide from their attackers. The second involves providing the villagers with 10 slingshots and then lethally targeting the 10 villagers who opt to arm themselves. Again, this option

---

<sup>142</sup> Lazar p.42. An alternative view would be to make the significance of vulnerability conditional on the target not being culpable, independently of considerations of liability.

<sup>143</sup> Based on a case raised by Jeff McMahan in discussion.

gives the victims a 5% chance of survival, since the villagers may be lucky enough to avoid being killed by slinging gravel into their assailants' eyes.

If it were true that attacks on individuals who are both vulnerable and non-liable are significantly and distinctly morally worse than attacks on individuals who exhibit only the property of non-liability, we would expect to judge that the bandits to act much more objectionably if they successfully kill 10 villagers via the first option rather than the second. But this isn't the case. Both options intuitively seem morally on a par. This result gives reason to doubt that the property of vulnerability has the moral significance required to constrain the non-reductive permission to kill in war and thereby provide a response to the war crimes objection.

Of course, perhaps one does not share this intuition. However, even if we grant that vulnerability has independent moral significance, it remains unlikely that it is sufficiently weighty to enable a defence of non-reductivism. To demonstrate this, consider the following variations on the *Bandits* example:

***Bandits 2:*** Exactly the same as *Bandits*, except that the second option (slingshots) involves killing one additional villager.

***Bandits 3:*** Exactly the same as *Bandits*, except that the 10 villagers targeted in the first option (hiding) have a 15% chance of survival.

In these variations it seems uncontroversial that the bandits would act more wrongfully, all-things-considered, if they opt for the second option rather than the first. This result suggests that whatever moral significance vulnerability may have, it can be overridden by relatively small variations in other morally relevant factors. Given this, it is hard to see how appealing to an independent constraint on harming the vulnerable can provide an alternative basis for the standard prohibition on targeting non-combatants in war, since this prohibition is meant to be extremely stringent.



A second type of challenge may be raised over whether the properties of vulnerability and non-vulnerability, even if sufficiently morally significant, successfully pick out the required distinction between non-combatants and combatants. The concept of vulnerability obviously requires much greater elaboration, but it seems natural to understand the core idea as grounded in an individual's inability or powerlessness to avoid threats of harm. There are at least two ways in which an individual may suffer from this inability. Most obviously, an individual may lack means of defence against a threat. Alternatively, an individual may lack an ability to remove themselves from a threat of harm, by fleeing for example. The intuition driving the distinction between combatants and non-combatants in terms of vulnerability rests on the seemingly plausible assumption that non-combatants are relevantly vulnerable to a far greater extent than combatants. This assumption explains why the humanitarian duty not to harm the vulnerable prohibits only attacks on the former and not the latter, thus providing a response to the war crimes objection.

However, it is not obvious that this is the case. Attacks on many combatants seem to constitute breaches of a humanitarian duty not to harm the vulnerable, if such a duty exists. Understanding vulnerability in terms of an inability to offer defence against harm, it seems clearly true that combatants are often completely unable to defend themselves from attack in war. In particular, consider the means by which modern wars are fought, at long range and with overwhelming force. The typical combatant harmed in war does not 'go down fighting'. Rather they are struck down by an unseen enemy whom they had no real possibility of defending themselves against. Consider, once more, the *Siege* cases. Imagine, plausibly, that participation in *Siege 1* involves shelling the enemy from distance. Does it really seem plausible to maintain that those targeted in *Siege 1* are not relevantly vulnerable, whereas those who would be targeted by the same methods in *Siege 2* are? Whether an individual is rendered vulnerable through an inability to resist being harmed seems to depend less on whether they fulfil the role of combatant or non-combatant and more on the manner in which force is used against them. The necessary

assumption that attacks on combatants rarely constitute a breach of a duty not to harm the vulnerable seems to trade on an implausibly romanticised view of in war in which opponents give their enemies 'a fighting chance'.

Drawing the desired distinction between combatants and non-combatants is also problematic if we understand vulnerability in terms of an individual's inability to remove themselves from a threat of harm. This is because in many cases non-combatants will be better able to remove themselves from threats than combatants are. Whereas non-combatants may have the opportunity to flee the fighting and potentially become refugees, this option is less available to combatants who will often face serious, perhaps lethal, sanction for fleeing the threats they face in battle. The point here is not that non-combatants and refugees are not vulnerable, but that if they are (and surely they are) then many combatants will also be vulnerable to an equivalent or greater extent, since non-combatants often possess a degree of control over their fates which combatants lack.

The preceding observations are designed to show that even if we grant the existence of an independent duty not to harm the vulnerable, it is unlikely that appealing to this duty will assist non-reductivists in arguing, *contra* the war-crimes objection, that their view can support both the permissive and restrictive components of the traditional *in bello* principle of discrimination. If this duty is able to constrain a non-reductive view from permitting attacks on non-combatants, it will also often prevent the view from permitting the targeting of combatants.

To conclude this section, it is worth making overt a very general objection that is present in some of my preceding arguments. This is that even if, *contra* my arguments, one accepts that the considerations canvassed above provide genuine moral reasons in favour of targeting non-liable combatants over non-liable non-combatants, this is not sufficient to establish that non-reductivism can be made safe from the war crimes objection. In order to do so, what must be shown is not just that these considerations may provide reasons, but that these reasons are sufficiently weighty to prohibit the targeting of non-liable non-combatants even when other considerations – the numbers of innocent lives at stake most

obviously – weigh in favour of doing so. In other words, it is not enough to show that a candidate consideration is morally relevant to the distribution of harm among non-liable persons, capable of being a moral tie-breaker when all other factors held equal. Rather, it must be demonstrated that the consideration generates reasons to target non-liable combatants over non-liable non-combatants that are extremely weighty, comparable to the reasons we have to target liable individuals over non-liable individuals. It is no less a challenge to demonstrate that a candidate consideration has this moral weight than it is to show that it has intuitive moral significance in the first place.

### **3.7 Conclusion**

I have argued that non-reductivism should be rejected on the ground that it cannot support the most intuitive prohibitions on intentional killing in war. These are most naturally explained in terms of the target-centred fact of non-liability, yet non-reductivists are committed to the view that this fact is not generally sufficient to render targeting impermissible. This is especially problematic for those who aim to motivate non-reductivism by arguing that reductivism renders many non-combatants liable to defensive killing. While I have not attempted to show that reductivism has the resources to avoid this implication, or even that it has to avoid it – another thesis in itself – I aim to have shown that non-reductivism is not a viable alternative.

So where does this result leave my wider inquiry into the defensibility of the authority criterion? As explained in Section 3.4, non-reductivism potentially provides the resources for a defence of the authority-criterion, in its permissive role against reductivist sceptics. Rejecting non-reductive views as unacceptable on independent grounds provides indirect support for a reductivist view and therefore further strengthens the reductivist case in favour of jettisoning the criterion entirely. The discussion here also provides an adequacy condition on any potential defence of the authority-criterion – it must explain why acting on behalf of a certain kind of entity in war generates additional permissions to cause harm in war, but which does

not permit the widespread intentional killing of non-labile non-combatants. In the next chapter I begin the positive task of providing a qualified defence of the authority criterion.

## Chapter 4 – Authority and Harm

### 4.1 Introduction

To summarise the story so far: I began by arguing that a major source of motivation for the inclusion of an authority criterion within just war theory is the intuition that acting on behalf of a certain kind of entity makes an important difference to whether individuals act morally permissibly by participating in war. More specifically, the thought is that individuals who act on behalf of certain entities acquire additional permissions to cause harm, which private individuals, acting in relevantly similar circumstances, would lack.

I then set out an influential reductivist position within contemporary just war theory, which holds that the authority criterion must be jettisoned. This conclusion follows from a certain conception of the relationship between war and all other cases of interpersonal violence, combined with a particular account of the range of justifications for killing and injuring that apply in these cases. For reductivists, acts of non-consensual killing and injuring in war, as in any other context, are morally permissible only when justified on the basis of liability or lesser-evil. On this view, the moral limits of conduct in war are set by the limits of conduct between private individuals. Reductivists thus conclude that the central intuition underlying the authority criterion – that those who act on behalf of certain entities gain additional permissions to cause harm – cannot be defended.

In the previous chapter, I provided indirect support for this conclusion by arguing that the main rival to reductivism – which holds that moral permissions to cause harm in war can be grounded in the relationships that obtain between co-members in certain kinds of association – must be rejected.

In this chapter I develop the resources for a qualified defence of the authority criterion. I do so by exploring the connections between two central topics in moral and political philosophy: the justification of interpersonal harming and the moral legitimacy of practical authority. Each

of these has received a huge amount of discussion in isolation, with normative and applied ethicists concentrating on the former and political and legal philosophers focusing on the latter, but comparatively little work has been done exploring the implications of certain views about authority for theories of permissible harming, and *vice versa*.<sup>144</sup>

As I aim to show, reflection on the relationship between these two topics reveals that certain common and intuitive views about the ethics of harming require revision. In particular, I argue that, when certain conditions are met, the command of an authority can provide an independent source of justification for causing harm. In doing so I provide the resources for denying the reductivist position that acts of non-consensual harming are necessarily impermissible in the absence of a liability or lesser-evil justification.

The arguments of this chapter are intended to provide a fairly general account of how considerations of practical authority are relevant to understanding the permissibility of causing harm. I postpone the further question of how these arguments apply to the particular case of war until Chapter 5, though I employ some war-based examples for illustrative purposes.

In sections 4.2–4.5 I clarify the specific questions to be addressed and set out two main claims I aim to defend. Sections 4.6–4.11 argue for the first claim, while Sections 4.12–4.16 argue for the second. Section 4.17 concludes.

## **4.2 Authority: The Basics**

A pervasive feature of the social world is that certain persons or institutions claim to possess practical authority over others. To do so is to

---

<sup>144</sup>An exception being a recent pair of articles by Malcolm Thorburn and John Gardner, although these focus primarily on the relevance of authority to criminal law defences for acts of harming, rather than to moral permissibility and justification. Malcolm Thorburn 'Justifications, Powers, and Authority', *Yale Law Journal* 117 (2008), 1070-1130; John Gardner 'Justification Under Authority', *Canadian Journal of Law and Jurisprudence* 23, No.1 (2010), 73-98. David Estlund also provides an important discussion, which this chapter builds upon. David Estlund 'On Following Orders in an Unjust War', *Journal of Political Philosophy* 15, No.2 (2007), 213-234.

claim the moral power to issue commands and, by doing so, place those to whom the command is directed under an obligation to obey. Paradigmatic examples include a parent directing their child to 'Clean up your room!', a policewoman ordering a car driver to 'Stop right there!', and an army officer commanding his troops to 'Hold your positions!'. At a more general level, states and legal systems claim the authority to place their subjects under obligations by enacting laws and through the pronouncements of officials. In all these cases the commander intends their directive to bring an entirely new reason for action into existence for those to whom it applies.

Despite the prevalence of authority claims, there is a clear philosophical puzzle here. How can someone conjure something as morally serious as an obligation merely by communicating her intention that it be the case?

Two features of authoritative commands further contribute to the puzzle. Firstly, commands are intended to provide subjects with reasons for action whose validity is not conditional on the pre-existing reasons for and against the performance of the act commanded. The subject of a command to  $\varphi$  gains a reason to do so in virtue of being commanded and not because of the merits of  $\varphi$ -ing. Valid commands thus create 'content-independent' reasons for action.

Secondly, commands are intended to play a particular role in the practical reasoning of the subject. When an authority issues a command to  $\varphi$ , the subject is not to treat that command as simply an additional reason to  $\varphi$  to be weighed alongside the range of pre-existing reasons for and against  $\varphi$ -ing. Instead, the command is meant to both provide the subject with a new reason to  $\varphi$  and also prevent the pre-existing  $\varphi$ -related reasons (or at least some of them) from bearing on the question of how she ought to act. This is known as the 'peremptory' feature of authoritative commands.

Once these two features are made overt, we can see that commands are problematic both in terms of their genesis and their status, purporting not only to create new reasons, but reasons that have an extremely privileged position in the practical reasoning of subjects. In a wide range of contexts, commands are meant to give subjects decisive reason to perform actions

they would have insufficient reason to perform – or even decisive reason not to perform – in the absence of that command.

The nature of authority claims raises two tasks for political and legal philosophy. The first lies in establishing the conditions – if any – under which claims to practical authority can be morally justified, so that a subject may have all-things-considered reason to obey the commands of an authority, independently of both their own assessment of right reason and of the actual balance of command-independent reasons. Answering this may be thought the ‘first question’ of political philosophy, and a major theme within the discipline is to explain how the possession of such power can be reconciled with important values that may be thought to conflict with it, such as individual liberty, moral equality, or rational autonomy.

Secondly, if the possession of authority can be justified this raises the further task of delineating its proper limits. Even if it is true that authorities are able to place subjects under content-independent duties, it is presumably not true that any and every command creates these duties, or that duties to obey are entirely unconditional and absolute. An account of authority’s justification, then, not only has to explain when obeying an authority is ever justified, but also when it is not.

These two tasks are fundamentally connected, in that an account of the limits of authority will fall out of an account of authority’s justification. The subject’s duty to obey an authority will be limited by the same reasons that justify the authority’s possession of normative power over the subject. Justification and limitation are thus two sides of the same coin.

### **4.3 Authority and the Ethics of Harming**

The core concern of those working on the ethics of harming is fairly self-explanatory. While causing physical harm to others is normally considered paradigmatically morally impermissible, we do not think that this is always the case. Under certain exceptional circumstances deliberately causing harm may be morally justified, even required. Theories of permissible harm



aim to provide a systematic account of all the normative factors that determine when such exceptions arise.

The topics of authority and the ethics of harming are routinely discussed in isolation from one another, by different branches of normative philosophy. This is understandable but unfortunate. Ultimately, theorists working in both these areas are engaged in the wider project of establishing what reasons moral agents have for action – with what individuals all-things-considered ought and ought not to do.

Those working on the ethics of harming are concerned with the question, ‘When, if ever, do individuals have undefeated moral reason to cause harm to others?’, which they aim to answer by providing a complete account of all the reasons for action capable of defeating whatever reasons normally render harming impermissible. Those working on the moral justification of authority, by contrast, are concerned with the different question, ‘When, if ever, do individuals have undefeated moral reason to obey the commands of another?’ The former are concerned with the justification of harm, the latter with the justification of obedience.

Set up this way, we can see that there are a range of possible cases in which determining whether an agent is justified in causing a particular harm – with whether she has undefeated reason to do so – requires an answer to an important third question, which combines our response to each of the above, namely, ‘When, if ever, do individuals have undefeated moral reason to obey commands which require them to cause, or refrain from causing, harm to others?’

These are cases in which the content of a command from a putative authority concerns the causing of harm. Term these ‘authorised harm’ cases. These are not merely hypothetical. For example, members of law-enforcement and, especially, military organisations are routinely commanded to cause harm, or refrain from causing harm, by those who claim practical authority over them.

Given the possibility of authorised harm cases, there is an important and open question here concerning the extent, if any, to which the presence of a command makes a difference to the moral reasons that apply to the

subject regarding the causing of harm. More strongly, we must ask whether the presence of a command can alter the moral status of an act of harming. This is raised most saliently by cases in which a subject is commanded to cause harm that, in the absence of the command, they would have decisive reason not to cause. In such cases there is, at very least, a *prima facie* tension between considerations of authority and standard theories of justified harm. Two putative sources of decisive practical reasons require opposing and mutually incompatible courses of action.

The best theory of permissible harming then requires a position on whether the commands of authorities can provide an independent source of reasons for and against harming and, if so, the extent to which these reasons may determine the moral status of such acts. The correct answer may well be that commands play no such role and make no difference to whether an agent is justified in causing harm. But this conclusion requires positive argument and engagement with theories of authority.

#### **4.4 The Ethics of Harm: *Completeness and Immunity***

I aim to demonstrate the relevance of authority to the ethics of harming, focussing on two topics of debate in this field. The first is the very general question of precisely which types of reasons are sufficient to justify acts of harming. Term this the question of the *range* of justifications for causing harm. The second aspect arises from a debate within the self-defence literature. It concerns the permissibility of defensively harming individuals who will otherwise cause harm to others, but who possess an all-things-considered justification for doing so. This issue rests on a more general question regarding how agents' reasons for harming are related and responsive to other agents' reasons for harming. Term this the question of *interaction*.

The standard account of why acts of harming are presumptively impermissible is that individuals possess basic rights against being harmed, which impose stringent duties on others to refrain from doing so. Given

this, it is typically thought that justifications for harming must take one of two basic forms.

Firstly, an act of harming can be justified by showing that its recipient has lost their normal right against its infliction, thus removing the constraint against harming them. There are a variety of mechanisms by which this can occur. An individual may voluntarily *waive* their right not to be harmed by consenting to its infliction, such as in cases of voluntary euthanasia, as well as certain kinds of sporting contests. The loss of rights against harm may also occur non-voluntarily. For example, an individual may lose this right because they *deserve* to suffer harm in virtue of some prior wrongdoing. This idea lies at the heart of retributivist justifications of punishment. It is also possible to harm an individual without acting against their rights if inflicting that harm is instrumental to averting a threat of wrongful harm for which they are appropriately responsible. This form of justification is known as *liability* and provides the standard justification for harming in self and other-defence.

Secondly, harm may be justified even when the recipient possesses a right against its infliction. In these cases, the reasons against harming provided by the right are overridden by weightier moral reasons in favour of causing harm. The most straightforward cases of this are those in which inflicting a harm against a right serves to avert an impartially considered much greater harm. For example, diverting a falling boulder from a path where it will crush ten strangers onto an alternate path where it will crush one stranger. In such cases there exists a *lesser-evil* justification for causing harm.

The preceding forms of justification may all be classed as grounded in *agent-neutral* reasons for action. As standardly understood, a reason is agent-neutral if it can be fully stated without making essential reference to a particular agent for whom it is a reason. Agent-neutral reasons thus apply to all agents universally. For example, the fact that Betty is experiencing undeserved suffering gives all agents a reason to see to it that Betty's suffering is relieved. Accordingly, justifications for harming grounded in agent-neutral reasons apply to all agents. For example, if inflicting minor

harm on one individual would avert a serious harm to many innocent people, then any appropriately motivated agent is able to act on this reason and justifiably inflict that harm. Similarly, if John is liable to be killed as a result of culpably posing a lethal threat to Bob, then this fact about John gives rise to a justification for killing him in defence of Bob that applies to all agents.<sup>145</sup>

Agent-neutral reasons stand in contrast to *agent-relative* reasons, whose articulation necessarily requires reference to a particular agent for whom they are reasons. Paradigmatic examples of agent-relative reasons are those grounded in special relationships and partiality. For example, the fact that Alex is Barry's child gives rise to a reason only *for Barry* to promote Alex's welfare, and not for all agents in general.<sup>146</sup>

Controversially, some argue that agent-relative reasons can provide an independent source of justification for causing harm. On one view of this type, morality affords agents a 'personal prerogative' to assign greater weight to their own interests, or the interests of those to whom they are specially related, than these interests would command from the impartial perspective.<sup>147</sup> In cases where these interests are seriously threatened, and acting against individuals' rights not to be harmed is necessary to protect them, this prerogative may provide the agent with an overriding reason to do so.<sup>148</sup> A second type of view, familiar from Chapter 3, locates an agent-relative form of justification in weighty associative duties, rather than

---

<sup>145</sup> This view of liability-based justifications as agent-neutral and universally applicable is probably the dominant position in the literature. See for example, Thomson 'Self-Defense', For discussion and criticism, see Cecile Fabre 'Permissible Rescue Killings', *Proceedings of the Aristotelian Society* 109 (2009), 149-164.

<sup>146</sup> For a lucid overview of the agent-neutral/agent-relative distinction, see Michael Ridge, 'Reasons for Action: Agent-Neutral Vs. Agent-Relative', in Edward N. Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition). Available at <http://plato.stanford.edu/entries/reasons-agent/>

<sup>147</sup> The notion of a personal prerogative comes from Samuel Scheffler, *The Rejection of Consequentialism* (Oxford: Clarendon, 1982).

<sup>148</sup> Jonathan Quong has done most to articulate and defend a view of this type. Quong 'Killing in Self-Defense', 507-537. See also, Fabre 'Permissible Rescue Killings'; Nancy Ann Davis 'Abortion and Self-Defense', *Philosophy and Public Affairs* 13, No.3 (1984), 175-207; McMahan 'The Basis of Moral Liability to Defensive Killing'; Jeff McMahan 'Debate: Justification and Liability in War', *Journal of Political Philosophy* 16, No.2 (2008), 227-244. Hobbes may be characterized as endorsing an extreme version of this view. For discussion, see Jeremy Waldron 'Self-Defense: Agent-Neutral and Agent-Relative Accounts', *California Law Review* 88 (2000), 711-750.

personal prerogatives.<sup>149</sup> On this view, participants in morally significant relationships incur extremely stringent duties to protect their associates from harm. Controversially, in cases where the associative duty comes into conflict with the general duty not to act against rights not to be harmed, the associative duty may be the weightier, thus rendering the harming all-things-considered justified as the lesser breach of duty.

A basic assumption running throughout the literature on harming is that the above forms of justification – consent, desert, liability, lesser-evil, personal prerogative or associative duty – or some combination of these, exhaust the range of justifications for causing harm. I will term this view *Completeness*, which can be stated in more general terms as:

***Completeness***: For any act of harming, if the recipient(s) of harm retain their rights not to be harmed and their rights are neither waived nor overridden by a weightier moral consideration, then there is decisive reason not to cause that harm and agents are morally unjustified in doing so.

The majority of the debate on permissible harming takes place within the confines of *Completeness*.<sup>150</sup>

I now turn to the question of *Interaction*, which concerns the permissibility of defensively harming individuals who will otherwise transgress other's rights not to be harmed, but who possess an all-things-considered justification for doing so.<sup>151</sup> As mentioned above, this issue rests on a more general question regarding how agents' reasons for harming are related and responsive to other agents' reasons for harming.

---

<sup>149</sup> Lazar 'Associative Duties and the Ethics of Killing in War'.

<sup>150</sup> For example, one of the most thorough recent discussions of the justification of harm, which 'points tantalizingly' towards the possibility of a 'grand unified theory' of permissible harming, consists almost entirely in an analysis of liability and lesser-evil justifications. David Rodin, 'Justifying Harm', *Ethics* 122, No.1 (2011), 74-110.

<sup>151</sup> I follow Frances Kamm in using the notion of a rights transgression to pick out actions that wrong an individual while remaining neutral on the question of whether the wronging is all-things-considered permissible or impermissible. Frances Kamm, 'Rights' in Jules Coleman and Scott Shapiro (eds), *The Oxford Handbook of Jurisprudence and Philosophy of Law* (Oxford: Oxford University Press, 2002), 476-513 at n.8.

An intuitive view is that there is a strong relationship here, so that if an agent is all-things-considered justified in causing harm then this, at the very least, provides a reason against the use of defensive force against that agent in order to prevent that harm from occurring. While this much may be uncontroversial, the debate is over the extent to which the justificatory burden is raised. Recently, some have argued for the stronger view that an agent's justification for harming provides decisive reason against defensively harming them. On this view, there can be no justified defence against the infliction of justified harm.<sup>152</sup> I will term this view *Immunity*:

***Immunity***: If an agent is justified in causing harm to others, then this justification defeats all the available reasons for defensively harming that agent. It is therefore morally unjustified to do so.

#### 4.5 Two Claims

I aim to show that taking a certain view of how authority can be justified and extending it to authorised-harm cases generates counter-examples to both *Completeness* and *Immunity*.

To home in on the issue, it is useful to consider two specific questions we can ask about the relationship between authority and harming, corresponding to the general questions of *range* and *interaction* outlined above.

The first concerns the extent – if any – to which the command of an authority may affect the moral status of an act of harming. More specifically, we ask whether the command of an authority can render an act of harming morally justified, when the act would be straightforwardly unjustified in the absence of the command. Can the command of an

---

<sup>152</sup> Tadros, *The Ends of Harm*, Ch.9. In his most recent work on this issue, McMahan also endorses a version of *Immunity*, albeit more tentatively. Jeff McMahan 'Self-Defense Against Justified Threateners', in Helen Frowe and Gerald Lang (eds), *How We Fight* (Oxford: Oxford University Press, 2014), 104-137. A view analogous to *Immunity* can be attributed to Locke, who writes that "Force is to be opposed to nothing, but to unjust and unlawful Force". John Locke, 'Two Treatises of Government', Peter Laslett (ed), (Cambridge: Cambridge University Press, 2002), p.402. For discussion, see Waldron, 'Self-Defense: Agent-Neutral and Agent-Relative Accounts'.

authority can give an agent an undefeated reason to cause harm, when they would otherwise have decisive reason not to do so?

We can also ask a second question, which is conditional on our response to the first. If it is true that the command of an authority can provide an agent with a justification for causing harm, we can then ask whether the existence of this justification makes a difference to whether other agents are permitted to defensively harm the authorised agent, when doing would be straightforwardly permissible in the absence of the authorised agent's justification.

I will defend a revisionary pair of responses to these two questions:

**First claim:** Under certain conditions, the command of an authority can provide an agent with a moral justification for an act of harming, when it is both true that the act transgresses a right and that the right is not overridden by a weightier moral consideration.

This claim denies *Completeness*, since it holds that an agent may have undefeated reason to cause harm even when none of the standard justifications outlined above obtain. It posits an additional and independent source of justification for harming. If defensible, this reveals that the *range* of justifications for harming is broader than commonly supposed by those working on the ethics of harm.

This strong response to the first question is tempered by a more modest response to the second:

**Second Claim:** The fact that an agent is commanded by an authority to cause content-dependently unjustified harm to others does not, in itself, raise the justificatory burden on defensively harming that agent.

This may seem like a more conservative and conciliatory claim about the relationship between authoritative commands and the morality of harming than the first, since it limits the normative effect of commands to inflict harm. However, this is not so. The second claim is also a revisionary

position within the ethics of harming when taken in conjunction with the first, since they together deny *Immunity*. If both claims are defensible, then there can be cases in which an agent is justified in harming another (in virtue of being commanded to do so), but this justification does not render the infliction of defensive harm on the authorised agent more difficult to justify than it would have been in the absence of their justification.

Before beginning a defence of these two claims, three clarifications are in order. Firstly, the claim that the commands of authorities can provide agents with undefeated reasons for causing harm refers to justification in the ‘fact-relative’ sense. On this view, commands bring genuine reasons for action into existence in the same way that, for example, valid promises do. Commands do not merely provide subjects with reasons for forming beliefs about the existence of authority-independent reasons, justifying action only in the ‘evidence-relative’ sense.<sup>153</sup>

Secondly, to avoid confusion, it is worth distinguishing my first claim from a different sense in which the commands of authorities may give rise to undefeated reasons to cause otherwise-unjustified harm. To demonstrate, consider a case in which a subject’s failure to obey an authority’s command to cause unjust harm will result in a bad consequence occurring. For example, that if disobeyed the authority will get angry and violently unleash her wrath on innocent people.<sup>154</sup> If this bad consequence is sufficiently grave then the subject may well be justified in complying with the command in order to prevent it. In such cases, while it is true that the command *results* in reasons that justify causing harm, the command *itself* does not create the reasons.<sup>155</sup> Rather, the existence of the command simply affects the causal architecture of the situation so as to activate a non-authority-based justification for harming, in this case a lesser-evil

---

<sup>153</sup> For the distinction between fact-relative and evidence-relative permissibility, see Derek Parfit, *On What Matters*, (Oxford: Oxford University Press, 2011), Ch.7. For detailed discussion and application to harming, see Tadros, *The Ends of Harm*, Ch.10.

<sup>154</sup> This is a version of the ‘Dictator’s Child’ example, discussed, independently, by David Estlund and David Enoch. David Estlund, *Democratic Authority* (Princeton: Princeton University Press, 2008), p.118; David Enoch, ‘Authority and Reason-Giving’, *Philosophy and Phenomenological Research* (forthcoming). The original example is attributed to John Deigh.

<sup>155</sup> In Enoch’s terminology, the authority only succeeds in ‘triggering’ reasons for the subject to act, rather than creating new reasons.



justification. The possibility of such cases does not challenge *Completeness* and therefore does not support my first claim.

Thirdly, the arguments of this chapter aims to show that examining the relationship between authority and harm yields interesting and important results for normative theory. It does so by elucidating the conditions under which the commands of an authority may make a difference to the reasons that apply to agents regarding harm. It does not tackle the empirical question of whether specific real-world authorities in fact satisfy these conditions. I address certain aspects of this question in Chapter 5.

## **4.6 Opposing the First Claim**

I anticipate many will find my first claim highly unintuitive, even repugnant. To begin a defence of this claim, which denies this common-sense reaction, I will outline three broad views about the justification and limitation of authority that are compatible with a commitment to *Completeness*. With this in place, we are in a better position to see what an account of authority must provide if my first claim is to be defended. Conversely, doing so also reveals which views of authority's justification and limitation must be defended in order to vindicate *Completeness*.

### *4.6.1 A Priori Philosophical Anarchism*

This first position denies my first claim at the most general level, by denying that the commands of an authority could *ever* create reasons for action. If true, this view guarantees that my first claim never gets off the ground and so ensures that *Completeness* (and, derivatively, *Immunity*) can be made safe from counter-examples that appeal to authority.

Philosophical anarchists object that commands' status as content-independent and peremptory reasons for action is incompatible with some fundamental value. On one influential version of this objection, obedience to an authority can never be justified because doing so is irreconcilable with the basic requirements of rational and autonomous agency.

The challenge can be stated in the form of a paradox, starting from the plausible assumption that agents should always aim to act in accordance with the balance of reasons that apply to them.<sup>156</sup> Given this, in cases where an authority commands us to act against the balance of reasons, obeying the command involves acting against reason. Reason then requires us not to obey. If, on the other hand, we are commanded to act as the balance of reasons recommends, then we ought to do so, but not for the reason that we have been commanded. In these cases, the command is redundant. Either way, the objection holds that treating commands as reasons for action involves directing one's rational attention away from the reasons that apply to oneself. The range of reasons for action is exhausted by the content-dependent. On this view, it is no wonder that we find claims to authority puzzling, since they are incoherent.

While philosophical anarchism may be considered an extreme view in the circles in which it is usually discussed, it nevertheless provides extremely strong support for *Completeness*. The case for anarchism might be strengthened once we see that it provides a defence of a very intuitive position within discussions of permissible harming. That said, this defence might be thought to kill the baby in order to save the bath water. If true, authorities *never* possess the moral power to impose *any* obligations on subjects.

Fortunately, the choice between endorsing philosophical anarchism or accepting my first claim is a false one. Those who wish to preserve *Completeness* may adopt one of two broad positions, which, if defensible, serve to defeat my first claim without endorsing the global scepticism of the anarchist. Both these views grant that there is some successful response to the anarchist challenge – so that authorities may possess the moral power to create content-independent obligations – but deny that this possibility entails that subjects can have decisive reason to obey commands to cause content-dependently unjustified harm.

---

<sup>156</sup> For the most influential formulation of the paradox, see Robert Paul Wolff, *In Defense of Anarchism*, (New York: Harper, 1970). For detailed discussion, see Scott Shapiro, 'Authority' in Coleman and Shapiro (eds), *The Oxford Handbook of Jurisprudence and Philosophy of Law*, 382-439.

Whereas the anarchist response protects *Completeness* by denying that the possession of authority can ever be moral legitimate, these views reject my first claim by appeal to the uncontroversial requirement that any plausible account of the justification of authority must place limits on the obligation to obey. Given this, it is natural to conclude that wherever the precise limits lie, commands to perform acts of content-dependently unjustified harming must surely exceed them, given the moral gravity of such actions.<sup>157</sup> Each of the following two defences of *Completeness* appeals to a distinct way in which authority may be morally limited.

#### 4.6.2 *The Invalidation View*

The first-view, which I shall term the ‘Invalidation View’, seeks to deny my first claim by appealing to the jurisdictional limits on authority. The idea here is that those who have authority do not possess it over any and every area of action. Authority is always held relative to a specific domain. To say that A has authority over B is always to say that A has authority over B *in domain X*. If A issues a command to B regarding domain Y, then the authority is acting *ultra vires* and its commands do not give B any content-independent reasons to act as commanded. For example, while a policewoman may be able to place me under a duty to cease walking by shouting ‘Stop right there!’, she is not able to place me under a duty to wash behind my ears by shouting words to that effect. While she may have authority over me in a certain range of domains, my personal hygiene is not one of them.

None of this is controversial. All plausible theories of authority recognise that authority claims are limited in terms of domain-specificity

---

<sup>157</sup> As Matthew Noah Smith puts it in a recent article, “The first characteristic of the obligation to obey the law is that there are very few limits on what an obligation to obey the law can require a subject to do. There are, of course, some limits. Presumably, *if obedience to the law requires commission of serious moral wrongs, then one is not obligated to obey the law*. But this limit is at the moral extremes.” Matthew Noah Smith, ‘Political Obligation and the Self’, *Philosophy and Phenomenological Research* 86, No.2 (2013), 347-375 at p.349 (emphasis added).

and aim to provide an account of when and why an authority is acting outside its legitimate sphere. The distinctiveness of the Invalidation View lies in its specific account of where these limits lie. On this view, when an authority commands an agent to perform an action that is content-dependently morally impermissible, the authority is acting *ultra vires*. Such commands are void and fail to provide subjects with any reason for action, in the same way that the policewoman's command to wash behind my ears fails to do so.<sup>158</sup>

In contrast to the full-blooded anarchist, who denies that commands ever create content-independent obligations, this view is more selective in its denial of content-independence. Whereas the anarchist asserts the primacy of the content-dependent reasons in all domains, the Invalidation View does so only in the domain of the moral reasons (or, more plausibly, the domain of moral reasons that pass some threshold of seriousness). On this view, commands may in general possess content-independent force, but this is conditional on the content of the command being pre-approved by the balance of moral reasons. Commands to act morally wrongfully are void, in the same way that promises to perform actions that are morally wrong may sometimes be invalidated by their content.<sup>159</sup>

This view then denies my first claim, since it would deem an authority commanding the infliction of content-dependently unjustified harm to have overstepped the scope of its legitimacy, thus rendering the command void. In these cases, the Invalidation View yields the same substantive verdict as the philosophical anarchist (that commands cannot provide an undefeated reason to cause content-dependently unjustified harm) and does so on the same basis (that such commands fail to generate reasons for action.)

---

<sup>158</sup> For a defence of this view, see Dudley Knowles, 'The Domain of Authority', *Philosophy* 82, No.1 (2007), 23-43.

<sup>159</sup> On the conditions of validity for promises with immoral content, see Seanna Shiffrin, 'Immoral, Conflicting and Redundant Promises' in R. Jay Wallace, Rahul Kumar, and Samuel Freeman (eds), *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon* (Oxford: Oxford University Press, 2011), 155-178.

#### 4.6.3 *The Pro Tanto View*

The Invalidation View provides one strategy for reconciling the moral legitimacy of authority with a defence of *Completeness* by appeal to authority's moral limits. Alternatively, one may also reject my first claim without endorsing either anarchism or the Invalidation View. Instead, one may accept that commands to inflict content-dependently unjustified harm fall within an authority's sphere of legitimacy and thus succeed in placing subjects under an obligation to obey, but deny that the existence of this obligation gives subjects an undefeated reason to cause harm. On this view, such commands are not void, but function only as *pro tanto* reasons for action, which may be defeated by stronger reasons against acting.<sup>160</sup> Again, we may invoke an analogy with the moral limits of promises. In cases where fulfilling a valid promise involves doing what is morally wrong, the promisor may remain under an obligation to keep their promise, but this is overridden by weightier reasons against doing so (though the obligation manifest itself in the form of an obligation to apologise to or compensate the promisee.)

The '*Pro Tanto View*', as I will call it, provides an intuitively plausible rebuttal of my first claim by combining two uncontroversial ideas. Firstly, just as any plausible theory of authority recognises that the reason-giving force of commands is limited by jurisdictional considerations, so too are they limited in terms of their weight. While the duty to obey may be stringent, it is not of absolute weight and may potentially be overridden by competing reasons.

Secondly, it is close to an axiom in moral philosophy that individuals' rights not to be seriously harmed provide the most stringent moral reasons available. Given this, it is natural to think that if the duty to obey an authority may be overridden in certain cases, it will be in precisely those cases in which obeying a command involves acting against these rights.

---

<sup>160</sup> McMahan, *Killing in War*, p.88. See also, Anna Stilz, 'Authority, Self-Determination and Community in Cosmopolitan War', n.22. In the surrounding text, Stilz also endorses a version of the Invalidation View.

The *Pro Tanto* View is more conciliatory towards my first claim than both philosophical anarchism and the Invalidation View, since it grants that commands may provide genuine reasons for causing content-dependently unjustified harm. But it remains in extensional agreement with both these views in denying that the presence of a command can justify acts of harming that would be unjustified in its absence. While commands may provide reasons to inflict a content-dependently unjustified harm, they crucially do not provide *undefeated* reasons, which is what my first claim requires.

Again, this point can be demonstrated by analogy with the moral limits of promising. In cases where fulfilling a valid promise involves doing what is seriously morally wrong, the promisor has a reason to fulfil their duty to the promises, but this reason is overridden by a weightier reason against doing so (though it may reappear in the form of a duty to compensate or apologise to the promisee).

#### **4.7 Service Accounts of Authority**

The previous section set out a range of views about the justification and limitation of authority that are entailed by a commitment to *Completeness*. If *Completeness* is true, then one of these views, or a view with a similar structure, must also be true. This is an interesting result in itself. A fairly uncontroversial assumption in one area of normative philosophy entails commitments to certain positions within another. More pertinently, we are now in a position to see what is required if my first claim is to be defended. A plausible account of authority is needed that shows all three of these views to be mistaken.

This section begins this positive task. This requires two components. Firstly, an account of how one person's possession of authority over another – understood as the moral power to create content-independent and peremptory obligations – can ever be justified. This is necessary to provide a response to the anarchist's general challenge. Secondly, and more specifically, it needs to be shown that commands to inflict content-

dependently unjustified harm do not exceed the moral limits of authority and may therefore provide undefeated reason for action. This component is required to provide a response to the two moderate objections to my first claim.

I will argue that a certain account of authority is able to meet both these requirements. This approach, influentially defended by Joseph Raz, holds that one person's possession of authority over another is justified when, and to the extent that, the authority's having this normative power serves the subject's ends.

Service-based accounts aim to provide a general model of how authority may be legitimate that meets the anarchist's challenge that obedience is fundamentally incompatible with acting in accordance with reason. A service-based view aims to show that reason and authority can be reconciled. I will argue that, if successful, service accounts also provide the resources to show that the two moderate objections to my first claim also fail. That is, if a service account succeeds in showing that subjects can ever have reason to obey an authority, then it also shows that subjects may have undefeated reasons to obey commands to cause content-dependently unjustified harm.

Raz's well-known 'service conception' of authority provides a starting point for our reflection. The service conception advances two main theses: The first – the 'pre-emption thesis' – provides a precise account of what the possession of practical authority consists in, by explicating the role that legitimate commands are meant to play in the practical reasoning of the subject.<sup>161</sup> The second – the 'normal justification thesis' – provides an account of how the possession of such authority can be justified.

According to the pre-emption thesis, the reasons provided by the command of an authority are intended to have a specific character. When an authority commands a subject to  $\varphi$  this does not simply provide another reason for  $\varphi$ -ing to be added to the pre-existing balance of reasons for and against  $\varphi$ -ing. Rather, the command both provides an additional 'first-order' reason for  $\varphi$ -ing and a 'second-order' *exclusionary* reason not to act

---

<sup>161</sup> Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986), p.46.

on the basis of the pre-existing first-order  $\varphi$ -related reasons (or, more accurately, not to act for a subset of these reasons). These pre-existing reasons are replaced or 'pre-empted' by the reason for action provided by the command.

The pre-emption thesis provides the first step in responding to the anarchist challenge. On this hierarchical account of practical reasoning, commands need not necessarily compete with the pre-existing first-order reasons for the rational attention of the subject, including the reasons that would otherwise count against the performance of the action commanded. In virtue of the command's existence, these reasons are excluded and no longer apply to the subject. This view provides the resources for a response to the anarchist's complaint that obedience entails acting against the reasons that apply to oneself, since one function of commands is to determine precisely which reasons apply.

Of course, the fact that commands are intended to play this precise role in the practical reasoning of the subject does not show that it is justified that they do so. This crucial second-step is provided by the normal justification thesis, which elucidates the conditions under which the possession of the moral power to create pre-emptive reasons is morally legitimate. According to the normal justification thesis:

the normal way to establish that a person has authority over another person involves showing that the alleged subject is likely better to comply with reasons which apply to him . . . if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons which apply to him directly.<sup>162</sup>

On this view, the possession of authority is justified in virtue of the rational benefit it provides to its subjects. An authority is entitled to create pre-emptive obligations only when their commands enable subjects to better achieve the aims that they have reason to achieve independently. The

---

<sup>162</sup> Raz, *The Morality of Freedom*, p.53.



normal justification thesis thus offers a broadly instrumental account of authority.<sup>163</sup>

Authorities can satisfy the test of normal justification in two main ways. Firstly, they can enable groups of individuals to better coordinate their behaviour, thus enabling them to resolve the various collective action problems they may encounter in pursuing valuable projects. By obeying an authority, each member can better comply with the reasons that apply to them – known as ‘dependent reasons’ – than they would if each tried to act on their own assessment of these reasons. Secondly, an authority may possess greater expertise than the subject regarding important matters, and thus be able to issue directives that more reliably track right reason than the subject is able to. In both cases, subjects can better conform to right reason by deferring to the authority and treating its directives as binding, than by acting on their own assessment of right reason.

We can now appreciate how a service account of authority provides a response to the anarchist objection that obedience is inherently irrational. On a service-based view, obedience can be shown to be rational on instrumental grounds. Given that agents have certain aims, they have sufficient reason to adopt the means to best achieving those aims. Obeying an authority may simply be the optimal means of achieving greater overall conformity with the reasons that apply to oneself.<sup>164</sup>

Furthermore, a service account explains why the commands of legitimate authorities have the particular role within the practical reasoning of the subject that the pre-emption thesis claims. Again, the reason for this is that it is rationally optimal for the subject to treat authoritative commands as wholly replacing their own assessment of right reason, rather than as *pro tanto* reasons to be weighed alongside this assessment.

---

<sup>163</sup> This is an overly simplistic way of putting things, since, as Daniel Viehoff has argued, a service-based account can also accommodate the view that subjects may have non-instrumental reasons to obey an authority. Daniel Viehoff, ‘Debate: Procedure and Outcome in the Justification of Authority’, *Journal of Political Philosophy* 19, No.2 (2011), 248-259. However, since the cases I discuss are not of this type, I will continue to characterise service accounts as providing an instrumental justification.

<sup>164</sup> Joseph Raz, ‘On Respect, Authority and Neutrality: A Response’, *Ethics* 120, No.2 (2010), 279-301 at p.299.

It is easiest to illustrate this point by considering a case of authority that is legitimated in terms of its superior expertise (another, though broadly parallel, story would have to be told in order to explain the pre-emptive quality of the commands issued by coordinative authorities.) To put things slightly crudely, authorities that are legitimated on the basis of expertise are those that are less likely to make mistakes as to what reason requires within an identifiable class of cases than the subject is. In such a scenario, if the subject allows the command to play a pre-emptive role in her practical reasoning, she will achieve an identical level of conformity with reason that the authority achieves. Now imagine, instead, that the subject adopts a mixed policy of reasoning, in which she forms her own assessment of the balance of dependent reasons and then adds some additional weight to the reasons that favour the course of action commanded by the authority. In a sub-class of cases, this extra weight will tip the balance in favour of acting as commanded. Within this sub-class she achieves the same rate of mistake that the authority achieves. But in others it will not, and the subject will act according to her own assessment. In this sub-class of cases, her rate of mistake will be higher than that of the authority. Across the whole class of cases, then, her rate of mistake will be greater than that of the authority. The weighing strategy can only serve to reduce her overall conformity with reason, compared to wholesale pre-emption. Instrumental reasoning itself dictates that valid commands play a pre-emptive role within the subject's practical reasoning, wholly replacing the reasons upon which they depend.<sup>165</sup>

---

<sup>165</sup> Raz, *The Morality of Freedom*, p.67-69. In addition to considerations of rational optimality, the pre-emptive character of valid commands can also be defended by an argument from double-counting. Since, on a service account, valid commands are wholly grounded in the dependent reasons that apply, it can be argued that the recipient of such commands cannot simultaneously be subject to the dependent reasons upon which the command is based, because the rational weight of these reasons has already been accounted for in producing the command. For commands to then compete with these dependent reasons would be to count the dependent reasons twice. To avoid double-counting, the command must either replace the reasons on which it depends, or not function as a reason at all. Since, by hypothesis, the subject will do worse in terms of conforming to right reason if the dilemma is resolved in the latter manner, reason itself dictates that legitimate commands have pre-emptive force. See Raz, *The Morality of Freedom*, pp.58-59

A service account also explains how mistaken commands – commands that fail to reflect the actual balance of dependent reasons in a particular case – may still provide subjects with undefeated reason to obey. This is because authorities do not need to be infallible in order to pass the test of normal justification. Subjects may be able to improve their overall conformity with reason in a particular domain by obeying a fallible authority, than by acting on their own assessment of right reason in each instance. In order to provide the service upon which their legitimacy rests, an authority need only be superior to the subject in terms of bringing about their conformity with right reason. The standard of normal justification is essentially comparative.

Crucially, subjects are only able to gain the rational benefits of obeying a (fallible) authority if the authority's commands are obeyed even when they fail to correctly reflect right reason. For, in order to avoid acting against right reason in particular cases in which the authority errs, the subject would have to assess each and every command against their own assessment of right reason and disregard those commands which conflict with this assessment. But by adopting this policy the subject would be forsaking the benefits of obedience. She overall does worse in terms of achieving conformity with right reason by doing so, since, by hypothesis, the authority in question meets the condition of normal justification in spite of its fallibility.

To demonstrate, consider a simple case of advisory authority grounded in expertise. Imagine that A is an expert regarding B in the domain of financial investments in the stock market. Within this domain, he is less likely to incorrectly assess right reason than B. In one instance, A may mistakenly direct B to invest in a certain company, which is, in fact, a poor investment. B will lose one hundred dollars by doing so. However, the only way that B can detect this mistake would be to go through the same consideration of financial factors that B went through in crafting his directive in each case. But if he does so, and acts on his judgements, he would overall make more mistakes in terms of maximising investment returns. B therefore has sufficient reason to act as A directs, including in

cases where A errs. The subject's aim of improving their conformity with reason itself requires that mistaken commands be obeyed, subject to certain limits to be discussed shortly.

An important feature of a service account is that the legitimacy of authority is inherently individualised. In order for a command to impose an obligation, the authority in question must not only meet the conditions of normal justification with regard to the particular domain of action in which it is issued, but also with regard to each person to whom it is issued. The very same authority may be legitimate in domain X with regard to Jones but not regarding Smith. This is because individuals are differently situated in many respects, affecting both the degree of conformity with reason that each can achieve independently and the kind of service that each may require in order to better conform. Given this, some individuals may be situated so that a particular authority is able to serve them regarding certain matters, whereas others are not. An important implication of this 'piecemeal' account of authority is that the authority of states and legal systems is likely to be far less extensive than that which they claim for themselves.

The test of normal justification provides a necessary, but not sufficient condition for the justification of authority. For sufficiency, two further conditions are required, which together account for authority's limits under a service-based account.

The first provides a principled restriction on the scope of authority's legitimacy, by specifying certain domains of reasons in which authority cannot be possessed. At first glance one might think this odd. If we think, plausibly, that rational agents always have the aim of improving their conformity with reason, then it looks like wherever we encounter reasons, we have the possibility of there being legitimate authority over those reasons under a service account. But this is not the case. Given the value of autonomy, there will be a range of domains in which agents' overriding rational aim is to choose for themselves, rather than achieve the optimal outcome. In these cases, achieving conformity with reason is incompatible with obeying an authority. Such domains are just not 'authority-apt'. These

include domains that intuitively fall under the umbrella of the liberal 'personal sphere', such as one's choice of leisure activity, romantic partner, religious affiliation, etc. In these cases, it may well be true that someone else is better placed than the agent with regard to which course of action will maximise their well-being or preferences, but this fact does not give them the authority to, for example, command the agent to play tennis rather than football after work, or marry Tim rather than Dorothy.

Given this, in order for A to possess authority over B in domain X, it must not only be true that B will better achieve conformity with the X-related reasons by obeying A than by acting on his own assessment of right reason, but must also be true that domain X is not one in which B has more reason to choose autonomously than achieve the optimal outcome. Raz terms this additional requirement the 'independence condition'.<sup>166</sup>

A service account not only places limits on authority in terms of domains of reasons, but also at the level of specific commands. As explained above, this view provides an account of how directives may still be valid and binding on subjects even when they fail to reflect right reason. A mistaken directive remains valid when the subject would have to forsake the benefits of obeying the authority in order to determine that a particular command is mistaken.

But, importantly, this rationale does not entail that *all* mistaken directives are binding in this way. Imagine a variation on the financial advisor example, in which A directs B to burn ten of his dollars. In this case, B can detect that this command fails to reflect right reason without having to go through a complex chain of financial reasoning. B can therefore disregard such directive without forfeiting the benefits of generally conforming to B's directives.

The same reasoning applies to service accounts of practical authority, which justify obedience only to the extent that is necessary for maximising

---

<sup>166</sup> Joseph Raz, *Between Authority and Interpretation*, (Oxford: Oxford University Press, 2009), p.137. The independence condition is also present in Raz's earlier writings, though less overtly. Raz, *The Morality of Freedom*, p.57; Joseph Raz, 'Facing Up: A Reply', *Southern California Law Review* 62 (1989), 1153-1235 at p.1180. For detailed discussion, see Adam Tucker, 'The Limits of Razian Authority', *Res Publica* 18, No.3 (2012), 225-240.

the subject's overall conformity with reason. Commands that require obedience beyond this point are invalid, providing no reasons for action, since there is no rational benefit to be gained by their having this status.

Importantly, whether a command's departure from right reason serves to invalidate it does not depend simply on how much of a mistake it is. In the example above, conforming to the first directive loses B ten times as much money as conforming to the second. Yet the first is valid and the second invalid. Instead, validity depends on the *type* of mistake. As Raz puts it, what matters is the clarity of a mistake, not its gravity.<sup>167</sup> Only clear mistakes invalidate commands, because only disobeying clearly mistaken commands is compatible with optimising conformity with reason.

The independence condition and the concept of clear mistakes thus provide two sources of principled limits on authority. A subject has all-things-considered reason to obey the command of an authority provided – (i) the authority meets the normal justification test with regard to that subject and the domain of reasons in which the command is issued, (ii) the domain is one in which the independence condition is satisfied, and (iii) the command is not clearly mistaken.

#### **4.8 The Authority View of Harm**

The preceding provided a brief overview of the justification and limitation of authority under a service account. If successful, the anarchist challenge can be met and one source of support for *Completeness* can be removed from contention.

In this section I argue that if we accept a service account of how the commands of authorities can ever create content-independent and preemptive reasons for action, it is a relatively short step to showing that the command of an authority may provide a subject with an undefeated reason to perform an otherwise-unjustified act of harming. I outline this argument below and discuss some of its intricacies in the next section, where I explain how it refutes the two moderate objections to my first claim.

---

<sup>167</sup> Raz, *The Morality of Freedom*, p.62.

A service-based view provides a very general account of the justification of authority – An authority A is legitimate with regard to subject B in domain X, provided that obeying A’s commands enables B to better conform to the X-related reasons that apply to them. The argument from this general account of authority to a defence of my first claim proceeds in four steps.

The first is to make one element of this three-place relation more specific: the domain in which the authority is legitimate. Presumably, unless extreme pacifism is true, there are possible domains in which acting in accordance with reason may involve causing harm to others. Term such domains ‘harm-apt’. I mentioned two possible examples in the introduction – the domains of military service and law enforcement. Agents in these domains (indeed, agents in all domains) have the rational aim of causing harm only when doing so is supported by the balance of moral reasons that apply to the distribution of harm. Indeed, we may take *Completeness* (or some sub-set of the justifications contained within *Completeness*) as providing an account of these reasons.<sup>168</sup>

The second step is to note that agents operating in these harm-apt domains may be differently situated regarding their abilities to reliably assess and successfully act on the reasons that govern the distribution of harm. Term this ‘agent-variability’.

The third step combines the first two. The facts of harm-aptness and agent-variability open up the possibility that agents operating within harm-apt domains may better achieve conformity with reason by obeying the commands of another, rather than by acting on their own assessment of right reason on a case-by-case basis. Such domains are not only harm-apt but also ‘authority-apt’. This conjunction of possibilities suffices to show how one agent may acquire legitimate authority over another with regard to the distribution of harm. If an agent will better avoid causing unjustified harm to others by obeying the commands of an authority, it seems

---

<sup>168</sup> On the assumption that the dependent reasons which determine the justification of harm are the same in all domains.

uncontroversial to say that this is what they all-things-considered ought to do.

A fourth and final step is required in order to derive the conclusion that the commands of an authority may give a subject undefeated reason to cause content-dependently unjustified harm. This is provided by the fact that, as explained above, authorities can be fallible yet legitimate on a service account of authority. On this view, subjects can be all things-considered morally required to obey commands that fail to reflect right reason. Hence, when an authority serves its subjects within a harm-apt domain, valid commands may include those that are mistaken and require subjects to distribute harm in ways that are not supported by the balance of dependent reasons. Term this four-stage argument the *Authority View of Harm*.

To demonstrate, it may be helpful to consider an example:

***Volcano Rescue:*** A volcano erupts in the middle of Nation A. In order to save as many lives as possible in the surrounding area, the lava flow needs to be diverted from areas of higher population density to lower. This requires Nation A's citizens to dig an integrated system of trenches, along which the lava can be redirected.

Let's imagine that Nation A's citizens will overall do better in terms of harming some to save others by obeying their government on matters of lava-redirection, compared to not obeying. Under the Authority View, Nation A's government acquires legitimate authority over its subjects regarding the domain of lava redirection on grounds of service. This service may be grounded in the government's expertise (it overall makes better decisions regarding lava redirection than its subjects, or makes its decisions faster) or its ability to achieve coordination among its subjects (because whether any individual's acts of trench-digging contribute to redirecting lava is dependent on what other trench-diggers are doing.)

The Authority View maintains that, under these circumstances, Nation A's citizens are morally required to obey the commands of their



government on matters of lava redirection, including in certain cases in which the command fails to reflect right reason (but not clearly so) and requires redirecting lava towards non-liable persons in the absence of a lesser-evil justification. This is because, as explained above, mistaken directives remain binding because, and to the extent that, the subject would have to forsake the benefit of obeying the authority in order to determine that a particular command is mistaken. Given that obeying their government's commands on matter of lava-redirection – including mistaken commands – is the citizens' best means of saving lives, instrumental reasoning dictates that this is what they morally ought to do, all things considered.

While a service account of authority may be complex, the inference from the plausibility of a service account to a defence of my first claim is fairly straightforward. If a service account succeeds in showing how a command can *ever* provide pre-emptive and content-independent reasons for action, it follows that commands can provide agents with undefeated reason to cause harms that they would have decisive reason not to cause in the absence of that command. In other words, certain acts of harming may be justified even when none of the justifications specified in *Completeness* obtain.

If correct, this shows that *Completeness* is too narrow an account of the range of justifications for causing harm. While *Completeness* may well be correct as an exhaustive account of the first-order, dependent reasons that can justify harming, there is a strong case to be made for an additional form of justification that is partly second-order in character. Whereas it is standardly held that justifications for harming fall into one of two categories – where the reasons against harming are either vitiated or overridden – the Authority View of Harm posits an additional form of justification, where the reasons against harming are defeated by *exclusion*.

## 4.9 Three Clarifications

The Authority View is derived by applying a service account of authority to a particular class of harm-apt domains. While this move is relatively straightforward in principle, certain features of harm-apt domains raise complications for the Authority View that require clarification.

### 4.9.1 *The Irrelevance of Certain Services*

On a service account of authority, the nature of the domain over which authority may be possessed is relevant to the kind of service that can morally justify authority within that domain.

To demonstrate, one common way in which an authority may serve its subjects on a certain matter is by reducing certain costs that subjects would incur by deliberating on the matter themselves. Deliberation often takes a lot of time and effort, which could be used to pursue other valuable projects. An authority may thus successfully serve its subjects by taking this deliberation out of the subject's hands.<sup>169</sup> This may be the case even if the authority's deliberations are less likely to track right reason than those of the subject. The deliberation costs that the subject avoids by obeying an authority may outweigh a degree of sub-optimality in terms of the quality of the authority's decisions. The subject will better conform to reason overall by obeying the authority than by not doing so, even if they would have reached better decisions by deliberating themselves.

However, this will not be true in domains where significant moral costs attach to sub-optimal decision-making. This is especially true of harm-apt domains. In these cases, the fact that a subject will avoid deliberation costs by obeying an authority does not justify doing so, because the subject has greater overall reason to achieve optimal outcomes within that domain than to avoid deliberation costs.

What this shows is that on a service-account, an authority may be morally legitimate with regard to one domain but not another, even if its

---

<sup>169</sup> Raz, *The Morality of Freedom*, p.75.

decisions are of equal quality in each domain. This is because certain means by which an authority may serve its subjects are more appropriate in certain domains than in others. So, when applying a service account of authority to harm-apt domains, it is important to note that certain services – such as the reduction of deliberation costs – will not be sufficient to justify authority.<sup>170</sup> In these domains, the justification of authority will largely depend on the service provided by the quality of an authority's directives.

#### 4.9.2 *The Option of Inaction*

The comparative nature of the justification of authority under a service account also requires clarification. As Raz formulates the normal justification thesis, a subject's obligation to obey an authority arises from the fact that the subject better complies with the relevant reasons by obeying compared to *trying to follow* the reasons that apply to them directly. This implies that the relevant comparison is between the level of compliance a subject would achieve by positively acting on their own assessment of right reason within a certain domain and the level of compliance they would achieve by obeying the authority.

However, this is an overly narrow account of the comparison class, since the subject may also have the option of adopting a policy of inaction with regard to a certain domain – neither acting on their own assessment of right reason nor obeying an authority. Importantly, it may be the case that the subject will better overall conform with reason via a policy of inaction, compared to obeying an authority, *even if* obedience is superior compared to the subject positively acting on their own assessment of right reason. Yet the narrow formulation of the comparison class implies that, under these conditions, the subject has undefeated reason to obey the authority.

That this cannot be correct comes out most starkly when a service account is applied to harm-apt domains. To demonstrate, imagine a group of individuals who are fairly poor at coordinating their actions with one

---

<sup>170</sup> Except perhaps in cases where deliberation gains come at no cost in terms of outcomes.

another in order to distribute harms. If they try to do so, they are very likely to cause more harm than good. By obeying an authority, they could improve their coordination and thereby do better at distributing harm. However, even at this improved level they are still likely to do more harm than good. Under these conditions, the narrow formulation implies they are required to obey the authority. But this is surely false. They have a morally superior option available to them – to refrain from distributing harm at all.

In order for authority to be justified in terms of service, then, obedience has to be superior compared to all the alternative feasible options available to the subject, and not just superior compared to the sub-set of those policies which involve the subject positively acting.<sup>171</sup> In terms of the Authority View of Harm specifically, this means that authorities will only be able to place subjects under undefeated obligations to cause harm within a harm-apt domain, when their having and exercising this moral power enables their subjects to achieve greater overall conformity with the harm-related reasons, compared to *both* acting on their own assessment of those reasons and adopting a policy of refraining from causing harm within that domain.

#### 4.9.3 Cumulative Effects

Finally, it is also worth clarifying what counts as a morally legitimating service according to the Authority View of Harm. On an intuitive understanding, an authority serves its subjects regarding a harm-apt domain if, in each case in which harm needs to be distributed, subjects are more likely to comply with right reason by obeying the authority, compared with not obeying.

However, this is insufficient to justify authority, since an authority may fail to serve its subject despite it being true that the subject will more likely

---

<sup>171</sup> Daniel Viehoff's revised version of the normal justification thesis captures this requirement. According to this formulation, an authority A is legitimate regarding subject S if "the subject S better conforms to reason that apply to it if it receives and obeys the directives of the authority A *than it would if it did not receive these directives or did not obey them.*" Daniel Viehoff, *The Authority of Democracy* (manuscript) Ch.5 (emphasis added).

comply with right reason in any particular instance. This is because a policy of distributing harms by obeying an authority may produce certain cumulative effects that are not present in any one case taken in isolation, but only when several iterated cases are taken together.

To demonstrate, consider a group of individuals acting together to distribute harms in the absence of an authority. Assume that by doing so they will mistakenly distribute harm in 10% of cases. In these cases, the burdens of their mistakes are randomly distributed among possible victims over time. Now, imagine that these individuals could achieve greater coordination by obeying an authority. If they do so, they will distribute harm mistakenly in only 5% of cases. However, the burdens of these mistakes will be concentrated on a particular sub-set of possible victims over time.

Given the possibility of cumulative distributive effects such as this, subjects may in fact fail to improve their overall conformity with reason by obeying an authority, even if they are less likely to make a mistake by obeying in any particular case. This is simply because the moral reasons that govern causing harm are sensitive to the distributive patterns that may be brought about by adopting certain procedures rather than others. For example, if the burdens produced by obeying an authority fall disproportionately on certain individuals on the basis of morally arbitrary characteristics – such as age, race or gender – then this gives rise to moral reasons against adopting that procedure on grounds of fairness or impartiality. Certain distributive patterns may also be morally objectionable on purely axiological grounds. For example, Larry Temkin has recently argued that it is a good-making feature of a state-of-affairs if a given amount of harm is spread as thinly as possible across a larger number of victims, rather than concentrated in a smaller number.<sup>172</sup>

So, in order to acquire the moral power to give subjects undefeated reasons to cause harm to others under the Authority View, an authority must enable its subjects to better comply with the full range of harm-

---

<sup>172</sup> Temkin labels this principle ‘Disperse Additional Burdens’. Larry Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford: Oxford University Press, 2012), Ch.3.

related reasons by obeying, including both cumulative and non-cumulative considerations. If subjects will worse conform to the culminative reasons by obeying, then the authority may fail to serve its subjects, even if the subjects will better comply with the non-culminative reasons.<sup>173</sup>

#### **4.10 The Moderate Objections Revisited**

I have argued that if a service account of authority is defensible, then it is a mistake to claim – as the *Pro Tanto* and Invalidation Views do – that a certain sub-set of commands (those to inflict content-dependently unjustified harms) necessarily fail to provide undefeated reasons for action. This section elaborates upon and refines the Authority View of Harm by explaining precisely why these two views fail.

##### *4.10.1 Against the Pro Tanto View*

According to the *Pro Tanto* View, commands to inflict content-dependently unjustified harms may give subjects reason to do so, but never an undefeated reason. While valid, such commands are necessarily overridden by the extremely strong reasons agents have not to inflict such harm. The content-independent reasons created by the command compete with the content-dependent reasons not to inflict the harm, and are defeated by weight.

However, under the Authority View, treating all such cases in terms of a competition of reasons is mistaken. As explained above, in order for authorities to successfully serve their subjects within a certain domain, their commands must have the status of pre-emptive reasons within the practical reasoning of the subject, wholly supplanting the dependent reasons on which they are based. Those subject to an authority that is justified on the basis of its service can only reduce their conformity with

---

<sup>173</sup> The converse may also be true, in that an authority may acquire legitimacy in virtue of its enabling a subject to better comply with the cumulative reasons, even if the subject does worse at the bar of the non-cumulative reasons by obeying.

reason by weighing its command against their own assessment of right reason.

This is equally true in harm-apt domains as in any other. In the kinds of cases in which the Authority View applies, when an authority is justified on service grounds regarding a harm-apt domain, and issues a mistaken (but not clearly mistaken) command to cause harm, the instrumental basis of that justification dictates that the dependent reasons which count against causing harm are defeated by exclusion, and so cannot be brought back so as to outweigh command.<sup>174</sup>

Importantly, this is not to say that valid and pre-emptive commands can't be overridden by weightier first-order reasons. This is perfectly admissible under a service account, provided that the reasons in question do not fall within the class of dependent reasons that the authority took into consideration in producing its command.<sup>175</sup>

To demonstrate, imagine that Smith has the aim of acting correctly on some morally important matter. The correct course of action in this domain is dependent on a complex trade-off between three distinct variables, X, Y and Z. Furthermore, imagine that Jones passes the test of normal justification regarding Smith in the domains of variable X and Y, but not Z. Jones then issues a command to Smith to act in a certain manner. This command then pre-empts Smith's assessment of variables X and Y. The reasons created by the command cannot, on pain of sub-optimality, weigh against the dependent reasons grounded in these two variables. But the reasons created by the command may perfectly permissibly be weighed against the reasons relating to variable Z. These dependent reasons are not summed up in Jones' command and are hence not excluded from featuring in Smith's practical reasoning. Furthermore, it is entirely possible that the non-excluded Z-related reasons are sufficiently weighty to defeat the first-

---

<sup>174</sup> The same conclusion can also be reached from the argument from double-counting. See n.165 above.

<sup>175</sup> Raz, *Between Authority and Interpretation*, pp.144-146.

order reason for action provided by the command, giving Smith all-things-considered reason to disobey.<sup>176</sup>

Given this, the Authority-View of harming is compatible with there being cases in which the *Pro Tanto* view provides an accurate picture of the normative situation. These will be cases in which three conditions hold. First, the authority issuing a valid command to cause content-dependently unjustified harm is legitimate regarding only a sub-set of the dependent reasons that govern causing harm. Second, the sub-set of reasons falling outside its sphere of authority count against causing harm in this case. Third, these reasons are sufficiently weighty to defeat the first-order weight of the command. In such cases, the subject will have both a content-independent reason to inflict the harm and a weightier countervailing reason not to do so.

However, the point here is that while there may be possible cases – or even many cases – of this type, it is not true that *all* cases in which a subject is commanded to cause content-dependently unjustified harm will have this structure. This is what the *Pro Tanto* View requires to be true if it is to refute my first claim and preserve *Completeness*.

All the Authority View requires in order to defend my first claim is the possibility of cases in which an authority possesses legitimacy over a sufficiently important sub-set of the reasons that govern the distribution of harm. When this condition is met, subjects may come to possess an undefeated reason to cause content-dependently unjustified harm.

#### *4.10.2 Against the Invalidation View*

The Invalidation View endorses a localised denial of content independence, which treats the reason-giving force of a command as conditional on its content not conflicting (or conflicting too seriously) with the balance of moral reasons. The scope of the moral power to require action is limited to those actions that it would not be (seriously) morally wrong to perform in

---

<sup>176</sup> This raises the difficult question of determining the positive first-order weight of preemptive reasons, which I cannot attempt to answer here.



the absence of a command. Commands to cause content-dependently unjustified harm are therefore void and my first claim false.

A service account of authority – on which the Authority View is based – provides a model of how commands may achieve the status of content-independent and pre-emptive reasons that both responds to the general anarchist challenge and generalises across domains of reasons. Given this, it is difficult to provide a principled rationale for the localised denial of content-independency required by the Invalidation View. If the subject's aim of improving their conformity with reason can explain why commands that require acting against the balance of content-dependent reasons can ever provide undefeated reasons for action, why should it not also do so regarding moral reasons, including those governing the distribution of harm? It would be arbitrary and *ad hoc* to simply carve off a subset of reasons as immune from a service-based form of justification.

This may, however, be too quick. Recall that under a service account of authority there are certain domains of action in which commands *are* necessarily void. These are the domains covered by the independence condition, in which agents have more reason to exercise autonomous choice than to achieve optimal outcomes. The invalidity of commands issued within these domains is not *ad hoc*. Given this, one way in which a version of the Invalidation View could be true, and therefore my first claim false, is if it were the case that the reasons governing the distribution of harm fell within a domain covered by the independence condition. If agents have more reason to distribute harms autonomously, than to maximise conformity with the balance of reasons, then commands to inflict unjustified harm *would* necessarily be invalid. They would be invalid for precisely the same reason that commands directing subjects to marry certain partners are invalid.<sup>177</sup>

---

<sup>177</sup> This revised version of the Invalidation View is broader than its original formulation. On the original view, it is the *moral content* of the act commanded which accounts for its invalidity. On the revised view, it is the *domain* in which the command is issued which accounts for its invalidity. The revised view is thus broader than the original, since it invalidates both commands to cause content-dependently unjustified harm *and* commands to cause content-dependently justified harm, whereas the original version invalidates only the former.

However, this is hard to believe. The distribution of harm seems a paradigmatic domain in which improving one's conformity with reason trumps the value of exercising one's autonomy. This is a domain whose authority-aptness is least controversial.<sup>178</sup> Appealing to the independence condition is thus ill-suited to rescuing the Invalidation View from the charge of arbitrariness.<sup>179</sup>

The Invalidation View may also be revised in a second direction in defence of the claim that commands to cause content-dependently unjustified harm are non-arbitrarily invalidated. This revision appeals to a further limit a service account places on the validity of commands. As explained above, even when an authority is generally legitimate with regard to a certain domain, this does not mean that all its commands issued in that domain will automatically be binding on the subject. While mistaken commands may provide agents with undefeated reasons for action, commands that are *clearly* mistaken are void and provide no such reasons.

Given this, while the Invalidation View may be mistaken to simply claim that commands are necessarily voided by their morally unmeritorious content, it may nevertheless be argued that commands with seriously immoral content (including commands to cause content-dependently unjustified harm) will *also* constitute clear mistakes. And whether a command is clearly mistaken *is* determined by its content. On this revised view, it is the clarity of the command's departure from right reason, and not its immorality *per se*, which accounts for its invalidity.<sup>180</sup>

---

<sup>178</sup> For example, Jonathan Quong, who is otherwise critical of service accounts, agrees that they provide a convincing account of the legitimacy of authority in the domain of the important moral reasons, such as those that arise from our basic rights and duties *vis-à-vis* one another. Jonathan Quong, *Liberalism Without Perfection* (Oxford: Oxford University Press, 2011), Ch.4. This domain will presumably include the moral reasons governing the distribution of harm.

<sup>179</sup> This objection does not require denying the idea that agents may have *some* reason to achieve autonomous choice over optimising outcomes in the domain of harm-distribution. It only requires denying the stronger position that agents have *most* reason to achieve autonomous choice over optimizing outcomes when it comes to distributing harms.

<sup>180</sup> Again, this revision results in a broader account of which commands are invalid compared to the original *Invalidation View*, since the set of commands that are clearly mistaken in virtue of their immoral content is a sub-set of the set of all clearly mistaken commands.

However, while this revision may avoid the charge of arbitrariness, it does so at the expense of providing an objection to my first claim. This is because the revised view is only incompatible with my first claim if it is necessarily true that any command with immoral content (or, more modestly, that any command to inflict a content-dependently unjustified harm) will also constitute a clear mistake. This is a very strong claim and one that does not seem true. Recall that in order for a particular command to constitute a clear mistake, the subject must be able to form the judgment that the command fails to reflect right reason without having to go through the same process of assessing the dependent reasons that the authority went through in producing the command.

Whether or not such a judgement can be formed is determined not only by the command's content, but also by the context in which the command is issued. Consider the example invoked above, in which A possesses authority over B in the domain of financial investments and mistakenly commands him to burn ten dollars. Although one might think that such a ludicrous command is intrinsically clearly mistaken, B is only able to form the judgement that this is the case by making a comparison between the act-type commanded [Burn ten dollars!] and the particular domain in which the command is being issued [Financial Investment].

To demonstrate, consider a fanciful variation on this case:

***Vending Machine Island:*** Bert and Ernie are washed up on a desert island with fifty one dollar bills each. The weather on the island is changeable and the only source of food is a vending machine. In order to survive the two weeks until help arrives, Bert and Ernie must each day decide how much money they will burn as tinder in order to start fires to keep themselves warm, and how much to use to feed themselves via the vending machine. These decisions are quite tricky, requiring expertise on meteorology, nutrition and mathematics. Bert is far better placed than Ernie with regard to assessing these reasons.

Given these facts, let us grant that Bert gains legitimate authority over Ernie in the domain of the use of the dollar bill reserve, given his superior expertise. Now imagine that Bert issues the command [Burn ten dollars!]. However, the command is mistaken. They would be better off spending it on Twinkies in this particular instance. However, in the context [*Vending Machine Island*], Ernie cannot form the judgement that the command is mistaken without going through same assessment of dependent reasons that Bert went through in producing the command. The command therefore remains binding on Ernie, giving him undefeated reason to burn ten dollars.

This example shows us that the very same command may constitute a clear mistake in one context, but not in another. Once we recognise this, we can see why the revised Invalidation View is implausible. In order to work it must presuppose that commands to inflict content-dependently unjustified harm constitute clear mistakes in all possible domains. In other words, it must be the case that there is a class of non-comparatively clear mistakes and that commands to inflict content-dependently unjustified harm are an instance of it. Perhaps there are instances of non-comparatively clear mistaken commands – such as commands to perform a logically impossible action [Do X and not-X! Now!]. But the idea that *any* command to inflict content-dependently unjustified harm can be clearly detected is surely false. If mistaken commands of the form [Burn ten dollars!] constitute clear mistakes only in certain contexts, then so too are mistaken commands of the form [Cause harm!]. Harm-apt domains are precisely those contexts in which such commands need not constitute clear mistakes.

Both the *Pro Tanto* and Invalidation Views are thus unsuccessful in refuting my first claim. Under the Authority View it may well be true that commands to inflict content-dependently unjustified harm are often invalidated or overridden, but it is not true that *all* such commands are so defeated. This is all that my first claim requires.

Importantly, I have not argued that these two views are mistaken because they misidentify ways in which the moral power of authorities is

limited. As we have seen, service accounts also impose limits on the duty to obey in terms of both the validity and weight of commands, as any plausible account of authority must do. Instead, the objections fail because they assume that the question of whether specific commands issued by a legitimate authority exceed these limits can be settled in the absence of a specific account of why that authority is legitimate. However, while intuitive, this is not true. To repeat a point made earlier in Section 4.2, an account of the moral limits of authority must be derived from an account of its justification. Whatever considerations account for why we should have authorities in the first place will determine the precise limits of the moral powers that legitimate authorities possess. Different accounts of authority will thus yield different answers to the above question.

On the view that authorities are justified because, and to the extent that, they serve their subjects' ends, the extent of the moral power to require action will be calibrated to what is required to successfully provide the relevant service. As I have argued, when the relevant service is that of enabling subjects to better conform to the reasons that govern the distribution of harm, an authority cannot provide this service unless its directives are capable of providing undefeated reasons for causing harm, even in certain cases where the command fails to reflect right reason.

#### **4.11 The 'Irrelevant Reasons' Objection**

Before moving on, it is worth considering an interesting objection that may be pressed against the Authority View of Harm, which targets the more general service-based account of authority on which it depends.<sup>181</sup> The objection focuses on the compensatory nature of authority on a service account. More specifically, on the fact that, on this view, subjects have reason to obey an authority because doing so enables them to better conform to reasons that they would not otherwise be able to, in virtue of certain *deficiencies* or *inabilities* on the part of the subject.

---

<sup>181</sup> This objection has been raised by Victor Tadros in personal communication. Another version of it is suggested in Kimberley Brownlee, 'The Responsibilities of Criminal Justice Officials', p.130.

Given this, one may object that the commands of service-based authorities cannot create practical reasons that are capable of morally justifying action. While there is a sense in which those who are subject to such an authority ought to obey its commands – since, by hypothesis, they will worse conform to reason by not obeying – the ‘ought’ in question is relativised to the subject’s abilities. However, it is argued, when it comes to moral justification we are concerned with what the subject ought to do in a different, unrestricted sense that is not relativised to their abilities.<sup>182</sup>

If this is correct, my argument in defence of authority-based justifications for harming rests on a mistake, since authoritative commands do not provide the kind of reasons that are capable of justifying any actions. The Authority View of Harm, and service accounts of authority more generally, trade on a conflation of two senses of justification. Term this the *Irrelevant Reasons* objection.

The underlying thought behind the objection is that an agent’s inabilities are irrelevant to whether they possess a moral justification for action. Term this the *Irrelevance Principle*. The principle can be illuminated and supported by considering certain cases, such as the following<sup>183</sup>:

***The Two Nazis:*** Steve is a guard in a Nazi concentration camp. If he follows his own judgement of right reason he will murder all the prisoners in the camp. Dave is the commander of the camp and is slightly less immoral than Steve. If Steve obeys Dave’s commands rather than act on his own judgment, he will murder only the adult prisoners and spare the children.

In this case it is clear that Steve is not morally justified in obeying Dave’s commands to kill the adult prisoners, even though it is true that he will act less wrongly by doing so than if he acts on his own assessment of right reason. While there is a sense in which he ought to murder only the adult

---

<sup>182</sup> Cf. Brownlee’s view that an official who causes or upholds injustice as a result of obeying an authority or procedure whose judgements are superior to her own “does not act as she ought, *even though she acts better than she would if she attended to her own judgment directly.*” Ibid. (Emphasis added)

<sup>183</sup> This example is Victor Tadros’s (personal communication).

prisoners *rather than* act on his judgment and murder all the prisoners, what he *actually* ought to do is not murder anybody. His inability to recognise and conform to moral reasons does not generate the kind of reasons for action that are relevant to moral justification.

In response to the Irrelevant Reasons objection, I argue that the Irrelevance Principle is too strong. It is not true that *all* inabilities are irrelevant to justification. Some are intuitively relevant. Furthermore, once we appreciate that the correct distinction is between inabilities that are justification-relevant and those that are justification-irrelevant, it is far from obvious that the kinds of inabilities which may generate reasons to obey an authority fall on the justification-irrelevant side of the distinction.

Most uncontroversially, an agent's physical inabilities are relevant to whether they are morally justified in acting. Consider the following case:

***Double Trolley 1:*** Dana is taking a walk along the street and comes across two trolley cars that are out of control. Trolley A is hurtling towards 1 innocent person, Trolley B is hurtling towards 5 innocent persons. Fortunately, each trolley may be diverted onto an empty side-track by pulling a lever, though there is only time to divert one trolley. Unfortunately, the lever to divert Trolley B is rusty and Dana is not strong enough to pull it.

It seems clear that Dana is morally justified in saving the 1. But her justification is certainly relativised to her abilities. If Dana had the physical ability to save the five by pulling the rusty switch she would be morally unjustified in saving the 1.<sup>184</sup> But since she lacks this ability, what she straightforwardly ought to do is save the single victim. This shows that some inabilities are perfectly admissible in determining whether an agent possesses a moral justification for action. The Irrelevance of Inabilities Principle is too strong, since it implies that Dana is unjustified in saving the single victim, which is false.

---

<sup>184</sup> Though some would deny this. See, most famously, John Taurek, 'Should the Numbers Count?', *Philosophy and Public Affairs* 6, No.4 (1977), 293-316.

At this point the objector may happily concede that not all inabilities are justification-irrelevant and offer a weakened version of the Irrelevance Principle. The revised principle holds that while agents' physical inabilities may be justification-relevant, their non-physical or cognitive inabilities are not. This seems sufficient to avoid implausible implications in cases such as *Double Trolley 1*, while maintaining the general objection to service accounts of authority.

However, once we allow that physical inabilities can be justification-relevant, a certain class of reasons to obey an authority may be excluded from the scope of the objection – those that arise from individuals' inability to successfully coordinate their actions with others. This is because the inability to achieve coordination seems relevantly analogous to a physical inability. A group's coordinative inability need not result from any cognitive inabilities on the part of any individual member. What is distinctive of coordination problems (and collective action problems more generally) is that they can arise even when participants are appropriately motivated and fully aware of all the relevant facts.

Furthermore, in the absence of a means of coordination, each participant's ability to bring about desirable outcomes is limited by external physical facts. While each participant in a situation requiring coordination may be physically able to perform the action that would be necessary to bring about a desirable outcome, whether or not that action is actually causally effective in bringing about the outcome depends on whether coordination with others is possible.

If the analogy between physical inabilities and the inability to coordinate is plausible, we may appropriately class the latter as justification-relevant. To demonstrate, consider a variation on the above case:

***Double Trolley 2:*** Dana is taking a walk along the street and comes across two trolley cars that are out of control. Trolley A is hurtling towards 1 innocent person, Trolley B is hurtling towards 5 innocent persons. Fortunately, each trolley can be diverted onto an empty side-track, though



there is only time to divert one trolley. Dana can divert Trolley A by pulling a well-oiled lever. However, in order to divert Trolley B, Dana must act together with four other potential co-rescuers. Each must pull a separate lever at a certain time and in a certain sequence. Unfortunately, due to time constraints and the distance between them, the five are unable to coordinate their lever-pulling with one another.

I contend that Dana is morally justified in saving the 1 in this case, for the same reason that she is justified in the original case. She simply lacks the ability to bring about the saving of the 5. If correct, this shows that moral justification may be relativised to agents' coordinative abilities. Furthermore, this point suggests that when an authority is legitimate on grounds of its ability to secure coordination, those subject to it can be morally justified in obeying its directives, despite the fact that their reasons to do so arise from an inability on their part.

Of course, this point only provides a partial response to the Irrelevant Reasons objection. Even if coordinative inability is relevantly analogous to physical inability and are therefore not problematic, one may still endorse the revised version of the Irrelevance Principle which holds that non-physical or cognitive inability are justification-irrelevant. In particular, one may hold that agents' suboptimal powers of reasoning are not relevant to whether they are morally justified in acting.<sup>185</sup> If this is correct, then the objection may still succeed in showing that subjects cannot be justified in obeying authorities that are legitimate on grounds of superior expertise, since their reasons to do so will be relativised to their cognitive inability or limitations.

Agent's non-physical inability undoubtedly seem justification-irrelevant in certain cases, such as *The Two Nazis*, in which the inability in question is a specifically *moral* disability, or where the agent is morally culpable for their inability to recognise, assess and conform to reasons. However, restricting the Irrelevant Reasons objection to non-physical

---

<sup>185</sup> Tadros focuses on this particular inability in presenting the Irrelevant Reasons objection.

inabilities is still too strong. It is implausible that the distinction between justification-relevant and justification-irrelevant inabilities neatly tracks the distinction between physical and non-physical inabilities.<sup>186</sup> To demonstrate, consider the following example<sup>187</sup>:

***Disease Cure:*** Dana spends her days working for a worthy and effective charity. She could, alternatively, sit down at her computer each day, type out the cure for a terrible disease, and send it to a top medical journal. The consequences of doing so would be overwhelmingly good. But Dana does not do so, not because she believes that curing terrible diseases would have suboptimal consequences, or because she has no desire to bring about overwhelmingly good consequences, but rather because she has no clue what the cure for these diseases is, or how to write them down.

In this case Dana is perfectly physically able to bring about overwhelmingly good consequences – all she has to do is push the buttons on her keyboard in whatever order spells out the cure for a deadly disease. Her inability to do so arises from a lack of expertise. If it is true that such inabilities are justification-irrelevant, then we must conclude that Dana is morally unjustified in going to work for the worthy charity each day, rather than disseminating the cures for terrible diseases. This seems implausible. Furthermore, generalising this conclusion renders practically all actions morally unjustified, since agents will be in an analogous position to Dana in the vast majority of cases, having the means of producing extremely good consequences within their physical abilities, which they do not act on.<sup>188</sup> I take it that few of us would be happy to accept this result. But avoiding it

---

<sup>186</sup> A further problem is that it may not be possible to cleanly distinguish between physical and non-physical inabilities in the manner suggested by the revised Irrelevance Principle. This is because many disabilities we may intuitively classify as non-physical – such as our limited powers of reasoning – will often be attributable to our physical limitations. For example, it seems rather arbitrary to classify my inability to run faster as a physical inability – and therefore justification-relevant – while holding that my inability to think faster is a non-physical inability – and therefore justification-irrelevant – given that my cognitive performance is presumably also limited by certain facts about my physiology.

<sup>187</sup> This example and the following paragraph draws on Dale Dorsey, 'Consequentialism, Cognitive Limitations and Moral Theory', in Mark Timmons (ed), *Oxford Studies in Normative Ethics: Volume 3* (Oxford: Oxford University Press, 2013), 179-202 at p.181.

<sup>188</sup> See *Ibid.*

requires accepting that agents' cognitive abilities are admissible to judgements of moral justification..

If we accept that non-physical inabilities may be justification-relevant, the objection that subjects necessarily lack a moral justification for obeying authorities that are legitimate on the basis of their superior expertise becomes correspondingly weaker. In these cases, the subject's reasons to obey the authority are relativised to their non-physical abilities, such as their powers of reasoning. But, as the example above suggests, this kind of relativity need not always be problematic.

I do not claim that these brief remarks decisively refute the Irrelevant Reasons objection. To do so would require a complete account of the distinction between justification-relevant and justification-irrelevant. This is an enormously complex and difficult task, which involves resolving some of the very deepest issues in normative ethics. What I have aimed to show is that the initial intuitive force of the objection rests on either a false or overly simplistic view of the relationship between abilities and moral justification. All justification-talk has to be relativised to the fact that we are limited and finite beings. The key question is which limitations and inabilities are relevant to judgements about moral justification and which are not. Furthermore, there is no obvious or natural account of how this distinction ought to be drawn. Once we appreciate the complexity of the issues at stake, we can see that the Irrelevant Reasons objection needs to do much more than simply point out that, on a service account, a subject's reasons to obey an authority are relativised to their abilities.

#### **4.12 Defending The Second Claim**

In the following section I shift focus from the question of the *range* of justifications for harming to that of how justifications for harming *interact* with each other. With respect to the morality of harm, the question of *interaction* is concerned with the extent to which an agent's possession of a justification for causing harm affects whether other agents are justified in defensively harming that agent.

Within debates on the ethics of self-defence, several theorists defend a version of the view I labelled *Immunity*:

***Immunity***: If an agent is justified in causing harm to others, then this justification defeats all the available reasons for defensively harming that agent. It is therefore morally unjustified to do so.

If my first claim is defensible, and the commands of authorities can provide an independent source of justification for causing harm, we can apply the question of *interaction* to this particular form of justification. More specifically, we are here concerned with the extent to which an agent's possession of an authority-based justification for harming affects whether other agents are permitted to defensively harm that agent.

I will defend the following account of *interaction* in the case of authority-based justification for causing harm:

***Second Claim***: The fact that an agent is commanded by an authority to cause content-dependently unjustified harm to others does not, in itself, raise the justificatory burden on defensively harming that agent.

This second claim, in conjunction with the first, denies *Immunity*. Together, they hold that there may be cases in which one agent is justified in causing harm (in virtue of a legitimate command to do so), but that the presence of this justification does not count against the infliction of defensive harm on that agent.<sup>189</sup>

---

<sup>189</sup> The *Second Claim* is compatible with the view that authorised harm-doers may not be *culpable* for threatening harm and that this may raise the justificatory burden for inflicting defensive harm upon them. The claim is specifically that the existence of the *justification* does not affect the normative situation of others, which is separate from the question of culpability.

#### 4.13 *Interaction* and Standard Justified Threat Cases

As argued above, when certain conditions obtain, agents who cause content-dependently unjustified harm in conformity with a command may be justified in doing so. Considered from the causal/temporal perspective in which their harms are imminent but yet to eventuate, these authorised harm-doers can be classed as a species of what are known as ‘justified threateners’ in the self-defence literature. Justified threateners are agents who act in such a way that, unless prevented, they will transgress others’ rights not to be harmed, but who possess all-things-considered sufficient reason to do so.<sup>190</sup>

To date, discussions of justified threateners and *interaction* have focussed on cases in which the threatener possesses a lesser-evil justification. Term these ‘standard’ justified threat cases. These provide a useful starting point for assessing the question of *interaction* in the case of threateners who possess an authority-based justification.

The classic case employed in discussions of *interaction* is that of the *Tactical Bomber*:

***Tactical Bomber***: A bomber crew of five combatants are on a mission to bomb a munitions factory as part of a just war. Bombing the factory will result in the deaths of five innocent bystanders as a side-effect. However, the good achieved by bombing the factory is sufficient to justify doing so as the lesser-evil. The five bystanders have access to an anti-aircraft gun and are able to shoot down the bomber before they drop their bombs.

The question here is whether the bystanders are permitted to defensively kill the bombers, given that the bombers are justified in causing their deaths. Many find it intuitive that the bystanders would be permitted to do so.<sup>191</sup> On this view, this is a case in which two parties simultaneously

---

<sup>190</sup> See McMahan, ‘Self-Defense Against Justified Threateners’ for a more detailed characterisation.

<sup>191</sup> McMahan, ‘The Basis of Moral Liability to Defensive Killing’; ‘Debate: Justification and Liability in War’; Steinhoff, ‘Debate: Jeff McMahan on the Moral Equality of Combatants’;

possess a justification for harming the other. However, several theorists have recently defended the view that *Immunity* holds in standard cases of justified threateners.<sup>192</sup> On this view, the fact that the bombers are justified in causing harm entails that defensively harming them is unjustified.

The debate between these two views ultimately turns on the position that each takes on the *range* of justifications for harming. As explained in Section 4.4, justifications for harming can be divided into those grounded in agent-neutral reasons and those grounded in agent-relative reasons. Agent-neutral justifications – such as liability and lesser-evil – apply to all agents, whereas agent-relative justifications – grounded in personal prerogatives and associative duties – apply only to specific agents.

If one takes the view that the *range* of justifications for harming is thoroughly agent-neutral, then a commitment to *Immunity* follows quite naturally. If the reasons that determine how harm all-things-considered ought to be distributed in any particular case apply equally to all agents, then this gives every agent the common aim of seeing to it that that this distribution comes about, or at least refraining from preventing it from coming about.<sup>193</sup> For the agent-neutralist, it is contradictory to hold that certain agents may be justified in bringing about one distribution of harm, while other agents are justified in bringing about an opposing distribution. A normative situation like this would imply that the very same pool of reasons could simultaneously recommend opposing and incompatible courses of action. On this view, there cannot be cases of ‘symmetrical justification’ or ‘conflicts of permissible harming.’<sup>194</sup>

---

David Mapel, ‘Moral Liability to Defensive Harm and Symmetrical Self-Defense’, *Journal of Political Philosophy* 18, No.2 (2010), 198-217; Rodin, ‘Justifying Harm’; Adil Ahmed Haque, ‘Rights and Liabilities at War’, in Paul Robinson, Stephen Garvey and Kimberly Ferzan (eds), *Criminal Law Conversations* (Oxford: Oxford University Press, 2009), 395-396; Adam Hosein, ‘Are Justified Aggressors a Threat to the Rights Theory of Self-Defense?’, in Frowe and Lang (eds) *How We Fight*, 87-103.

<sup>192</sup> Tadros, *The Ends of Harm*, Ch.9; McMahan, ‘Self-Defence Against Justified Threateners’.

<sup>193</sup> For the characterisation of agent-neutrality in terms of common aims, see Derek Parfit, *Reasons and Persons*, (Oxford: Oxford University Press, 1984), p.27.

<sup>194</sup> For example, Victor Tadros defends a thoroughly agent-neutral view of permissible harming and explicitly appeals to this view in order to reject the possibility of conflicts of permissible harming (excluding cases in which the conflict is itself valuable – in sporting contests for example). Tadros, *The Ends of Harm*, Ch.9.

Conversely, if one endorses a broader account of the *range* of justifications for harming which includes agent-relative forms of justification, then one need not endorse *Immunity*. On this view, the reasons that determine how harm ought to be distributed are not common to all agents. There are forms of justification that apply only to specific agents. Given this commitment, it need not be contradictory to hold that, in cases like the *Tactical Bomber*, one party is justified in bringing about one distribution of harm whereas another party is justified in bring about an opposing distribution. While the bombers may possess a lesser-evil justification for causing harm, the innocent bystanders may, for example, be justified in resisting on the basis of a personal prerogative to assign greater weight to their own interests. Hence, on an account of the *range* of justifications for harming which includes agent-relative justifications, *Immunity* need not be true and cases of symmetrically justified harming may be possible.<sup>195</sup>

#### 4.14 Authorised Threateners and *Immunity*

Let us now turn to the question of *interaction* in the case of threateners who possess an authority-based justification for causing harm.

It is tempting to endorse *Immunity* in these cases and treat authorised agents as equivalent to standard justified threateners such as the *Tactical Bombers*. On this view, if an agent possesses an authority-based justification for causing harm, then other agents lack justification for defensively harming the authorised agent. This view is appealing because it generates the intuitively right result in a range of cases, such as the following:

***Police Officer:*** A police officer acts to arrest an individual as a result of a command to do so from a legitimate authority. However, the command is mistaken (but not clearly so) and the prospective arrestee is innocent.

---

<sup>195</sup> For further discussion, see Waldron, 'Self-Defense: Agent-Neutral vs Agent-Relative Accounts'.

Intuitively, it seems impermissible for the mistaken arrestee to use defensive force against the police officer. Combining the authority-view of harming with *Immunity* provides a neat and principled explanation for why this is so. The Authority View allows us to characterise the police officer as posing a justified threat to the innocent arrestee, despite the fact that the harm is not justified on the basis of content-dependent reasons.<sup>196</sup> The addition of *Immunity* allows us to make the argument that the police officer's justification for harming the arrestee serves to defeat the reasons that the arrestee would otherwise have for using defensive force against the police officer, thus rendering resistance unjustified.<sup>197</sup> Furthermore, this analysis also yields the intuitively right result in a variation on the case:

***Vigilante:*** A private citizen acts to carry out a citizen's arrest on the basis of a reasonable and good-faith suspicion that the arrestee will otherwise commit a serious crime. However, they are mistaken and the prospective arrestee is innocent.

In this case it does seem intuitively justified for the arrestee to use defensive force against the vigilante. Again, combining the Authority View with *Immunity* provides a neat explanation of why this is so. Though both the police officer and the vigilante threaten an identical harm, only the police officer possesses a justification for doing so, since only the police officer threatens harm in conformity with a legitimate command. Though the vigilante may reasonably believe that they are justified in harming the

---

<sup>196</sup> Those who endorse *Completeness* would seem to lack the resources to account for the intuition that the police officer acts with justification. They either have to treat the police officer as merely excused (making it difficult to explain why resistance would not be justified) or else adopt an evidence-relative account of justification (on the assumption that the command gives the officer sufficient reason to believe that the arrest is justified on content-dependent grounds).

<sup>197</sup> This is not to deny that there are alternative explanations of the intuition that it is impermissible for the innocent arrestee to resist that are compatible with *Completeness*. For example, defence may be futile or counter-productive, given that other police officers will act to make the arrest even if the initial arresting officer is successfully resisted. However, such explanations offer an unsatisfactorily contingent defence of the impermissibility of resistance. Thanks to Jeff McMahan and James Lenman for raising this point.



arrestee, they in fact lack an undefeated reason to do so.<sup>198</sup> Since this analysis classifies the vigilante as a species of unjustified threatener, *Immunity* does not apply to his case and resistance may then be justified (subject to the usual requirements of necessity and proportionality).

However, consideration of other cases strongly suggests that *Immunity* cannot be the right approach to *interaction* in the case of authorised-threateners. Consider the following:

***Combatants***: A group of combatants act to annex an area of territory belonging to a neighbouring state as a result of legitimate command to do so. However, the command is mistaken (but not clearly so) and the invasion is unjustified.<sup>199</sup>

In this case it seems clearly justified for those threatened by the authorised combatants to use defensive force against them. It also seems uncontroversial that disinterested third-parties would be justified in doing so on their behalf. However, applying *Immunity* to the case of authorised threateners generates precisely the opposite result. If it is true, as I have argued, that authorised-threateners are appropriately classed as a species of justified threatener, and if it is true that the infliction of defensive harm on justified threateners is unjustified, then it follows that those threatened by the authorised combatants would be unjustified in using defensive force against them. Surrender would be morally required, which is highly counter-intuitive.<sup>200</sup>

---

<sup>198</sup> Evidence-relative accounts of justification will thus struggle to explain the intuitive differences between the *Police Officer* and *Vigilante* case. Since, on this view, both agents threaten the innocent with harm on the basis of a reasonable belief, each should be classed as a justified threat.

<sup>199</sup> Estlund, 'On Following Orders in an Unjust War' also argues for the possibility of cases of this sort.

<sup>200</sup> The right to resist all forms of military aggression has recently come under sustained criticism, so it is not necessarily counter-intuitive to claim that resistance may be unjustified in a case like the one I describe here. See, especially, Rodin, 'The Myth of National Defence'. However, what is counter-intuitive, even on the most pacifistic views, is the conclusion that resistance may be unjustified *in virtue of the fact* that the aggressors act with justification. Hence, the oddity of the conclusion generated by applying *Immunity* in the *Combatants* case can be appreciated regardless of one's position on the extent of the right to resist military aggression.

Applying *Immunity* to the case of authority-based justifications for harming then results in a dilemma. Since, by hypothesis, both the police officer and the combatants possess the same form of justification for harming, we cannot claim that *Immunity* applies to one but not the other. Either *Immunity* holds in both cases – generating an implausible result in the *Combatants* case – or fails to apply in both cases – generating an unintuitive result in the *Police Officer* case.

#### **4.15 Authorised Threateners and Agent-Relativity**

I propose an account of *interaction* for authority-based justifications which aims to avoid the above dilemma. The proposal has two parts. Firstly, it argues that *Immunity* does not apply in the case of authority-based justifications for harming, thus avoiding the first horn of the dilemma. Secondly, it avoids the second horn by providing an alternative and non-*ad hoc* account of why defence may be unjustified in cases such as the *Police Officer*. This section focuses on the first argument – that *Immunity* does not apply to authority-based justifications.

Recall the above discussion of the relationship between views about the *range* and *interaction* of justifications for causing harm. Those who take an entirely agent-neutral view of the *range* of justifications are typically committed to *Immunity* as an account of *interaction*, whereas those who accept the existence of agent-relative forms of justification are able to deny *Immunity*.

Given this, a promising strategy for denying that *Immunity* applies to authority-based justifications is by showing that it is a species of agent-relative justification. If authority-based justifications consist in reasons for action that apply only to the authorised agent, and not to others, then *Immunity* need not follow from the possession of such a justification, since agent-relative reasons need not affect the normative situation of other agents. The possession of an agent-relative justification for bringing about a certain distribution of harm does not entail that other agents also have reason to bring about that distribution. In fact, the reasons that apply to

others may justify them in harming the authorised-threatener in order to prevent the threatened harm from eventuating.

The argument for this view is tentative, but fairly straightforward – authority-based justifications fit the standard characterisation of agent-relative reasons. One can only give a full statement of the reason for action provided by a legitimate command by making explicit and ineliminable reference to a particular agent for whom it is a reason. When an authority issues a command to  $\varphi$ , this is not intended to bring a new reason for  $\varphi$ -ing into existence for all agents everywhere, but only for the addressees of the command. The agent-relativity of the reasons created by legitimate commands is particularly salient under service accounts of authority, where the legitimacy of authority must be demonstrated anew with regard to each putative subject and their particular circumstances. On views of this type, a subject can only be shown to have an obligation to obey by making ineliminable reference to specific facts *about that subject*.

If this characterisation of authority-based justifications as agent-relative is defensible, we can explain why *Immunity* need not hold in authorised threatener cases such as *Combatants* – thus avoiding the first horn of our dilemma. Though the combatants possess undefeated reason for causing harm, the reasons do not affect the normative situation of others. This is true only of agent-neutral forms of justification. Given this, the victims may resist their aggressors in precisely the same way that they would be permitted to do had the aggressors lacked their authority-based justifications. This supports my *Second Claim* – that an agent’s possession of an authority-based justification for posing a threat of harm does not in itself raise the justificatory burden for defensively harming that agent.

It may be objected here that my *Second Claim* is too strong.<sup>201</sup> The objection proceeds from the following assumption – that all agents have a *pro tanto* reason to promote all other agents’ conformity with reason, which may include enabling them to be served by legitimate authorities. Given this, it seems to follow that in cases like *Combatants* the prospective

---

<sup>201</sup> This objection has been put to me, independently, by Massimo Renzo and Yitzhak Benbaji.

victims have a *pro tanto* reason not to resist the authorised threateners, since doing so prevents the authorised threateners from conforming to the reasons that apply to them provided by their authority. Hence, my *Second Claim* is wrong to hold that authorised-threatener's possession of justification does not count against resistance on the part of their victims.

However, I do not think that this conclusion follows from the assumption. It may well be true that I have the aim of promoting all other agents' conformity with the reasons that apply to them, and that doing so may involve bringing it about that others *are subject* to legitimate authorities. But it does not follow from this that I have a reason to promote others obeying commands from those authorities that fail to reflect the balance of dependent reasons.<sup>202</sup> While *the subject* of the mistaken command has an undefeated reason to obey it, they do so only because a policy of general obedience is an optimal, though imperfect, strategy *for them* to achieve greater overall conformity with their ultimate reasons. When the strategy goes awry in particular cases, I should be guided by the subject's ultimate reasons, not their instrumental reasons. Hence, in cases like *Combatants*, the victims' aim of promoting others' conformity with reason does not give them a reason to refrain from resisting their attackers. If anything, it gives them a reason to resist.<sup>203</sup>

---

<sup>202</sup> Provided that the mistaken directive is not also binding on me. This is an important caveat, which I discuss in more detail in the next section.

<sup>203</sup> One further possibility is worth considering, which is that there may be value in an individual's conforming to instrumental reasons, over and above the ultimate values that justify the existence of these reasons. For example, children and students may have instrumental reasons to obey the directives of, respectively, their parents and teachers (in order to, say, acquire virtue or knowledge), but there may also be value in the obedience itself (perhaps the hierarchical relationship is itself valuable, or perhaps obedience is constitutive of showing respect for the parent or teacher).

If this is true, then, *contra* what I have claimed above, there may be possible cases in which agents' aim of promoting all other agent's conformity with reason does give them a *pro tanto* reason not to prevent subjects from obeying authorities, even in cases where the authority's command fails to reflect right reason.

Of course, we may be sceptical that there are such values, or that this value is present in the kind of cases which we are concerned with here, where the authority-subject relationship is less intimate and where the domain of authority concerns the distribution of harm. But I am willing to accept that *if* such value were present, then all agents may have a *pro tanto* reason not to interfere with a subject obeying a command to cause content-dependently unjustified harm (though this reason would almost always be overridden by reasons in favour of preventing unjust harm).

I believe that this concession is consistent with what I claimed in the above text – that it does not *follow* from the fact that a subject is justified in obeying a mistaken command

Before moving on, it is worth highlighting two implications of the authority-based case for agent-relativity sketched in this section. Firstly, the argument suggests that agent-relative reasons can be generated even if it is true that the dependent reasons, upon which legitimate commands are based, are entirely agent-neutral. Even in a world populated solely by agent-neutral reasons, authorities may still be able to serve their subjects by issuing commands enabling them to achieve greater conformity with those (agent-neutral) reasons. But the reasons created by these commands will be agent-relative. Legitimate commands may be understood as a species of what John Gardner and Brad Hooker have independently termed ‘derivative’ agent-relative reasons.<sup>204</sup> These are reasons for action that are specific to certain agents, but whose normative force is derived solely from their role in enabling the agent to achieve greater overall conformity with the ultimate, agent-neutral reasons.

Secondly, the Authority View of harm provides a novel argument for the existence of agent-relative justifications for harming and, hence, for the possibility of conflicts of permissible harming, which is immune from objections pressed against standard, partiality-based accounts of agent-relative justification. These two forms of agent-relative justification – partiality-based and authority-based – are entirely separable, in that one can reject the former while endorsing the latter. Hence, it is not only a commitment to *Completeness* that is undermined by reflection on authority.

---

that others have a reason not to prevent them from doing – since there will be cases (perhaps the majority of cases) in which the subject has undefeated instrumental reasons to obey, but there is no intrinsic value realised by obedience.

Furthermore, and more importantly, I believe that this concession is consistent with my core claim that *Immunity* does not apply to authority-based justifications for harm, even if it shows that my *Second Claim* – interpreted strictly – requires modification. It remains the case that the subject’s *justification* for obeying a command to cause unjust harm does not give other agents reason to refrain from defensively harming them. In a case in which some intrinsic value is realised by a subject obeying such a command, this value does not justify the subject in doing so. Her justification lies in her instrumental reasons. So, even if we accept that, in such cases, agents do have a *pro tanto* reason not to defensively harm the subject, this reason is not grounded in the same reasons that justify the subject in causing unjust harm. It remains the case that *the fact that the subject is justified* does not raise the justificatory burden on defensively harming them, even if other facts – such as the intrinsic value realised by the subject’s obedience – may give rise to reasons against defensively harming the subject.

<sup>204</sup> John Gardner, *Offences and Defences* (Oxford: Oxford University Press, 2007), p.65. Brad Hooker, *Ideal Code, Real World* (Oxford: Oxford University Press, 2000), p.110.

If the Authority View of harm is defensible, a commitment to the agent-neutrality of justifications for harm and to the impossibility of conflicts of permissible harming are also not tenable.

I will return to the issue of the agent-relativity of authoritative commands in Sections 5.10 and 5.11.

#### **4.16 Authority and Constraints**

The preceding section provided an argument for why *Immunity* need not hold in the case of authority-based justifications for harming, thus allowing us to avoid the first horn of the dilemma encountered earlier.

An additional argument is required in order to avoid the second horn – that defensively harming authorised agents is justified in cases such as *Police Officer*, in which resistance seems intuitively impermissible. This conclusion seems to follow from the denial of *Immunity* in the case of authority-based justifications.

This final section provides an additional argument for why resisting an authorised threatener may be impermissible in cases of this type. The argument needs to provide a plausible, non-*ad hoc* account of the conditions under which resisting an authorised threatener is unjustified which does not depend on the fact that the threatener possesses justification. This would simply be a tacit appeal to *Immunity*, which must be avoided on pain of returning to the first horn of the dilemma.

Thus far, I have focussed on one important normative consequence of legitimate commands – the creation of a justification for causing harm where none existed antecedently. In order to explain why resisting authorised threateners is sometimes impermissible we need to look at the wider range of normative consequences that commands are able to effect. In particular, in addition to providing agents with decisive reasons for performing content-dependently unjustified action, commands may also create decisive reasons for refraining from performing content-dependently justified actions. If the idea of authority-based justifications

for harming is defensible, the possibility of authority-based *constraints* should also be.

Once we recognise that legitimate commands may be capable of effecting this additional form of deontic alteration, we have an account of why, in cases like *Police Officer*, it may be impermissible to resist an authorised threat. The key feature of such cases is that the command that harm be caused is addressed to *both* the agent who carries out the harmful action *and* the agent who will suffer the resulting harm. Given this, the command affects the normative situation of both agents. The command [Joe be arrested!] may give the police officer a decisive reason to inflict the harm of arrest on Joe *and* give Joe a decisive reason not to exercise his normal right of self-defence.<sup>205</sup> The idea here is that authorities acquire these twin moral powers because their having them serves their subjects, by enabling them to achieve the coordinative and adjudicative benefits of a system of law.<sup>206</sup> It is this dual exercise of authority, I contend, which explains why resisting authorised-threateners is impermissible in certain cases.

To clarify, the notion of an authority-based-constraint is not a tacit reaffirmation of *Immunity*. On this view, it is *not* the fact that that the authorised-threatener possesses a justification for causing harm which explains why their victim is not justified in resisting in cases like *Police Officer*. This is what *Immunity* would hold, since *Immunity* is an account of how justifications *interact* interpersonally. Rather, according to the authority-based-constraints explanation, *both* the authorised agent's justification for causing harm and their victim's lack of justification for

---

<sup>205</sup> Of course, this reason will not always be decisive. If the harm faced by Joe were more serious – long term imprisonment for example – then Joe may be morally justified in resisting, despite the fact that he is subject to an authoritative command that he be harmed in this way. While the legal authority may be legitimate regarding Joe, in virtue of enabling to better comply with his reasons to, for example. coordinate with others and adjudicate disagreements impartially, these reasons are not absolute and may be overridden by countervailing considerations that fall outside of this domain, such as the costs Joe would have to bear by obeying the authority.

<sup>206</sup> For a discussion of how authorities may successfully serve their subjects on grounds of adjudication, see Viehoff, 'Debate: Procedure and Outcome in the Justification of Authority'.

resisting are explained by the same antecedent condition – a legitimate command issued by an authority. Each of these normative effects are independent of one another, in the sense that the existence of one does not entail the existence of the other. There is no *interaction* between an authorised-threatener’s justification for harming and the normative situation of their victim, though the two are related by a shared antecedent.

This can be demonstrated by pointing out that there may be cases in which subjects are prohibited from using defensive force by the command of an authority, but where the agent who threatens them lacks any justification for doing so. Consider the following case:

***Invasion***<sup>207</sup>: A small island nation (A) is facing wholly unjustified annexation by a more powerful neighbour (B). A’s government possesses legitimate authority over its citizens regarding the domain of national defence and, after assessing the expected costs and benefits, commands its citizens not to resist B’s agents.

In this case, under the Authority View, resistance is rendered impermissible by the authority’s command. But it is clearly not the case that the prohibition on resistance is tracking justification on the part of the aggressors, since the aggressors lack any justification. Cases such as this demonstrate that the authority-based constraints are entirely separable from the threatener’s justification and, therefore, that no tacit appeal to *Immunity* is being made

#### **4.17 Conclusion**

Reflection on the relationship between the justification of harm and the justification of authority reveals that certain widely held views about the ethics of harm – *Completeness* and *Immunity* – are only tenable in conjunction with the view that the command of an authority cannot give

---

<sup>207</sup> This case is inspired by Anna Stilz’s description of the annexation of Hawaii by the United States in 1893. Stilz ‘Authority, Self-Determination and Community in Cosmopolitan War’.



agents undefeated reasons to cause content-dependently unjustified harm. I have argued that taking a plausible and influential approach to the justification of authority and extending into the domain of harm shows this view to be mistaken and that, therefore, *Completeness* and *Immunity* are also mistaken. The best theory of permissible harm will need to make space for both authority-based justifications and authority-based constraints. In the following chapter I assess the implications of the Authority View of Harm for theorising about the ethics of war.

## Chapter 5 – Authority and War

### 5.1 Introduction

In the previous chapter I set out and defended the Authority View of Harm, which aimed to demonstrate that, under certain conditions, the commands of authorities may alter the deontic status of acts of harming. My aim in this chapter is to apply this fairly abstract account of the relationship between the justification of authority and the morality of interpersonal harming to the specific case of warfare and to draw out its implications for just war theorising. In particular, I assess the extent to which the Authority View provides the resources for a rehabilitation of the authority criterion.

I begin by providing a rough account of the conditions under which an authority may be legitimate regarding warfare under the Authority View and the normative consequences that follow from the possession of such authority (Section 5.2). I then explain how these normative consequences enable a qualified defence of the authority criterion in its permissive role (Section 5.3) and, to a lesser extent, its restrictive role (Section 5.4). Most saliently, the Authority View holds that, in certain cases, individuals may be morally justified in obeying commands to participate in unjust wars.

In Section 5.5 I compare and contrast the Authority View with reductivist approaches to the ethics of war. I argue that while the two positions come apart in important respects, the Authority View is broadly compatible with the basic commitments underlying reductivism. I emphasise this point in Section 5.6, where I point out that, contrary to what is often supposed, the view that the permissibility of conduct in war is authority-sensitive need not invoke the idea that combatants who act under authority are thereby less morally responsible for their actions. Under the Authority View individuals remain fully responsible moral agents even when they are subject to authoritative institutions.

In the remainder of the chapter I clarify and refine the Authority View in light of two objections that may be pressed against it. In Section 5.7 I outline an objection which holds that the Authority View, even if correct,

has very little practical relevance, because no real-world authorities will in fact satisfy the conditions it requires for having legitimacy regarding warfare. In response (Section 5.8), I argue that authority over warfare can be justified from much more modest assumptions about the character and competence of existing political authorities than one might think. This is because authority over warfare need not be all-or-nothing. Rather, an authority may be legitimate regarding certain aspects of warfare and not others, depending on the facts of the case. In Sections 5.9 and 5.10 I refine my initial presentation of the Authority View in light of this ‘fine-grained’ approach to justifying authority over warfare.

In Section 5.11 I outline a second objection, which argues that the Authority View, when applied to war, falls foul of the war crimes objection that I pressed against non-reductivist views in Chapter 3. The objection draws its force from the fact that both the Authority View and non-reductivist views offer agent-relative forms of justification for killing in war. In response (Section 5.12), I argue that there is a subtle, but important, difference between the forms of agent-relativity that underpin each view, which shows that that Authority View is not susceptible to the war crimes objection to nearly the extent that non-reductivism is. However, I concede that there will be special cases in which the Authority View does justify the targeting of non-liable non-combatants, but I argue that this conclusion is defensible in these cases. Section 5.13 concludes.

## **5.2 The Authority View of War: A First Cut**

According to the Authority View of Harm, defended in the previous chapter, under certain conditions, the command of an authority is capable of altering the deontic status of acts of harming. This view was derived by applying a more general service account of authority to a particular class of harm-apt domains. According to resulting Authority View of Harm:

***The Authority View of Harm:*** A has authority over B regarding a domain of reasons that govern the distribution of harm, if B achieves greater overall

conformity with these reasons by obeying A's commands than by not obeying.

On this view, when an authority is legitimate regarding a subject within a harm-apt domain, that authority has the moral power to give that subject undefeated reasons to cause or refrain from causing harm. Importantly, this may be true even in cases in which the agent would otherwise have sufficient or decisive reason not to act as commanded. This is because authorities do not need to be infallible in order to provide their legitimating service. The commands of legitimate authorities remain binding even when they are (non-clearly) mistaken and fail to reflect right reason. Commands are thus capable of providing an independent source of both justification for causing harm and constraints on causing harm.

We can draw out the implications of the Authority View of Harm for just war theorising by a further stage of domain-specification. For our purposes, we are interested in one particular harm-apt domain – that of the reasons that govern the distribution of harm *by means of war*. According to the resulting *Authority View of War*:

***The Authority View of War***: A has authority over B regarding the domain of warfare, if, by obeying A's commands, B achieves greater overall conformity with the reasons that govern the distribution of harm by means of war than by not obeying.

The Authority View of War provides a very general account of what the possession of war-waging authority consists in and the normative effects that follow from its possession. To possess legitimate authority regarding war is to possess the moral power to provide subjects with undefeated content-independent reasons for action in the domain of warfare, including actions that involve causing serious harm to others.<sup>208</sup> I will offer a more

---

<sup>208</sup>For a different application of a service account of authority to warfare, see Adil Ahmed Haque, 'Law and Morality at War', *Criminal Law and Philosophy* 8 (2014), 79-97. Haque is concerned with the authority of the laws of war, whereas I focus primarily on the authority of leaders and governments.

refined version of the Authority View of War in Section 5.9 in light of the discussion to follow.

### **5.3. A Qualified Defence of the Permissive Authority Criterion**

With this account of legitimate war-waging authority to hand, we are now in a position to provide a qualified defence of the traditional just war authority criterion. I start with the permissive role of the criterion. According to the permissive interpretation of the criterion, as formulated in Chapter 2:

***The Permissive Authority Criterion:*** If an agent participates in war on behalf of an entity that possesses a certain set of properties, then that agent is subject to more extensive permissions to cause harm than if they fought privately, or on behalf of an entity which lacks that set of properties.

The Authority View of War provides an account of the properties that are capable of generating this normative effect, through the notion of an authority-based justification for causing harm. If an authority is legitimate over a subject regarding the domain of reasons that determine right action in war, that authority's commands may provide the subject with decisive reason for causing harm, even when that harm would not be justified on the basis of the command-independent reasons. More specifically, acts of killing and injuring in war may be morally justified even when the recipient of harm possesses a right against its infliction and their right is neither waived nor overridden by weightier reasons in favour of harming.

The most salient implication of the Authority View of War is that the command of an authority may provide an agent with a justification for participating in a war that is *ad bellum* unjustified. If a subject will better conform to reason by obeying an authority on matters of *jus ad bellum* compared to not obeying, then they have all-things-considered reason to obey the commands of that authority regarding participation in wars. Importantly, this includes commands that fail to reflect right reason

(though not clearly so), and require participation in wars that are *ad bellum* unjustified. I discuss this implication in more detail in Section 5.8.

It is worth making clear that, on this view, it is the command of a legitimate authority *itself* that provides the subject with a justification for causing harm, by placing them under an undefeated obligation to do so. The view is not, as some have suggested, that the normative effect of participating in war on behalf of a legitimate authority is to alter the conditions of permissible self-defence. On this alternative view, agents who obey legitimate commands to fight in unjust wars are permitted in causing harm to their opponents because they retain their normal rights of self-defence, which they would have forfeited if they fought in the absence of authority.<sup>209</sup> The Authority View, by contrast, makes no appeal to considerations of self-defence. It holds that authorised agents may be justified in causing unjust harm in war, *despite* lacking a permission to do so on grounds of self-defence.<sup>210</sup>

#### 5.4 A Qualified Defence of the Restrictive Authority Criterion

I now turn to the question of whether the Authority View of War can provide a defence of the authority criterion in its standard restrictive role. As formulated in Chapter 2, the restrictive interpretation of the criterion can be understood as follows:

***The Restrictive Authority Criterion:*** In order for a war to be justified, it must be initiated and fought by an entity that possesses a certain set of properties.

---

<sup>209</sup> Estlund, 'On Following Orders in an Unjust War', p.226. (Though Estlund qualifies this at p.228.) See also, Fabre, *Cosmopolitan War*, pp.160-161.

<sup>210</sup> The Authority View is therefore consistent with McMahan's claim that "a person can have no right of defense against a threatened harm to which he has made himself liable." *Killing in War*, p.14. What it denies is that a person's liability for threatening an unjust harm entails that they have no permission to cause that harm, since it posits an additional form of justification that obtains independently of considerations of liability and self-defence.

The Authority View does not support this interpretation of the criterion. It provides an account of what legitimate war-waging authority consists in and the normative consequences that follow from its possession. It makes no particular claims about the normative consequences of its *absence*.<sup>211</sup> More specifically, it does not claim that it is impermissible for agents to initiate and wage wars in the absence of a legitimate authority.

In fact, more strongly, the Authority View *cannot* support such a claim. As explained in Chapter 2, the restrictive authority criterion rests on a certain view about the range of justifications for causing harm that apply to individuals *qua* private actors, independently of authoritative institutions – that these justifications are insufficient to justify the scale and type of harming constitutive of war. Given the justificatory structure of service accounts of authority, the Authority View of War is incompatible with this position. Service accounts are based on what Stephen Perry has termed the ‘Political Subordination Thesis’, which holds that the only reasons which may be ultimately be appealed to in justifying authority are those which apply to individuals in their private capacity.<sup>212</sup> Given this, in order for an authority to legitimately command subjects to perform certain acts it must be the case that the act is of a type that *could* be justified solely on the basis of these private reasons. Hence, if it were true that the reasons that apply to individuals *qua* private actor are incapable of justifying the waging of wars, the command of an authority would not be able to render participation in war permissible on service-based grounds. In other words, if pacifism is true at the level of private morality, it remains true at the level of political morality. In the terminology that I have adopted, if this conception of permissible interpersonal harming is true, warfare would not be a harm-apt domain.

This point highlights the key difference between the Authority View of war and the conception of war and authority present in classical just war

---

<sup>211</sup> Beyond the trivial point that, in the absence of authority, the permissibility of agent’s conduct in war will be determined solely by authority-independent reasons.

<sup>212</sup> Stephen Perry, ‘Political Authority and Political Obligation’, in Leslie Green and Brian Leiter (eds), *Oxford Studies in Philosophy of Law: Volume 2* (Oxford: Oxford University Press, 2013), 1-74 at p.58. For Raz’s endorsement of the thesis, see *The Morality of Freedom*, p.72.

theory, discussed in Chapters 1 and 2. While both views hold that individuals who fight on behalf of a legitimate war-waging authority enjoy a more extensive range of permissions to cause harm, the classical view denies the Political Subordination Thesis. Since, on this view, the waging of wars cannot be justified in terms of permissible interpersonal harming, the role played by authority in justifying participation cannot be explained in terms of private morality. On the classical view, if acting under authority serves to justify actions in war that would not otherwise be morally permitted, it must be because actions carried out under authority are governed by moral principles that are *sui generis*.

This discussion also reveals an interesting tension between my qualified defence of the permissive authority criterion put forward in the previous section and the standard, restrictive understanding of the criterion's role. The affirmation of the former entails the denial of the latter. In order for individuals to be justified in causing harm in war in virtue of an authoritative command – as the former claims – it has to be true that individuals could potentially be morally justified in inflicting warlike violence in the absence of an authority – as the latter denies.

However, the inability of the Authority View to support the restrictive authority criterion need not count against the view. As explained Chapter 2, the permissive and restrictive interpretations of the criterion make logically independent claims, in that a commitment to one does not entail any commitment to the other. Furthermore, as we saw, there are powerful objections to the view that aggregates of private individuals cannot fight justified wars, encapsulated in the First Spectrum Argument. Given this, the fact that the Authority View does not, indeed cannot, rule out the possibility of justified private war may count in its favour. Reductivists are correct to argue that the traditional restrictive criterion is indefensible and ought to be jettisoned.<sup>213</sup>

---

<sup>213</sup> Despite the incompatibility of the Authority View with the restrictive criterion, it is worth pointing out that under certain contingent circumstances, the Authority View could support something like this version of the criterion. For example, it could be the case that a group of individuals, acting on their own assessment of the reasons that govern the initiation and waging of war, would achieve a sufficiently poor level of success at conforming to reason that they would be morally required to adopt a blanket policy of



The Authority View of War does, however, contain the resources for a revised and weakened interpretation of the restrictive authority criterion. This is provided by the notion of an authority-based constraint on causing harm, as outlined in Section 4.15. Under the Authority View, an authority that is legitimate in the domain of the reasons governing the distribution of harm by means of war possesses the moral power to give its subjects decisive reasons *not* to initiate or participate in wars. This may be so even in cases in which the war satisfies all the conditions of *jus ad bellum* and participation would be straightforward justified on the basis of the authority-independent reasons. The Authority View is thus able to provide an additional and independent constraint on participation in war. A group of individuals may possess a just cause and the means to prosecute it at a proportionate cost, and yet be under an undefeated obligation to refrain from doing so.<sup>214</sup>

One might think the notion of an authority-based constraint on participation in war has very narrow scope, since it only applies to agents who are subject to authorities that are legitimate over the domain of warfare. It therefore cannot really provide a defence of the restrictive authority criterion, which claims to provide a necessary condition for the justification of all wars. However, we can reformulate the criterion as follows:

---

never resorting to war. In other words, given their abilities they ought to be contingent pacifists. Furthermore, it could be the case that an authority is able to serve these individuals regarding the *ad bellum* reasons, thereby enabling them to achieve a level of conformity with reason above the threshold below which pacifism is mandatory. Under these background conditions, it would perhaps be true that these individuals would only be permitted to resort to war if sanctioned by the authority. In other words, these individuals, in these circumstances, would be subject to something like the restrictive authority criterion. But, of course, these conditions are highly stipulative and do not hold universally. The possibility of such cases therefore does not provide a defence of the traditional restrictive authority criterion, which is intended to apply to all wars.

<sup>214</sup> Again, clearly mistaken commands do not impose such obligations, even when issued by an authority that is generally legitimate. In such cases, it is possible that the use of force may be justified even in the face of a command not to do so. To give a speculative example, it may have been the case that UN peacekeeping forces in Rwanda in early 1994, led by Gen. Romeo Dallaire, would have been justified in using force to prevent the unfolding genocide, despite being ordered not to intervene by their superiors. If we grant the assumption that the UN leadership at the time possessed legitimate authority over its agents on service-based grounds, such a command may still have been invalidated as clearly mistaken, given the obviousness of the facts on the ground. For a first-hand account, see Romeo Dallaire, *Shake Hands With the Devil: The Failure of Humanity in Rwanda* (Toronto: Random House, 2003).

***The Revised Restrictive Authority Criterion:*** An agent's participation in a war is justified only in the absence of an authoritative command to refrain from doing so.

This negative formulation of the criterion is attractive in that it is universal in scope – placing an authority-based requirement on permissible participation in all wars – while not ruling out the possibility of justified wars fought by individuals in their private capacity. This is because there are two ways in which the criterion can be satisfied. Firstly, a war may be sanctioned by an agent or institution with the appropriate war-making authority. Secondly, a war may be fought by a group of individuals who are not subject to a legitimate war-making authority. In both kinds of case, the potential belligerents are not subject to an authority-based constraint on participation.

## **5.5 The Authority View of War and Reductivism**

This qualified defence of the authority criterion stands in opposition to the sceptical arguments offered by reductivists in support of jettisoning the authority criterion from just war theorising. In particular, my defence of the permissive interpretation of the criterion denies the core reductivist claim that acts of killing and injuring in war are only permissible given the existence of a liability or lesser-evil justification, thus providing a qualified defence of the permissibility of fighting in unjust wars. In this section I clarify the Authority View of War by drawing out more precisely the points at which it diverges from reductivism. I aim to show that, despite their important differences, my defence of the authority criterion shares some of the core commitments of a reductivist view.

As argued in Chapter 1, a major source of motivation for the inclusion of an authority criterion within the just war tradition is a commitment – either implicit or explicit – to the view that conduct in war is morally exceptional, governed by a distinct, and more permissive, set of moral

norms than those that apply to ordinary cases of interpersonal harming. Given this, we saw in Chapter 2 that the reductivist commitment to the Continuity Thesis – which holds that killing and injuring in war are governed by precisely the same moral principles that govern killing and injuring in all other contexts – undermines one powerful source of conceptual pressure for including an authority criterion within just war theory.

My defence of the authority criterion is wholly compatible with the Continuity Thesis. It does not argue that war, or political action more generally, is in any sense morally *sui generis*. Rather, it aims to defend the criterion by applying a very general account of how one individual's authority over another can be justified – which does so in terms of the reasons that apply to individuals in their private capacity – to the particular domain of warfare. The argumentative strategy has been to show that if we accept a service-based account of legitimacy in the more mundane domains of action in which authority is usually discussed, then there is no principled basis for denying that subjects may have duties to obey in harm-aft domains, including the domain of warfare. In fact, this argument only goes through given the truth of the Continuity Thesis. To deny that a service account of authority is applicable to the domain of warfare would require one to argue that war is morally anomalous in some important respect, thus violating the Continuity Thesis.

Furthermore, the employment of a service account in defence of the authority criterion is consistent with the individualistic and aggregative approach to the ethics of warfare endorsed by reductivists. On such an account, authority claims are evaluated solely on the basis of the relationship between the authority and each individual over whom it claims the power to impose obligations. The relationships *between* subjects plays no independent role here. The extent of an authority's legitimacy is simply the sum of its legitimacy over individuals. As Raz puts it, "it is an advantage of the analysis...that it is capable of accounting for authority

over a group on the basis of authority relations between individuals.”<sup>215</sup> So, while my defence of the authority criterion places substantial moral significance on the fact that wars are often waged by political authorities, which reductivists are deeply hostile to, it does so while maintaining a commitment to the view that individuals are the ultimate unit of moral concern.<sup>216</sup>

Where the Authority View and reductivism do come apart sharply is over the truth of the Content Thesis – the view that liability and lesser-evil justifications exhaust the range of justifications for killing and injuring in war. The Authority View denies the Content Thesis, since it holds that subject of authorities may, under certain conditions, be obligated to cause harm that is not justified on the basis of these twin principles. But if the arguments of Chapter 4 are successful, this is the correct conclusion to draw. The Content Thesis is false, for the same reason that the broader view that I termed *Completeness* is false. Reductivists, and theorists of the morality of harming more generally, hold a mistakenly narrow view of the range of justifications for causing harm.

The Authority View does not deny the reductivist insight that certain actions in war commonly assumed to be permissible – such as participating in an unjust war – involve extremely grave rights violations. Rather, it denies the intuitive assumption upon which the Content Thesis rests – that commands to perform such actions cannot give subjects undefeated reason to obey. The Authority View sets out the conditions under which such commands need not exceed the moral limits of the duty to obey.

However, while the Authority View denies that liability and lesser-evil exhaust the range of justifications for killing and injuring in war, the Content Thesis may still play an important justificatory role within the Authority View. On a service-based view, authorities are justified in virtue of their enabling subjects to better conform to the reasons that apply to

---

<sup>215</sup> Raz, *Morality of Freedom*, p.71.

<sup>216</sup> The Authority View is thus consistent with McMahan’s position that “There is nothing in the relations that people establish *among themselves* within a state that can extend their permissions to harm people outside the state”. McMahan, ‘War’, p.310 (emphasis added). What the Authority View denies is that the relations between each individual and a common authority cannot extend their permissions to cause harm.

them independently. But service accounts are entirely neutral with respect to the question of what these substantive first-order reasons consist in.<sup>217</sup> Given this, the Authority View of War is compatible with the Content Thesis being correct as an exhaustive account of the *authority-independent* reasons that govern the permissibility of causing harm in war. While it denies that liability and lesser-evil exhaust the range of justifications for harming, it can accept the view that *all* justifications for harming in war are ultimately, if sometimes indirectly, grounded in those two principles.

This potential justificatory relationship between the notion of an authority-based justification for harm and the forms of justification specified in the Content Thesis highlights a key difference between the Authority View and alternative approaches to the morality of war that also aim to deny the Content Thesis, such as the associative-duty view of killing in war discussed in Chapter 3. These views aim to deny the Content Thesis by positing the existence of additional justifications for harming at the level of the first-order, authority-independent reasons. They thus reject the Content Thesis at a more fundamental level than the Authority View need do.

Of course, since service accounts are neutral regarding the reasons that apply to subjects independently, the Authority View is also compatible with accounts of the ethics of war that deny the Content Thesis. For example, one could endorse both the associative-duty view and the Authority View of War. On this combination of views, killing in war can be justified in virtue of the strength of individual's duties to protect their associates, and an authority may acquire legitimacy over the domain of warfare by enabling its subjects to better discharge these duties. So, although I have stressed the (qualified) compatibility of the Authority View of War with a reductivist outlook, this should not be taken to mean that one has to be a reductivist in order to accept the Authority View.

I conclude this section by providing a deflationary explanation of why the Second Spectrum Argument for the Content Thesis, set out in Chapter 2, *seems* highly persuasive. Recall, in McMahan's spectrum example, we

---

<sup>217</sup> On this point, see Raz, 'Facing Up', pp.1184-1185

consider an escalating range of cases, starting with a small-scale case of interpersonal harming involving a handful of individuals exercising their rights of self- and other-defence. We then, by a process of aggregating such cases, move along a spectrum of cases which eventually reach a level of violence constitutive of a state of war.

Reductivists argue that if the range of justifications for harming is exhausted by considerations of liability and lesser-evil in the small-scale cases, then the same is true of harming in a state of war – in other words, that the Content Thesis is true. All that changes as we move across the spectrum is that the cases become more complex, organised and protracted, but the basic moral principles are invariant. Given that the conditional claim seems highly plausible, as does the antecedent, reductivists conclude that we must reject the idea that individuals enjoy any additional permissions to cause harm in wars in virtue of fighting on behalf of a certain kind of entity.

The Authority View obviously rejects this conclusion. It accepts the conditional claim but denies the antecedent, holding that the range of potential justifications for harming is broader across the whole spectrum of cases. It does so by positing the possibility of authority-based justifications for harming. The truth of the antecedent merely *seems* highly plausible in the small-scale cases because these are case in which considerations of authority have little practical relevance. In these cases, as a contingent matter, authority-based justifications for harming are very unlikely to be activated. It is hard to see how an authority could potentially serve its subjects regarding the reasons that govern the distribution of harm in small-scale self-defence cases. In fact, it may well be a necessary condition of permissible self-defence that the relevant authorities are not able to intercede. Such cases, while being harm-apt, are unlikely to be *authority-apt* on a service-based view.<sup>218</sup> It is this contingent feature of the small-scale cases at the beginning of the spectrum, I contend, which accounts for

---

<sup>218</sup> This is not to deny that there *could* potentially be small-scale defence cases in which an authority is well placed to serve its subjects. My point here is just that the more small-scale and unstructured the case is, the less likely it is that an authority will be able to provide this service.

the intuitive plausibility of the claim that the range of possible justifications for harming in these cases is exhausted by considerations of liability and lesser-evil.

However, as we move across the spectrum and the cases become more complex, protracted and organised, the situation becomes increasingly authority-apt. At some point along the spectrum, an authority may be well placed to serve its subjects regarding the distribution of harm, by enabling them to coordinate their actions or by offering expert guidance. When this is so, a pre-existing but latent form of justification for causing harm may be activated.

## **5.6 Authority and Responsibility**

In this section I distinguish the Authority View of War from an alternative argument that is often made in support of the conclusion that individuals may be permitted to fight in unjust wars in virtue of doing so under authority. On this alternative view, the source of an authority-based permission to cause harm lies in the fact that individuals acting under authority are not responsible for their actions in the same way that they would if they were acting privately.<sup>219</sup> The thought here is that responsibility for the subject's actions lies solely with the authority under which they act, thereby absolving the subject of wrongdoing if their actions are unjust, including participation in an unjust war.

In contrast, the Authority View makes no such claims about diminished or transferred responsibility. In fact, the Authority View is premised on the idea that subjects remain fully responsible moral agents even when obeying an authority. On this view, the command of an authority provides reasons for action, which the subject recognises and responds

---

<sup>219</sup> This seems to have been Augustine's view, see *The City of God Against the Pagans*, R.W. Dyson (ed), (Cambridge: Cambridge University Press, 1988), Bk.1. Ch.21. For discussion, see Langan, 'The Elements of St. Augustine's Just War Theory'. Graham Parsons has recently put forward an explicit version of the view that the normative effect of a fighting on behalf of an is to negate the subject's moral responsibility for acts of wrongdoing. Graham Parsons, 'Public War and the Moral Equality of Combatants', *Journal of Military Ethics* 11, No.4 (2012), 299-317.

appropriately to by obeying. This emphasis on reason-responsiveness is hard to square with a denial of the subject's responsibility.

This feature of the Authority View highlights another respect in which it is compatible with a core commitment of the reductivist view. Reductivists are keen to insist that individuals remain fully autonomous and responsible moral agents even if they are members of hierarchical organisation such as the military, and have a duty to avoid unjustly harming and killing others which cannot be evaded by claims of superior orders. The Authority View is in complete agreement with this. Where the Authority View diverges from reductivism is in claiming that, under certain conditions, obeying an authority is the appropriate means of discharging this duty.

As well as being consistent with reductivism on this point, the Authority View of War also has a number of independent advantages over the alternative 'Responsibility View'. Firstly, the claim that those acting under authority lack moral responsibility for their actions lacks intuitive plausibility. Normally, a successful defence of non-responsibility to a charge of wrongdoing involves claims to limited cognitive capacity, irrationality or insanity.<sup>220</sup> Yet (at least the majority of) combatants are clearly capable of rational deliberation and have control over their own behaviour. In the absence of a detailed account of how responsibility can be negated by acting under authority, the Responsibility View is unsatisfying.<sup>221</sup> The Authority View, by contrast, does need to provide any such account.

Secondly, even if such an account can be provided, the Responsibility View is unable to support the conclusion that those who participate in an unjust war possess a genuine moral justification for doing so. The fact that one would not be responsible cannot provide one with a reason to fight in an unjust war, since the fact that one would not be responsible is not a

---

<sup>220</sup> See, Gardner, *Offences and Defences*, pp.77-90.

<sup>221</sup> Frances Kamm has defended the view that permissibility may be sensitive to the moral responsibility of others, but the cases she discusses are disanalogous to those under discussion here. Frances Kamm, *Intricate Ethics: Rights Responsibilities and Permissible Harm* (Oxford: Oxford University Press, 2007), Ch.10. For criticism of aspects of Kamm's position, see Jeff McMahan, 'Responsibility, Permissibility and Vicarious Agency', *Philosophy and Phenomenological Research* 80, No.3 (2010), 673-680.



reason to do anything. At best, the Responsibility View can give rise only to an excuse for fighting in an unjust war on the basis of a legitimate command, not a permission or justification.<sup>222</sup> The Authority View, by contrast, is straightforwardly able to claim that those who fight in unjust wars under authority may be justified in doing so, since it treats authoritative commands as providing reasons for action, rather than defeaters of moral responsibility.

Thirdly, the Responsibility View is vulnerable to the war-crimes objection that I pressed against non-reductive approaches to the morality of war in Chapter 3. The problem is that if acting under authority serves to negate an agent's responsibility for killing non-labile combatants, such as by fighting in an unjust war, it is unclear why acting under authority does not also negate responsibility for the targeting of non-labile non-combatants. I discuss the war-crimes objection in relation to the Authority View in detail in Sections 5.10 and 5.11.

Fourthly, the Responsibility View runs into an additional problem when we consider the question of the permissibility of using defensive force against those who threaten unjust harm but do so under authority. A major topic of debate within the self-defence literature concerns the permissibility of defensively killing individuals who causally threaten lethal harm to innocent people, but are not morally responsible for doing so. Most theorists hold that a degree of moral responsibility for posing a threat is a necessary condition for liability to defensive force.<sup>223</sup> Defensive harms inflicted on non-responsible threateners are therefore unjust and subject to a much more stringent burden of justification than defensive harm inflicted on morally responsible threateners. If, as the Responsibility View claims, those who participate in unjust wars under authority are not responsible for doing so, this suggests that authorised threateners should be understood as a species of non-responsible threatener. However, this leads

---

<sup>222</sup> McMahan, *Killing in War*, pp.89-90. For discussion of the relationship between responsibility and excuse, see Gardner, *Offences and Defences*, Ch.4.

<sup>223</sup> See, for example, McMahan, 'The Basis of Moral Liability to Defensive Force'; Rodin, *War and Self-Defense*, Ch.4; Quong, 'Killing in Self-Defense'; Michael Otsuka, 'Killing the Innocent in Self-Defense', *Philosophy and Public Affairs* 23, No.1 (1994), 74-94. For dissent, see Thomson, 'Self-Defense'.

to some implausibly pacifistic implications. It implies that the victims of unjust military aggression may be morally required to suffer the threatened harm, if avoiding it involves intentionally killing authorised threateners. Furthermore, it generates different results in cases that are morally similar, since the permissibility of resisting aggression will turn on whether the threateners act on behalf of an authority or not, even if the threatened harms are identical.

Again, the Authority View avoids this problem. Since it treats authorised threateners as fully responsible agents, they are straightforwardly able to render themselves liable to defensive force when they threaten unjust harm to others. As argued in Section 4.14, under the Authority View the fact that an agent possesses an authority-based justification for causing unjust harm does not, in itself, raise the justificatory burden for defensively harming that agent, since such the reasons that justify the agent in obeying are fully agent-relative. An agent's lack of responsibility, on the other hand, gives rise to an agent-neutral reason against inflicting defensive harm on them that applies to all agents, including their victims.

### **5.7 The *A Posteriori* Objection**

Thus far, my qualified defence of permissive authority criterion has primarily been constructed in opposition to reductivist just war theorists, who defend the intuitive position that the command of an authority cannot possibly render justified acts of harming that would otherwise be impermissible.

In this section I consider a different form of challenge, which accepts the notion of an authority-based justification for harming in principle, but holds that their possibility has very little practical significance for just war theorising. This is because, the objection holds, it is extremely unlikely that real-world authorities of the kind we are familiar with – states most obviously – will *in fact* satisfy the conditions required to generate such justifications. According to this objection, my defence of the permissive authority criterion, while theoretically distinct, will be practically

equivalent to a reductivist view. In particular, critics of this stripe argue that realistic states lack the moral power to give their subjects decisive reason to fight in unjust wars.<sup>224</sup> While authority-based justifications for participating in unjust wars are possible, no individuals in fact possess such a justification.

The difference between these two forms of objection has a structural similarity to the distinction between *a priori* and *a posteriori* arguments for philosophical anarchism – the view that no states are legitimate and that no individuals have political obligations. Whereas *a priori* anarchists hold that states *necessarily* lack legitimacy and that the justification of political obligations is *impossible*, *a posteriori* anarchists accept that states could potentially be legitimate under certain conditions, but argue that certain contingent, yet deeply pervasive, features of states prevent them from satisfying these conditions.<sup>225</sup> We may then characterise the current objection to my defence of the permissive authority criterion as a much more limited form of *a posteriori* anarchism. Whereas the *a posteriori* anarchist denies that states possess general authority over their subjects, the *a posteriori* objection to my proposal denies only that states have the authority to place their subjects under undefeated obligations to cause content-dependently unjustified harm in war, such as by participating in wars that are unjust. Since, according to the Authority View of War, authorities acquire this moral power because they successfully serve their subjects in the domain of the reasons that govern warfare, the *a posteriori* objection holds that no real-world authorities successfully provide such a service. In other words, it holds that individuals' optimal policy for achieving conformity with the war-related reasons will not be that of obeying the commands of their state.<sup>226</sup>

---

<sup>224</sup> McMahan, *Killing in War*, pp.69-70. Jeff McMahan and Thomas Hurka have also raised versions of this objection in correspondence and in conversation, respectively.

<sup>225</sup> For this distinction, see A. John Simmons, 'Philosophical Anarchism' in John T. Sanders and Jan Narveson (eds), *For and Against the State: New Philosophical Readings* (Boston: Rowman and Littlefield, 1996), 19-40

<sup>226</sup> This is one way of interpreting Uwe Steinhoff's claim that, "the deliberation process of a single individual may at times be much more balanced than a public political discourse infected with patriotism, or a decision process supported by a bureaucracy and advisory staff made up of yesmen and opportunists." *The Ethics of War and Terrorism*, p.20 n.34

The objection can be stated in weaker or stronger forms, depending on the extent of one's scepticism regarding the ability of states to serve their subjects. In the strongest possible terms, one might hold that states generally fail to serve their subjects in all domains.<sup>227</sup> A weaker view may accept that states successfully serve their subjects in certain domains, but deny that this is true of the harm-apt domains, including the domain of warfare. This view accepts that real-world authorities may successfully create decisive reasons to action on grounds of service, but denies that this ever in fact extends to acts of causing harm. The weakest version of the objection accepts that real-world authorities may serve their subjects in a wide-range of domains, including some that are harm-apt, but denies that states successfully serve their subjects regarding the particular domain of warfare. This view grants that authority-based justifications for harming are realistically possible, but denies that this is true in the case of war.

The weakest version of the objection is the most powerful, for two reasons. Firstly, it makes the most modest empirical claim necessary to reach its conclusion. Secondly, there may plausibly be domain-specific reasons for thinking that real-world authorities fail to successfully serve their subjects regarding warfare.

The objection can be pressed in two different ways, corresponding to two dimensions along which authorities may fail to serve their subjects regarding warfare. Firstly, authorities may fail to issue their directives on the basis of a good-faith assessment of the relevant reasons that apply to their subjects. Their claim to legitimacy over the domain of warfare then fails at the first hurdle. Term this the 'maleficence objection'. Secondly, even if making the appropriate efforts, authorities may fail to serve their subjects by doing an insufficiently good job of assessing the relevant reasons and crafting directives designed to improve their subjects' conformity with them. They lack the competence required to pass the test of normal justification regarding the domain of warfare, thus failing the test of normal justification. Term this the 'incompetence objection'. In light of

---

<sup>227</sup> This strong claim does not entail a commitment to *a posteriori* anarchism, since it may be combined with the view that states are legitimate on non-service-based grounds.

this distinction, we can interpret McMahan's complaint that states and their leaders are typically "transparently immoral, dishonest and intellectually bovine"<sup>228</sup> as combining both the maleficence and incompetence objections.

As mentioned, there are reasons to think that these objections are particularly applicable to the domain of war. One important way of motivating the maleficence objection is by pointing out that war, unlike other harm-apt domains, primarily involves inflicting harms on individuals who are not members of one's political community.<sup>229</sup> Given this, it may plausibly be argued that since states and their leaders are accountable only to their own citizens, they are heavily disincentivised from giving the appropriate weight to the interests of outsiders who will be harmed by resorting to war.<sup>230</sup> Furthermore, since states are highly militarised, investing huge sums of money on acquiring the means of fighting wars, one may argue that states are incentivised to make use of these means, independently of the merits of their use.<sup>231</sup> Relatedly, it has been argued that the war-making apparatus of states are disproportionately under the influence by private interests that profit from going to war, thus further increasing the likelihood that states will resort to war unjustly.<sup>232</sup> All these factors, and many more, fuel the objection that states typically lack authority over warfare in virtue of their failure to assess and act upon the right reasons.<sup>233</sup>

---

<sup>228</sup> McMahan, *Killing in War*, p.3.

<sup>229</sup> Civil wars being a possible exception. Though in such cases, where a political community has fissioned to the point of civil war, it seems doubtful that belligerents consider each other co-members with equal standing.

<sup>230</sup> Michael Huemer makes the point that in general individuals have a lesser aversion to causing the deaths of those who they perceive as different or alien, and that this fact makes states more prone to go to war unjustly. Huemer, *The Problem of Political Authority*, p.237.

<sup>231</sup> Huemer, *The Problem of Political Authority*, p.306. This echoes Kant's worry that states having standing armies creates a grave risk of abuse. Immanuel Kant, 'Perpetual Peace: A Philosophical Sketch', in H.S Reiss (ed) *Kant: Political Writings*, 2<sup>nd</sup> Edition, (Cambridge: Cambridge University Press, 1991), 93-139.

<sup>232</sup> See, for example, Andrew Fiala, *Against Religion, Wars, and States: The Case for Enlightened Atheism, Just War Pacifism, and Liberal Democratic Anarchism* (Plymouth: Rowman and Littlefield, 2013), Ch.7.

<sup>233</sup> Interestingly, many who have pressed versions of the maleficence objection take it to support a non-absolutist form of pacifism. For example, Rawls writes that large and powerful states "are in some circumstances so likely to be unjust that one is forced to conclude that one must abjure military service altogether." John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971), p.381. See also, Andrew Fiala, *Practical*

The incompetence objection may also apply to a greater degree in the domain of war. Again, the fact that war involves interacting with outsiders provides support for this, since it may be argued that states are much less well placed with regard to assessing the behaviour of outsiders than of members. Their domestic competence at assessing right reason thus does not entail competence in terms of the reasons that govern interactions with members of other states. In addition, it has been argued that certain structural features of states impair their ability to make good decisions regarding war. For example, Michael Huemer argues that since states have a monopoly on war-making within their territory, the normal market mechanisms that lead to improved performance do not apply. Unlike service providers in other areas, states lose no market share by making poor decisions about war.<sup>234</sup> But perhaps the most powerful form of the incompetence objection is simply to point to the numerous examples in which states have made morally calamitous decisions regarding war.

## 5.8 Response

### 5.8.1 Three Concessive Remarks

In responding to the *a posteriori* objection, I begin with three broadly concessive remarks. Firstly, it is important to stress that in arguing for the possibility of authority-based justifications for participating in unjust wars, I am not attempting to provide a defence of the orthodox Equality Thesis, which holds that participation in unjust wars is *generally* permissible.<sup>235</sup> This conclusion would only follow from the Authority View of War if two empirical assumptions hold. Firstly, that all war-making entities – or, more modestly, all recognised states – possess legitimacy over their subjects in the domain of warfare. Secondly, that every command to participate in an

---

*Pacifism* (New York: Algora Publishing, 2004); Saba Bazargan, 'Varieties of Contingent Pacifism' in Frowe and Lang (eds), *How We Fight*, 1-17.

<sup>234</sup> Huemer, *The Problem of Political Authority*, p.237.

<sup>235</sup> Though my view may provide a maximally charitable account of why many people find the equality thesis plausible – it reflects the common (though often mistaken) conviction that their states have legitimate authority over them regarding warfare.

unjust war issued by such authorities is not invalidated as clearly mistaken. These assumptions are incredibly demanding and certainly false. So, if the Authority View were intended as a defence of the general permissibility of participation in unjust wars, the *a posteriori* objection would be decisive. Under the Authority View, an authority's possession of the moral power to require participation in unjust wars is dependent on some fairly specific background conditions obtaining, which can only be established on a case-by-case basis. The *a posteriori* objection is correct to conclude that such authority will be lacking in many cases, but this should be no more surprising than the fact that various other conditions of just war theory – just cause most obviously – are often not satisfied.

Secondly, it is worth emphasising that shifting from an *a priori* to an *a posteriori* argument against the Authority View of War remains a major concession on the part of the reductivist, since it involves giving up a commitment to the Content Thesis and admitting the possibility of authority-based justifications for causing harm. Given this, even if the *a posteriori* objection is successful in showing that there are no actual legitimate war-making authorities, the Authority View still succeeds in showing that considerations of authority cannot be eliminated from just war theorising in the way that reductivism claims. It remains the case that, in order to determine whether an agent acts permissibly in war, it is necessary to establish whether or not that agent acts on behalf of a legitimate war-making authority. The authority criterion remains an ineliminable part of just war theory, even if it is paired with a deep scepticism as to the existence of such authorities.

Thirdly, and relatedly, even if the *a posteriori* objection is correct to hold that existing political entities lack the requisite authority, it is less plausible to maintain that such authorities could *never* exist. It does not seem completely utopian to think that future intuitional reform could produce authorities that are able to successfully serve their subjects within the domain of warfare. The topic of institutional reform has been relatively

unexplored by just war theorists.<sup>236</sup> One way of understanding the view presented here is as providing a sketch of the ideal to which our war-making institutions should aspire. So, while reductivists could be correct in holding that no current war-making entities possess the kind of authority capable of generating moral justifications for participating in unjust wars, they should still endorse taking steps towards creating institutions which would possess such authority. Hence, the success of the *a posteriori* objection would not entail that the Authority View does not have practical relevance.

### 5.8.2 Fine-Grained Domains – *Jus ad Bellum*

The foregoing remarks sought to show that considerations of authority remain relevant for just war theorising even if the *a posteriori* objection is successful. There is a world of difference between legitimate authority over warfare being impossible and its merely being highly unlikely in present circumstances. In what follows I argue that the possibility of legitimate war-making authority, and of authority-based justifications for causing harm in war, are not as unlikely as the *a posteriori* objection alleges. I do not attempt to resist the *a posteriori* objection on empirical grounds. This can only be done on a case-by-case basis. Instead, I offer some refinements of the Authority View of War with the aim of providing principled grounds for resisting the objection.

The force of the *a posteriori* objection comes from the natural assumption that in order to acquire the relevant legitimacy over the domain of war, an authority must successfully serve its subjects regarding the whole range of reasons that govern the distribution of harm by means of war. Most obviously, including the reasons that determine whether wars possess a just cause or not. This assumption sets a fairly high bar for acquiring legitimate war-making authority. The higher the bar, the more

---

<sup>236</sup> Exceptions include Jeff McMahan, 'The Prevention of Unjust Wars' in Yitzhak Benbaji and Naomi Sussmen (eds), *Reading Walzer* (Abingdon: Routledge, 2013), 233-255 and Allen Buchanan 'Institutionalizing the Just War', *Philosophy and Public Affairs* 34, No.1 (2006), 2-38.



powerful the *a posteriori* objection is. I aim to resist the *a posteriori* objection by denying that the assumption is true. Authorities need not satisfy such a demanding standard of service in order to acquire the moral power to create authority-based justifications for causing harm in war.

The *a posteriori* objection rests on a holistic conception of the task that an authority must perform in order to possess this moral power. However, this conception is mistaken. Recall, on a service-based view, authority is piecemeal with respect to domains of reasons. An authority's legitimacy must always be demonstrated anew with respect to each domain over which it is claimed. Importantly, the scope of these domains is not fixed at a certain level of description. Complex activities are usually governed by distinct sub-domains of reasons, corresponding to the different kinds of considerations that go in to determining what ought to be done, all-things-considered. For example, the practical matter of which trade policy to adopt with a specific country is determined by a range of distinct considerations, such as the economic benefits, the political consequences, the social impact, etc.

In order to acquire legitimacy regarding a certain activity, an authority need not serve its subjects regarding each and every one of the sub-domains of reasons that govern a particular activity. Just as the fact that an authority does not serve me regarding the domain of the X-related reasons does not entail anything about whether it serves me regarding the domain of the Y-related reasons, the fact that an authority does not serve me regarding the  $X_m$ -related sub-domain of the X-related reasons does not entail that it fails to serve me regarding the  $X_n$ -related sub-domain of the X-related reasons. When an authority serves its subjects regarding a sub-domain of reasons, its legitimacy is limited to the class of reasons contained within that sub-domain. But its commands are still *relevant* to the question of what ought to be done regarding the wider macro-domain of reasons and may, in certain cases, *decisively* settle that question.

This fine-grained approach to individuating domains of authority applies equally to the complex activity of warfare and provides the resources for a principled response to the *a posteriori* objection. As

mentioned earlier, the key target of the *a posteriori* objection is the claim, which I have defended, that an agent may be justified in participating in an unjust war in virtue of a command to do so. Given this, I will begin by explaining how the fine-grained approach allows this claim can be derived from fairly modest assumptions about states and their ability to serve their subjects. The key point is that an authority may successfully serve its subjects regarding a particular sub-domain of the *jus ad bellum* reasons, while failing to serve them regarding a distinct sub-domain. For example, an authority may fail to serve its subjects regarding the reasons that determine whether wars have a just cause, but successfully serve them on the matter of whether wars have a reasonable chance of success. This limited war-making authority is sufficient to generate authority-based justifications for participating in an unjust war.

When an authority is legitimate regarding only a sub-domain of *jus ad bellum*, its commands provide pre-emptive reasons for action, but only exclude those reasons that fall within that sub-domain. When it commands a subject to participate in a war, the subject is rationally free – indeed, morally required – to weigh the command against the reasons for and against participation arising from *ad bellum* considerations over which the authority is not legitimate. If the war is unjustified on the basis of reasons that fall outside the scope of the authority's legitimacy, these reasons not to participate will override the subject's obligation to participate. They all-things-considered ought not to participate. However, if the war is *ad bellum* unjustified only in virtue of considerations that fall *within* the sub-domain of reasons over which the authority is legitimate, the subject will have undefeated reason to participate in that unjust war if commanded to do so. To return to the above example: If an authority possesses legitimacy over a subject regarding the success-related *ad bellum* reasons, and issues a mistaken (but not clearly so) command to participate in a war that is unjust in virtue of lacking a reasonable prospect of success, then that subject has an authority-based justification for participating.

Once we recognise that authorities may serve their subjects regarding specific components of *jus ad bellum* and not others, and that wars can be

unjustified for a wide variety of reasons, we can generate the conclusion that there may be authority-based justifications for participating in unjust wars from much more modest assumptions about the ability of real-world states to successfully serve their subjects.

In fact, the required assumptions may be even more modest, since the relevant sub-domains may be more fine-grained than that of each requirement of *jus ad bellum*. Each of these sub-domains may itself be made up of distinct sub-domains of reasons, each of which is capable of determining whether that *ad bellum* requirement is satisfied. For example, whether a war has a reasonable prospect of success may be determined by a range of different considerations – the fighting ability of one’s troops, the level of enemy resistance, the ability of the economy to support the war, weather conditions, etc. – each of which is potentially determinative of whether a particular war satisfies the success condition.

Given this, an authority need only be legitimate over one of these narrower sub-domains in order to possess the moral power to impose undefeated obligations to participate in an unjust war. More formally, in order to acquire this moral power, an authority only needs to possess legitimacy over a *sufficiently important* sub-domain of the *ad bellum* reasons that determine the justification of resorting to war. An *ad bellum* sub-domain is sufficiently important, in the relevant sense, if the reasons contained within that sub-domain are potentially determinative of whether the resort to war is all-things-considered justified or unjustified.

Before moving on, it is worth considering a possible objection, which holds that these refinements of the Authority View involve a sleight-of-hand, since it is a far graver moral wrong to participate in a war that lacks a just cause than one whose injustice rests on other grounds, such as lacking a reasonable chance of success or failing to be the last resort. This is because in the first case, but not the second, one will be intentionally killing the non-liege. Given this, the objection holds, I haven’t really made good on my promise to show that we can generate authority-based justifications for causing seriously morally wrongful harm in war from fairly modest assumptions about states and their virtues.

There are three responses to this. Firstly, on certain accounts of liability, such as McMahan's, one can only be liable to harm that is both necessary and effective in averting a threat of unjust harm. Hence, one would be intentionally killing the non-liable in either type of unjust war.<sup>237</sup> Secondly, even if this difference in liability holds, killing large numbers of liable persons when doing so is either unnecessary or ineffective remains something that we have strong moral reasons to avoid doing. This is especially so when those who are liable possess excuses for their role in the existence of unjust threats. Thirdly, independently of whether participation in the second type of case involves intentionally killing non-liable combatants, such action also involves causing substantial collateral harm to innocents. Hence, even if it is true that participation in the first type of unjust war involves a graver degree of moral wrongdoing than in the second, both involve the commission of acts of very seriously wrongful harming.

### 5.8.3 Fine-Grained Domains – *Jus in Bello*

While the permissibility of participation in an unjust war is an important and high profile implication of the Authority View – one that proponents of the *a posteriori* objection understandably focus on – these cases far from exhaust its practical applications. Taking a fine-grained approach to individuating domains of war-waging authority also reveals that authority-based justifications for causing harm in war are possible, even if the authority in question lacks legitimacy over *any* aspect of *jus ad bellum*

This is because the moral status of actions performed in war are not fully determined by considerations that fall with the scope of *jus ad bellum*, though their status may of course be dependent on this domain of reasons. Whether specific actions in war are morally permissible is also determined by additional, less macro-level considerations. For example, the

---

<sup>237</sup> McMahan, *Killing in War*, pp.8-10. Some have found this implausible, see Joanna Mary Firth and Jonathan Quong, 'Necessity, Moral Liability and Defensive Harm', *Law and Philosophy* 31, No. 6 (2012), 673-701. For further discussion, see Frowe, *Defensive Killing*, Ch.4

permissibility of certain courses of action within an *ad bellum* justified war may depend upon: the degree to which particular targets are militarily valuable, whether specific mission or battles are necessary for achieving the war's aims, whether the aims of the war have been sufficiently advanced to cease fighting, etc.

The key point is that an authority may be sufficiently well placed to serve its subjects regarding distinctively *in bello* sub-domains of the war-related reasons, even if it fails to serve them regarding the *ad bellum* reasons.<sup>238</sup> Furthermore, these sub-domains also have the feature of being *sufficiently important*, in that the reasons contained within that sub-domain are potentially determinative of whether acts of causing harm in war are all-things-considered justified or unjustified. Given this, if an authority successfully serves its subjects regarding one of these sub-domains, it acquires the moral power to place its subjects under undefeated content-independent obligations to cause harm in war. Its subjects are bound to obey its commands within that sub-domain, including those that fail to reflect right reasons (though not clearly so). This is compatible with the authority lacking the moral power to require participation in unjustified wars.

Once it is recognised that authority over warfare can be piecemeal rather than holistic, the possibility of authority-based justification for causing harm in war can be generated from much more modest assumptions about the ability of real-world states to serve their subjects.

#### *5.8.4 Authority and Coordination*

This conclusion can also be supported by pointing out that authorities can serve their subjects in different ways. The thrust of the *a posteriori* objection focuses on showing that states fail to serve their subjects regarding war in virtue of their lack of expertise on relevant matters. But expertise is not the only ground, or indeed the main ground, for justifying

---

<sup>238</sup> David Estlund also notes that considerations of war-waging authority need not be restricted to the *ad bellum* domain. Estlund, 'On Following Orders in an Unjust War', pp.229-231

authority on a service-based view. In addition, authorities may serve their subjects on grounds of coordination. In cases where acting rightly depends upon coordinating one's actions with others, subjects may best comply with reason by obeying a common authority, rather than by each trying to act on the balance of reasons directly. This can be true even when an authority is fallible, and issues commands that fail to reflect right reason.

Though it is surprisingly rarely noted in the literature on war ethics, an obvious morally relevant feature of warfare is that it requires huge amounts of coordination.<sup>239</sup> Given this, an important ground of authority over warfare on service grounds will be that of achieving coordination. Many *in bello* sub-domains of the sort highlighted above will be authority-apt for precisely this reason. Importantly, grounding authority over warfare on coordination-based service further supports the possibility of authority-based justifications for harming in war. For, in order to achieve the rational benefits of having a coordinative authority over war, it will be necessary for subjects to obey its commands across a generalised range of cases. This may include cases where the authority issues mistaken commands that require causing harm that is not justified on the basis of the content-dependent reasons.

So, while one might argue that states typically lack the required expertise over warfare required for authority, it seems less plausible to hold that they do not have the ability to successfully coordinate the actions of their subjects in war. Again, the practical significance of the Authority View of War can be derived from fairly modest assumptions about states and their abilities.

To sum up, I have sought to blunt the force of the *a posteriori* objection by showing that conditions under which authorities may acquire the moral power to give subjects undefeated reasons to cause content-dependently unjustified harm in war are less demanding than one might assume. Correspondingly, the *a posteriori* objector will need to adopt a more extensive – and therefore less plausible – sceptical position regarding the capabilities of real-world states to serve their subjects in order to defend

---

<sup>239</sup> An exception being Lazar, 'Necessity in Self-Defence and War'.

the view that authority-based justifications for causing harm in war are extremely unlikely to arise.

### **5.9 The Authority View of War: A Second Cut**

The preceding discussion reveals that my initial characterisation of the Authority View of War requires revising, since it assumes a rough-grained view of the domain of reasons over which authorities must serve their subjects in order to acquire legitimacy over war. Instead, as argued above, we should endorse a fine-grained approach. The notion of ‘a domain of warfare’ should be thought of as a heuristic for the collection of sub-domains of reasons that together govern right action regarding war. To reflect this, the Authority View requires the following reformulation:

**The Revised Authority View of War:** A has authority over B regarding the domain of warfare, or *a sufficiently important sub-domain of warfare*, if, by obeying A’s commands, B achieves greater overall conformity with the reasons that govern the distribution of harm by means of war than by not obeying.

### **5.10 An Asymmetry**

A fine-grained approach to specifying domains of authority over warfare also shows that the moral power to impose authority-based constraints on participation in just wars is subject to a lower justificatory burden than the moral power to create authority-based justifications for participation in unjust wars.

This is because, in order for the command of an authority to provide a decisive reason for refraining from participating in a war, that authority need only possess legitimacy with regard to a particular sub-domain of *jus ad bellum* considerations. This is because a war’s satisfaction of each aspect of *jus ad bellum* requirement is necessary for participation to be justified. Hence, an authoritative command to refrain from fighting is sufficient to give its subject a decisive reason not to participate. By contrast, a command

to fight in an unjust war issued by an authority that is legitimate regarding a sub-domain of *jus ad bellum* is not sufficient to give subjects a decisive reason to participate. This is because the reason for participation provided by the command may still be defeated by reasons not to participate arising from the war's failure to satisfy additional *ad bellum* considerations that fall outside the authority's sphere of legitimacy. In order for the command to justify participation in an unjust war, that war must also satisfy these additional *ad bellum* requirements.

### **5.11 The War Crimes Objection Revisited**

In this section and the next I consider a different objection to the Authority View of War, which holds that it falls foul of the same war crimes objection that I pressed against non-reductive approaches to the ethics of war in Chapter 3. The objection held that non-reductive views cannot support the highly intuitive and stringent constraint on targeting non-liaible non-combatants in war. Since I argued that this objection warrants rejecting those approaches, its applicability to the Authority View of War would pose a serious problem.

As explained in Chapter 3, the source of the war crimes objection lies in the agent-centeredness of non-reductivist defences of the permissibility of killing in war. The problem is that if the relevant agent-centred considerations – such as one's associative duties to protect – are sufficient to justify the intentional killing of non-liaible combatants, then those very same considerations are also sufficient to justify the intentional killing of non-liaible non-combatants.

In Chapter 4, I argued that authority-based justifications for causing harm should be understood as a species of agent-relative justification. Furthermore, in this chapter I have argued that, under certain circumstances, individuals may possess authority-based justifications for participating in unjust wars, and thus for killing non-liaible combatants. Hence, given that both the Authority View and non-reductivist views offer agent-relative justifications for killing non-liaible combatants, one may



reasonably object that the war crimes objection applies equally to each. The Authority View of War inherits the very same problem that afflicts the non-reductivist. More specifically, the objection is that if, as I have claimed, the command of an authority may be sufficient to justify the intentional killing of non-liable combatants, then that authority's commands to target non-liable non-combatants will also justify doing so. Since we want to avoid this implication, it is argued, we must reject the Authority View of War.

## **5.12 Response**

My response to this objection has two parts. Firstly, I argue that despite their structural similarity, there is a relevant difference between the Authority View and non-reductivist views, which show that the Authority View does not justify the targeting of non-liable non-combatants to nearly the same extent as non-reductive views. Secondly, I concede that there may be possible cases in which the command of an authority does provide an agent with justification for targeting non-liable non-combatants, but argue that, within this narrow range of cases, this conclusion is defensible.

### *5.12.1 Distinct Forms of Agent-Relativity*

The objection is correct to highlight that both the Authority View and non-reductivist views offer justifications for the intentional killing of non-liable combatants in war that invoke agent-relative reasons. However, there is an crucial distinction to made between the two views, in terms of whether the agent-relative reasons that provide the justification are intrinsic reasons or instrumental reasons.

On a non-reductivist view, killing the non-liable is justified in terms of intrinsic or non-instrumental reasons. For example, on the associative-duty variant of non-reductivism, the constraint on killing a non-liable combatant is defeated by being overridden by a weightier agent-relative reason – the duty to protect one's associates from serious harm. This reason is grounded

in the intrinsic value of the relationship, and not in any further value this is realised by protecting one's associates.

By contrast, on the Authority View, valid commands create instrumental reasons for action. The subject's reason to obey the command arises from the fact that a policy of obedience is an optimal means for achieving conformity with the ultimate, non-instrumental reasons that apply to her. There need be no intrinsic value in obedience itself. This is true of valid commands that require the killing of non-labile combatants.

This distinction between these two types of agent-relative justification is important, since it provides the Authority View of War with resources for avoiding the war crimes objection that are not available to non-reductivists. The key point is that the instrumental structure of the Authority View allows for commands to be invalidated by their content, even when issued by a legitimate authority. These are commands that constitute what I earlier called 'clear mistakes'. To recapitulate, commands are clearly mistaken if the subject can form the judgement that the command fails to reflect right reason without having to go through the same process of reasoning that the authority went through in producing its directive. Under these conditions the subject can disobey the command without forsaking the rational benefits that the authority is able to provide. There is no instrumental value in obeying such commands.

The application of war crimes objection to the Authority View rests on the claim that an authority's having the moral power to obligate its subjects to intentional kill of non-labile combatants – such as by requiring participation in an unjust war – entails its having the moral power to obligate its subjects to intentionally kill non-combatants. The distinction between clearly and non-clearly mistaken commands shows that this claim is false.

As explained above, in order to possess the former moral power, an authority must successfully serve its subjects regarding a sufficiently important sub-domain of *jus ad bellum*. When this condition is met, subjects can be required to obey commands that are mistaken and require participation in unjust wars. This is the case when the subject cannot form

the judgement that the war is unjust without going through the same process of *ad bellum* reasoning that the authority went through in producing its command to fight. However, if the same authority issues a command to target non-combatants in war, its subjects *are* able to form the judgement that the command fails to reflect right reason without having to engage in any *ad bellum* reasoning. This is because the permissibility of targeting non-combatants is not dependent on *ad bellum* considerations in the same way that targeting combatants is. It is generally impermissible to target non-combatants regardless of whether one's war is justified or unjustified. Given this difference, it is sufficient for the subject to be able to form the judgement that a command requires the targeting of non-combatants for that command to be invalidated as clearly mistaken.<sup>240</sup>

This shows that the conditions of legitimacy sufficient for justifying the targeting of non-lieable combatants are not sufficient to justify the targeting of non-lieable non-combatants. The fact that an authority is legitimate over matters of *jus ad bellum* does not entail that its commands to target non-combatants create any reasons for action, let alone undefeated reasons. This is because an authority that is legitimate regarding the resort to war only has the moral power to require act-types that would be permissible in a justified war.<sup>241</sup> The war crimes objection can thus be rebutted. Commands to target non-lieable non-combatants will invariably be invalid, even when issued by an authority that possesses the moral power to require participation in unjust wars.

It is worth stressing that this response to the war crimes objection is not available to non-reductivists, since they justify the permission to kill non-lieable combatants non-instrumentally. There is no analogous mechanism for vitiating the agent-relative reason for killing the non-lieable when the crosshairs are shifted from combatants to non-combatants. For example, on the associative-duty view, the fact that a combatant can directly form the judgement that the individual they would have to kill in

---

<sup>240</sup> This point provides one explanation of why the convention of combatants wearing uniforms may have moral significance.

<sup>241</sup> This point is also made by David Estlund. 'On Following Orders in an Unjust War', pp.215-216.

order to protect their associate is a non-labile non-combatant, rather than a non-labile combatant, does not change the fact that their duty to protect is weightier than their duty not to kill the non-labile. An omniscient and infallible agent would remain justified in killing non-labile non-combatants, if necessary to discharge their duty to protect.

So, despite defending agent-relative justifications for killing non-labile non-combatants in war, the Authority View does not permit the targeting of non-labile non-combatants to nearly the same degree as non-reductivist views.

### *5.12.2 Accepting the Implication*

I have argued that in cases in which the Authority View justifies obeying commands to target non-labile combatants, it does not follow that the view will justify obeying commands to target non-combatants. However, while this suffices to rebut the war crimes objection, this does not mean that the Authority View of War will never have the implication that commands to target non-labile non-combatants are valid and give the subject undefeated reasons for action.

This follows from the general point, made in Section 4.10.2, that there are very few 'intrinsically' clearly mistaken commands, which are necessarily invalid across all possible domains in which an authority may serve its subjects. Instead, determining whether a command constitutes a clear mistake or not involves a comparison between the content of the command and the specific domain in which it is issued.

This point applies to my response to the war crimes objection, in which I argued that commands to target non-labile non-combatants will be invalidated as clear mistakes. Given the comparative nature of clear mistakes, this will only be true when the command is issued by an authority that is legitimate regarding a particular class of domains – such as domains of *jus ad bellum* considerations. When authorities are legitimate only regarding matters of *jus ad bellum*, its commands to target non-

combatants constitute clear mistakes because such commands couldn't reflect right reason *within that domain*.

However, as explained in Section 5.8, domains of authority regarding war are not restricted to *ad bellum* domains. Authorities may also potentially serve their subjects regarding matters of *jus in bello*. This opens up the possibility that there are domains of authority over warfare in which commands to target non-labile non-combatants will not constitute clear mistakes. To demonstrate, consider the following scenario:

***Sneaky Combatants***: Nation A is fighting a just war against Nation B. In order to gain a military advantage, Nation B's combatants disguise themselves as civilians, mix themselves with the civilian population, and launch attacks from within that population.

Furthermore, imagine that Nation A's military commanders are better placed with regard to distinguishing Nation B's combatants from Nation B's non-combatants than Nation A's combatants are. Nation A's combatants will better distribute harms in accordance with right reason by obeying their military leaders than by acting on their own assessment of right reason. Given these facts, under the Authority View of War, it is possible that Nation A's commanders acquire legitimacy over the *in bello* sub-domain of distinguishing between actual and merely apparent non-combatants. When an authority is legitimate over this particular domain, commands that are mistaken, and require the targeting of individuals who are in fact genuine non-combatants will not necessarily constitute clear mistakes. They may remain valid and provide decisive reasons for action.<sup>242</sup>

---

<sup>242</sup> Estlund also suggests, though very tentatively, that commands which require violating principles of *jus in bello* could be binding, "depending, of course, on many more details of the particular case." 'On Following Orders in an Unjust War', p.231. I hope to have provided an account of what these particular details might consist in. However, Estlund does seem ambivalent on this issue, since he later concludes that his view "leaves the possibility of an authoritative command that violates the rules of war, but only modestly." (p.233). The 'only modestly' caveat would seem to exclude cases of intentionally killing non-labile non-combatants, which are an extreme violation of *jus in bello*. My view is therefore stronger than Estlund's on this point. Furthermore, as examples of modest

This is because, in order to determine that such a command is mistaken, the subject *would* have to go through the same process of reasoning regarding distinguishing actual and apparent non-combatants that the authority went through in issuing its command. Avoiding acting against reason in these particular cases would require the subject to forsake the service that the authority is able to provide. The subject therefore has all-things-considered reason not to try to do so and instead to obey the authority, even in cases where doing so involve targeting non-labile non-combatants.

It must then be conceded that, under certain highly specific circumstances, subjects will have authority-based justifications for targeting non-labile non-combatants. However, I contend, in these cases this conclusion should be accepted. When the Authority View does yield the conclusion that an authority has the moral power to give its subjects decisive reason to target non-combatants, it does so only because, and to the extent that, the authority's having this power enables the subject to better overall avoid targeting non-labile individuals. This point helps further distinguish the Authority View from non-reductive views on the question of targeting non-combatants. Its instrumental structure not only means that it will permit the targeting of non-combatants to a far less extent than non-reductive views, it also means that, when it does so, it does so for the right kind of reasons.

### **5.13 Conclusion**

I have sought to show that applying the Authority View of Harm developed in the previous chapter to the specific case of war provides the resources

---

violations of *jus in bello* which may be justified in virtue of an authoritative command, Estlund uses the case of giving prisoners of war two meals a day rather than three in order to conserve supplies, and the case of keeping prisoners awake for twenty-four hours in order to gain information that could save lives (p.232). As Estlund describes them, both these cases seem intuitively justified on the basis of authority-independent reasons, and hence do not provide examples of cases in which the command of an authority serves to justify a genuine moral violation of *jus in bello*, even if they are violations of the laws of war.

for a qualified defence of the authority criterion. In particular, I have offered a defence of the claim that individuals who participate in wars on behalf of a legitimate war-waging authority are subject to a more extensive range of permissions to cause harm in war than those who fight in the absence of authority. One important implication of my argument is that, under certain conditions, individuals may be justified in obeying commands to participate in wars that are in fact unjust.

However, while I have aimed to rehabilitate an idea that runs throughout the just war tradition, my view remains revisionary in several respects. Firstly, to reiterate, the Authority View does not claim that those who fight on behalf of a legitimate authority are subject to *sui generis* moral principles. Rather, I have sought to show that additional permissions to cause harm can be justified by reference to the same principles that govern all other forms of interpersonal harming. It thus fits into an attractively unified view of morality.

Secondly, my account of legitimate war-making authority avoids the statist bias that afflicts standard interpretations of the authority condition. The conditions of legitimacy required under the Authority View are purely normative, making no reference to formal or legal properties such as statehood. Given this, there is no principled reason why non-state organisations could not qualify as legitimate authorities (though it may be easier for an organisation to satisfy the normative conditions if it also possesses the formal properties). Furthermore, according to the Authority View, it is not a necessary condition of justified war that it be waged by a legitimate authority. Hence, even non-state groups that fail to possess legitimate authority are not barred from waging just wars. All that follows from this failure is that their members cannot be justified in fighting *in virtue of possessing an authority-based justification*, not that they cannot be justified in fighting *tout court*.

Thirdly, and conversely, my account of war and authority does not claim that all states, or even all morally decent states, have legitimate authority regarding all aspects of warfare. Whether or not a state possesses the relevant authority is determined by whether it passes a fairly

demanding normative test, and there is no reason to hold a presumption that states will satisfy its conditions. This conclusion echoes the more general view, endorsed by the majority of theorists of authority, that states will be typically lack the degree of authority over their subjects that they claim to possess.<sup>243</sup>

Fourthly, on the view that I have advanced, what counts as having legitimate authority regarding warfare is a more complex and subtle matter than is commonly assumed. Rather than being all-or-nothing, authority over war comes in degrees. An authority may thus be legitimate with respect to certain war-related matters and not others.

Finally, while the Authority View defends a version of the authority criterion that is weaker in many respects than the criterion contained in mainstream just war theory, there is one important respect in which it is stronger: It denies the common view that commands to violate the rules of war – the prohibition on targeting non-combatants most saliently – necessarily exceed the moral limits of the duty to obey an authority.

---

<sup>243</sup>See, for example, Leslie Green, *The Authority of the State* (Oxford: Oxford University Press, 1988); A. John Simmons, 'Political Obligation and Authority' in Robert L. Simon (ed) *The Blackwell Guide to Social and Political Philosophy* (Oxford: Blackwell Publishers, 2002), 17-37; Joseph Raz, *The Authority of Law: Essays on Law and Morality* (Oxford: Oxford University Press, 1979), Ch.12.



## Conclusion

I conclude by providing a brief summary of the key ideas defended in the thesis. In the first chapter I argued that the authority criterion plays a much more expansive role within mainstream just war theory than is usually acknowledged. As standardly understood, the authority criterion functions solely as a condition of *jus ad bellum*, providing a necessary condition for justifying the resort to war. On this view, in order for a war to be justified, it must be initiated and waged by an entity that possesses a certain set of relevant properties. By contrast, on the revised interpretation that I argue for, the criterion also plays a crucial *in bello* role, capturing the idea – running throughout the just war tradition – that individuals who fight on behalf of certain kinds of entities enjoy additional permissions to cause harm in war that are unavailable to private actors.

With this interpretation in place, I then demonstrated how it provides a more accurate understanding of the challenge posed by irregular belligerency to mainstream just war theory. For if it is true that fighting on behalf of a certain kind of entity give rise to additional permissions to kill and injure, then just war theorists must provide an account of precisely which violence-using entities can and cannot generate these permissions, which both tracks our intuitions about particular cases and, more importantly, *explains why* acting on behalf of certain entities has this impressive normative effect. Hence, far from being an interesting but peripheral issue, the challenge of irregular belligerency strikes at the heart of the most central theoretical commitments of mainstream just war theory.

In the second chapter I began the task of assessing whether the authority criterion can be given a principled moral defence, by considering powerful objections to the criterion raised by a reductivist approach to the ethics of war. As I explained, the central commitment of a reductivist view is the claim that the moral permissions to cause harm that apply to individuals *qua* private agents – more specifically, liability and lesser-evil justifications – are both sufficient and necessary for justifying the resort to

war and the constituent acts of killing and injuring necessary for its prosecution. On this view, the *status* of a belligerent entity has no independent moral significance. The sufficiency claim entails rejecting the standard, restrictive role of the authority criterion. The necessity claim entails rejecting the non-standard, permissive role that I identified in Chapter 1.

While mainly expository in nature, this chapter aimed to illuminate this important and increasingly influential view by showing that, of all the revisionary aspects of a reductivist view, its rejection of the authority criterion may plausibly be considered its most radical break with the just war tradition.

While I accepted the reductivist critique of the standard, restrictive authority criterion, the following three chapters of the thesis investigated the prospects for rehabilitating the authority criterion, in its permissive role.

To this end, in Chapter 3 I considered a rival 'non-reductivist' approach to the ethics of war. Proponents of non-reductivism hold that it is a mistake to morally analyse war solely in terms of aggregated liability and lesser-evil justifications, since war differs from private violence in some morally relevant respect. Non-reductivists typically locate this difference in the fact that war involves the use of force by groups of individuals engaged in morally significant relationships with one another. A central claim of non-reductivism is that these relationships are independently capable of generating permissions for killing and injuring in war, thus denying reductivism and, potentially, providing the resources for a defence of the idea that individuals who fight on behalf of certain kinds of entities gain additional permissions to cause harm in war.

However, I argued that non-reductivism must be rejected, since it faces an unpalatable dilemma: If successful in showing that acting on behalf of a certain type of entity gives rise to additional permissions to cause harm in war, non-reductivists must also jettison the most intuitive *restrictions* on conduct in war – most saliently, the prohibition on intentionally killing morally innocent non-combatants. As I explained, this problem arises

because non-reductivists endorse an agent-centred, rather than target-centre, justification for killing non-liable persons in war. Since, on this view, it is facts about those *causing* harm, rather than *suffering* it, which justifies killing the non-liable, the view lacks the resources to discriminate between the permissibility of killing non-liable combatants and non-liable non-combatants.

In defending this conclusion I assessed various alternative strategies for grounding the principle of discrimination in war that non-reductivists might invoke in order to avoid the dilemma and argued that they all fail. The factors identified are either morally irrelevant, insufficiently moral weighty, or fail to track the required distinction between combatants and non-combatants.

My arguments in favour of rejecting non-reductivism thus provided strong indirect support for reductivism – by eliminating its main rival – and, by extension, cast further doubt on the defensibility of the authority criterion.

In the fourth chapter I provided the groundwork for an alternative defence of the permissive authority criterion, by exploring the connections between two central topics in moral and political philosophy: the morality of interpersonal harming on the one hand, and the moral legitimacy of authority on the other. Whereas those working on the ethics of harm are concerned with the conditions under which individuals are justified in causing harm to others, the theorist of authority is interested in the conditions under which individuals are morally required to obey the commands of another.

I considered two important questions that arise at the intersection of these two topics. The first concerned the extent to which individuals may be morally justified in obeying commands that require causing serious physical harm to others. The second concerned the permissibility of imposing defensive harm on those who threaten unjust harm, but who do so in conformity with a legitimate command to do so.

In response to the first question, I argued that, under certain conditions, the command of an authority can provide agents with an independent moral

justification for causing serious physical harm to others. Controversially, I argued that this may be true even in cases where the harm would be straightforwardly unjustified in the absence of the command. My argument for this position drew on a service-based account of authority, according to which authorities possess the moral power to place their subjects under obligations to obey when, and to the extent that, the subjects better conform with reasons that apply to them by obeying the authority, compared to not obeying. This view thus offers a broadly instrumental justification of the duty to obey.

I argued that the very same reasoning holds in the case of the reasons that govern the distribution of harm. If individuals will overall better distribute harms in accordance with right reason by obeying an authority, compared to not obeying, then obedience is morally required (and therefore justified). Importantly, this includes cases in which the authority issues commands that are mistaken and require distributing harm in ways not supported by the balance of authority-independent reasons. In such cases, I argue, the subject of the command may still have an undefeated reason – an instrumental reason – to obey the command.

I termed this argument the ‘Authority View of Harm.’ If successful, it yields two important results. Firstly, it shows that the range of possible justifications for harming is broader than standardly assumed within the literature on the morality of harm. In addition to standard justifications for harming – in which the right not to be harmed either fails to obtain or is overridden – the Authority View posits the existence of what I termed *authority-based* justifications for harm. Secondly, it reveals that certain common and intuitive views about the moral limits of the duty to obey are mistaken. While it is true that commands are limited both in terms of their validity and weight, it is not true that commands which require causing serious unjust harm to others necessarily exceed those limits.

I then turned to the second question, concerning the permissibility of imposing defensive harm on agents who possess an authority-based justification for causing unjust harm. In response, I argued that, in such cases, the fact that the threatener possesses a justification does not, in

itself, increase the justificatory burden for imposing defensive harm on them. The reason for this, I contended, is that authority-based justifications are fully agent-relative, grounded in reasons for action that apply only to the subject of the command.

Again, this conclusion yields two important results. Firstly, an on-going debate within the morality of harming literature concerns the existence and scope of agent-relative justifications. Typically, these justifications are defended by appeal to considerations of permissible partiality or self-preference. If my arguments are correct, there may be an additional species of agent-relative justification for harming that is not grounded in considerations of partiality and, as such, may not be susceptible to same objections that are typically levelled against the notion of agent-relative justifications. Secondly, and relatedly, the authority-based argument for agent-relative justifications enables a novel defence of the possibility of 'conflicts of permissibility', in which two agents may each be justified in bringing about opposing and incompatible distributions of harm.

Finally, while I denied that an agent's possession of an authority-based justification itself counts against imposing defensive harm on them, this does not mean that the commands of authorities cannot raise the justificatory burden for acts of defensive harm. In the final section of the chapter I argued that the same argument for the existence of authority-based justifications also supports the notion of an *authority-based constraint*, where the command of an authority places the subject under an undefeated obligation to refrain from causing harm that would otherwise be justified.

The Authority View of Harm provides a fairly general account of how judgements about the morality of causing harm are sensitive to considerations of authority. In the fifth chapter I applied the Authority View to the specific case of warfare and assessed its implications for the ethics of war.

I argued that the twin notions of authority-based justifications and constraints provide the resources for a qualified defence of both the permissive authority criterion and a reformulated version of the restrictive

authority criterion. When it is the case that individuals will better distribute harms by means of war by obeying an authority, compared to not obeying, then that authority acquires the moral power to give those individuals undefeated reasons to either cause or refrain from causing harm in war. The most salient (but by no means only) implication of this argument is that individuals may be justified in fighting in unjust wars, or unjustified in fighting in just wars, in virtue of being commanded by a legitimate war-waging authority.

While I have argued for a rehabilitation of the authority criterion against its reductivist critics, I have stressed that the view I have defended does not endorse the idea – endorsed, either explicitly or implicitly, throughout the just war tradition – that war (or political action more generally) is some sense morally special, so that ordinary moral principles are either supplanted or overridden.

Much of the scepticism regarding for the authority criterion follows from the fact that the criterion seems to jar so heavily with a commitment to the liberal and moderately cosmopolitan view that individuals are the ultimate unit of moral concern and possess basic rights not to be harmed, whose stringency not affected by group membership. In response, I have sought to show that there is a plausible view of why individuals may be morally required to obey authorities that is both compatible with these commitments and yields some interesting and controversial results about the morality of war.

## Bibliography

Aboodi, R., Borer A. and Enoch D. 'Deontology, Individualism, and Uncertainty: A Reply to Jackson and Smith', *Journal of Philosophy* 105, No.5 (2008), 259-272

Anscombe, G.E.M. 'War and Murder', in *The Collected Papers of G.E.M. Anscombe. Volume Three: Ethics Religion and Politics* (Oxford: Blackwell, 1981), 51-61

Augustine, *The City of God Against the Pagans*, R.W Dyson (ed) (Cambridge: Cambridge University Press, 1988)

Altman, A. and Wellman, C.H. 'From Humanitarian Intervention to Assassination: Human Rights and Political Violence', *Ethics* 118, No.2 (2008), 228-257

Arneson, R.J. 'Just Warfare Theory and Non-Combatant Immunity', *Cornell International Law Journal* 39 (2006), 663-688

Bass, G. 'Jus Post Bellum', *Philosophy and Public Affairs* 32, No. 4 (2004), 384-412

Bazargan, S. 'The Permissibility of Aiding and Abetting Unjust Wars', *Journal of Moral Philosophy* 8, No.4 (2011), 513-529

Bazargan, S. 'Varieties of Contingent Pacifism' in Helen Frowe and Gerald Lang (eds), *How We Fight* (Oxford: Oxford University Press, 2014), 1-17

Bazargan, S. 'Complicitous Liability in War', *Philosophical Studies* 165 (2013), 177-195

Begby, E. Reichberg G. and Syse, S. 'The Ethics of War. Part II: Contemporary Authors and Issues', *Philosophy Compass* 7, No.5 (2012), 328-347

Berdal, M. 'The 'New Wars' Thesis Revisited' in Huw Strachan and Sibylle Scheipers (eds), *The Changing Character of War* (Oxford: Oxford University Press, 2011), 109-133

Brandt, R.B. 'Utilitarianism and the Rules of War', *Philosophy and Public Affairs* 1. No.2 (1972), 145-165

Brooks, R.E. 'War Everywhere: Rights, National Security Law, and the Law of Armed Conflict in an Age of Terror', *University of Pennsylvania Law Review* 153, No.2 (2004), 675-761

Brownlee, K. 'Responsibilities of Criminal Justice Officials', *Journal of Applied Philosophy* 27, No.2 (2010), 123-139

Buchanan, A. 'The Internal Legitimacy of Humanitarian Intervention', *Journal of Political Philosophy* 7, No.1 (1999), 71-87

Buchanan A. 'Institutionalizing the Just War', *Philosophy and Public Affairs* 34, No.1 (2006), 2-38

Chan, D.K. *Beyond Just War: A Virtue Ethics Approach* (Basingstoke: Palgrave MacMillan, 2012)

Christopher, P. *The Ethics of War and Peace: An Introduction to Legal and Moral Issues*, 3<sup>rd</sup> Edition (New Jersey: Prentice Hall, 2004)

Coady, C.A.J. 'The Status of Combatants' in David Rodin and Henry Shue (eds), *Just and Unjust Warriors: The Moral and Legal Status of Soldiers* (Oxford: Oxford University Press, 2008), 153-175

Coates, A. *The Ethics of War*, (Manchester: Manchester University Press, 1997)

Dallaire, R. *Shake Hands With the Devil: The Failure of Humanity in Rwanda* (Toronto: Random House, 2003)

Davis, N.A. 'Abortion and Self-Defense', *Philosophy and Public Affairs* 13, No.3 (1984), 175-207

Dorsey, D. 'Consequentialism, Cognitive Limitations and Moral Theory', in Mark Timmons (ed), *Oxford Studies in Normative Ethics: Volume 3* (Oxford: Oxford University Press, 2013), 179-202

Enoch, D. 'Authority and Reason-Giving', *Philosophy and Phenomenological Research* (forthcoming)

Estlund, D. 'On Following Orders in an Unjust War', *Journal of Political Philosophy* 15, No.2 (2007), 213-234



Estlund, D. *Democratic Authority* (Princeton: Princeton University Press, 2008)

Evans, M. 'Moral Theory and the Idea of a Just War', in Mark Evans (ed), *Just War Theory: A Reappraisal*, (Edinburgh: Edinburgh University Press, 2005)

Fabre, C. 'Cosmopolitanism, Just War Theory and Legitimate Authority', *Ethics and International Affairs* 84, No. 5 (2008), 963-976

Fabre, C. 'Guns, Food and Liability to Attack in War', *Ethics* 120, No.1 (2009), 36-63

Fabre, C. 'Permissible Rescue Killings', *Proceedings of the Aristotelian Society* 109 (2009), 149-164

Fabre, C. *Cosmopolitan War* (Oxford: Oxford University Press, 2012)

Fabre, C. 'Internecine War Killings', *Utilitas* 24, No,2 (2012), 214-236

Fabre, C. 'Cosmopolitanism and Wars of Self Defence' in Cecile Fabre and Seth Lazar (eds), *The Morality of Defensive War* (Oxford: Oxford University Press, 2014), 90-114

Fabre, C. 'War Exit', *Ethics* (forthcoming)

Ferzan, K. 'Justifying Self-Defense', *Law and Philosophy* 24, No.6 (2005), 711-749

Ferzan, K. 'Self-Defense and the State', *Ohio State Journal of Criminal Law* 5, No.2 (2008), 449-504

Fiala, A. *Public War, Private Conscience: The Ethics of Political Violence*, (London: Continuum, 2010)

Fiala, A. *Against Religion, Wars, and States: The Case for Enlightened Atheism, Just War Pacifism, and Liberal Democratic Anarchism* (Plymouth: Rowman and Littlefield, 2013)

Finlay, C. 'Legitimacy and Non-State Political Violence', *Journal of Political Philosophy*, 18, No.3 (2010), 287-312

Firth, J.M. and Quong, J. 'Necessity, Moral Liability and Defensive Harm', *Law and Philosophy* 31, No. 6 (2012), 673-701

Fletcher, G. 'Defensive Force as an Act of Rescue', *Social Philosophy and Policy* 7, No. 2 (1990), 170-179

Fletcher, G. *Romantics at War: Glory and Guilt in the Age of Terrorism*, (Princeton: Princeton University Press, 2002)

Foot, P. *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford: Oxford University Press, 1978)

Fotion, N. 'Two Theories of Just War', *Philosophia* 34 (2006), 53-64

Fotion, N. *War and Ethics: A New Just War Theory*, (London: Continuum, 2007)

Frowe, H. *The Ethics of War and Peace* (Abingdon: Routledge, 2011)

Frowe, H. 'Self-Defence and the Principle of Non-Combatant Immunity', *Journal of Moral Philosophy* 8, No.4 (2011), 530-546

Frowe, H. *Defensive Killing: An Essay on War and Self-Defence* (Oxford: Oxford University Press, forthcoming)

Gardner, J. *Offences and Defences* (Oxford: Oxford University Press, 2007)

Gardner, J. 'Justification Under Authority', *Canadian Journal of Law and Jurisprudence* 23, No.1 (2010), 73-98

Gardner, J. 'Criminals in Uniform', in Anthony Duff, Lindsey Farmer, Sandra Marshall, Massimo Renzo and Victor Tadros (eds), *The Constitution of Criminal Law* (Oxford, Oxford University Press, 2012), 97-118

Glover, J. *Causing Death and Saving Lives* (London: Penguin, 1977)

Goodin, R.E. *What's Wrong With Terrorism?* (Malden, MA: Polity Press, 2006)

Green, L. *The Authority of the State* (Oxford: Oxford University Press, 1988)

Haque, A.A. 'Rights and Liabilities at War', in Paul Robinson, Stephen Garvey and Kimberly Kessler Ferzan (eds), *Criminal Law Conversations* (Oxford: Oxford university Press, 2009), 395-396

Haque, A.A. 'Law and Morality at War', *Criminal Law and Philosophy* 8 (2014), 79-97

Hare, R.M. 'Rules of War and Moral Reasoning', *Philosophy and Public Affairs* 1. No.2 (1972), 166-181

Heinze, E.A. and Steele, B.J. 'Introduction: Non-State Actors and the Just War Tradition', in Eric A. Heinze and Brent J. Steele (eds), *Ethics, Authority and War* (Basingstoke: Palgrave MacMillan, 2009), 1-20

Holliday, I. 'When is a Cause Just?', *Review of International Studies* 28, No.3 (2002), 557-575

Honore, T. 'The Right to Rebel', *Oxford Journal of Legal Studies* 8, No.1 (1988), 34-54

Hooker, B. *Ideal Code, Real World* (Oxford: Oxford University Press, 2000)

Hosein, A. 'Are Justified Aggressors a Threat to the Rights Theory of Self-Defense?', in Helen Frowe and Gerald Lang (eds) *How We Fight*, (Oxford: Oxford University Press, 2014), 87-103

Huemer, T. *The Problem of Political Authority: An Examination of the Duty to Obey and the Right to Coerce* (Basingstoke: Palgrave Macmillan, 2013),

Hurka, T. 'Liability and Just Cause', *Ethics and International Affairs* 20 (2007), 199-218

Johnson, J.T. 'Aquinas and Luther on Sovereign Authority', *Journal of Religious Ethics*, 31, No, 1 (2003), 3-20

Kaldor, M. *New and Old Wars: Organised Violence in a Global Era*, 2<sup>nd</sup> Edition (Cambridge: Polity Press, 2007).

Kamm, F. 'Rights' in Jules Coleman and Scott Shapiro (eds), *The Oxford Handbook of Jurisprudence and Philosophy of Law* (Oxford: Oxford University Press, 2002), 476-513

Kamm, F. *Intricate Ethics: Rights, Responsibilities and Permissible Harm* (Oxford: Oxford University Press, 2007)

Kamm, F. *The Moral Target*, (Oxford: Oxford University Press, 2012)

Kant, I. 'Perpetual Peace: A Philosophical Sketch', in H.S Reiss (ed) *Kant: Political Writings*, 2<sup>nd</sup> Edition (Cambridge: Cambridge University Press, 1991), 93-139

Knowles, D. 'The Domain of Authority', *Philosophy* 82, No.1 (2007), 23-43

Kutz, C. 'The Difference Uniforms Make: Collective Violence in Criminal Law and War', *Philosophy and Public Affairs* 33, No.2 (2005), 148-180

Lang Jr, A.F. ,O'Driscoll. C., and Williams, J. 'Introduction' in *Just War: Authority, Tradition and Practice*, Anthony F. Lang Jr, Cian O'Driscoll and John Williams (eds), (Washington DC: Georgetown University Press, 2013), 1-19

Langan, J. 'The Elements of Augustine's Just War Theory', *Journal of Religious Ethics* 12, No. 1 (1984), 19-38

Lazar, S. 'The Responsibility Dilemma for Killing in War: A Review Essay', *Philosophy and Public Affairs* 38, No,2 (2010), 180-213

Lazar, S. 'Necessity in Self-Defense and War', *Philosophy and Public Affairs* 40, No.1 (2012), 3-44

Lazar, S. 'Associative Duties and the Ethics of Killing in War', *Journal of Practical Ethics* 1, No.1 (2013), 6-51

Lazar, S. 'National Defence, Self-defence, and the Problem of Political Aggression', in *The Morality of Defensive War*, Seth Lazar and Cecile Fabre (eds), (Oxford: Oxford University Press, 2014), 11-39

Lichtenberg, J. 'How to Judge Soldiers Whose Cause is Unjust' in Rodin and Shue (eds), *Just and Unjust Warriors: The Moral and Legal Status of Soldiers* (Oxford: Oxford University Press, 2008), 112-129

Locke, J. 'Two Treatises of Government', Peter Laslett (ed), (Cambridge: Cambridge University Press, 2002)

Luban, D. 'Just War and Human Rights', *Philosophy and Public Affairs* 9, No.2 (1980), 160-181

Luban, D. 'War Crimes: The Laws of Hell' in Larry May (ed), *War: Essays in Political Philosophy* (Cambridge: Cambridge University Press, 2008), 266-288

Mapel, 'Moral Liability to Defensive Harm and Symmetrical Self-Defense', *Journal of Political Philosophy* 18, No.2 (2010), 198-217

May, L. *War Crimes and Just War* (Cambridge: Cambridge University Press, 2007)

May, L. *After War Ends: A Philosophical Perspective*, (Cambridge: Cambridge University Press, 2012)

McMahan, J. 'Killing, Letting Die, and Withdrawing Aid', *Ethics* 103, No.2 (1993), 250-279

McMahan, J. 'Innocence, Self-Defense and Killing in War', *Journal of Political Philosophy* 2, No. 3 (1994), 193-221

McMahan, J. 'Self-Defense and the Problem of the Innocent Attacker', *Ethics* 104, No.2 (1994), 252-290

McMahan, J. 'The Ethics of Killing War', *Ethics* 114, No.4 (2004), 693-733

McMahan, J. 'War as Self-Defense', *Ethics and International Affairs* 18, No.1 (2004), 13-18

McMahan, M. 'The Basis of Moral Liability to Defensive Killing', *Philosophical Issues* 15 (2005), 386-405

McMahan, J. 'Just Cause for War', *Ethics and International Affairs* 19, No.3 (2005), 1-21

McMahan, J. 'Just War' in Robert E. Goodin, Philip Pettit and Thomas Pogge (eds), *A Companion to Political Philosophy* 2<sup>nd</sup> Edition, (Oxford: Blackwell, 2007), 669-677

McMahan, J. 'Collectivist Defences of the Moral Equality of Combatants', *Journal of Military Ethics* 6, No.1 (2007), 50-59

McMahan, J. 'Précis of *The Morality and Law of War*', *Israel Law Review* 40, No. 3 (2007), 310-323

McMahan, J. 'The Morality of War and the Law of War', in Rodin and Shue (eds), *Just and Unjust Warriors: The Moral and Legal Status of Soldiers* (Oxford: Oxford University Press, 2008), 19-43

McMahan J. 'Debate: Justification and Liability in War', *Journal of Political Philosophy* 16, No.2 (2008), 227-244

McMahan, J. *Killing in War*, (Oxford: Oxford University Press, 2009)

McMahan, J. 'Intention, Permissibility, Terrorism and War', *Philosophical Perspectives* 23 (2009), 345-372

McMahan, J. 'The Morality of Military Occupation', *Loyola International and Comparative Law Review* 31 (2009), 7-31

McMahan, J. 'Responsibility, Permissibility and Vicarious Agency', *Philosophy and Phenomenological Research* 80, No.3 (2010), 673-680

McMahan, J. 'Who is Morally Liable to be Killed in War', *Analysis* 71, No.3 (2011), 544-559

McMahan, J. 'War' in David Estlund (ed), *The Oxford Handbook of Political Philosophy*, (Oxford: Oxford University Press, 2012), 298-318

McMahan, J. 'The Prevention of Unjust Wars' in Yitzhak Benbaji and Naomi Sussmen (eds), *Reading Walzer* (Abingdon: Routledge, 2013), 233-255

McMahan, J. 'What Rights May We Defend By Means of War' in Cecile Fabre and Seth Lazar (eds), *The Morality of Defensive War* (Oxford: Oxford university Press, 2014), 115-158

McMahan J. 'Self-Defense Against Justified Threateners', in Helen Frowe and Gerald Lang (eds), *How We Fight* (Oxford: Oxford University Press, 2014), 104-137

McPherson, L. 'Innocence and Responsibility in War', *Canadian Journal of Philosophy* 34, No.4 (2004), 485-506

McPherson, L. 'Is Terrorism Distinctively Wrong?', *Ethics* 117, No.3 (2007), 524-546

Meisels, T. 'In Defence of the Defenceless: The Morality of the Laws of War', *Political Studies* 60, No 4 (2012), 919-935

Moellendorf, D. *Cosmopolitan Justice* (New York: Basic Books, 2001)

Moellendorf, D. 'Jus ex Bello', *Journal of Political Philosophy* 16, No. 2 (2008), 123-136

Moseley, A. 'The Philosophy of War', *Internet Encyclopedia of Philosophy*, available at <http://www.iep.utm.edu/war/>

Neff, S.C. *War and the Law of Nations: A General History*, (Cambridge: Cambridge University Press, 2005)

Norman, R. *Ethics, Killing and War* (Cambridge: Cambridge University Press, 1995)

O'Donovan, O. *The Just War Revisited* (Cambridge: Cambridge University Press, 2003)

Orend, B. 'Jus Post Bellum', *Journal of Social Philosophy* 31, No.1 (2000), 117-137

Orend, B. *Michael Walzer on War and Justice*, (Cardiff: University of Wales Press, 2000)

Orend, B. 'War' in *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), Edward N. Zalta (ed), available at: <http://plato.stanford.edu/archives/fall2008/entries/war/>

Otsuka, M. 'Killing the Innocent in Self-Defense', *Philosophy and Public Affairs* 23, No.1 (1994), 74-94

Overland, G. 'Killing Soldiers', *Ethics and International Affairs* 20, No.4 (2006), 455-475

Palmer-Fernandez, G. 'Cosmopolitan Revisions to Just War Theory', in Michael Boylan (ed), *The Morality and Global Justice Reader* (Boulder: Westview Press, 2011), 325-342

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

Parfit, D. *On What Matters*, (Oxford: Oxford University Press, 2011)

Parsons, G. 'Public War and the Moral Equality of Combatants', *Journal of Military Ethics* 11, No.4 (2012), 299-317

Pattison, J. 'When Is it Right to Fight? Just War Theory and the Individual-Centric Approach', *Ethical Theory and Moral Practice* 16, No.1 (2013), 34-54

Perry, S. 'Political Authority and Political Obligation', in Leslie Green and Brian Leiter (eds), *Oxford Studies in Philosophy of Law: Volume 2* (Oxford: Oxford University Press, 2013), 1-74

Pinker, S. *The Better Angels of Our Nature: A History of Violence and Humanity* (London: Penguin Books, 2011)

Quinlan, M. and Guthrie, C. *Just War* (New York: Bloomsbury, 2007)

Quinn, W. 'Actions, Intentions and Consequences: The Doctrine of Doing and Allowing', *The Philosophical Review* 98, No.3 (1989), 287-312

Quinn, 'Actions, Intentions and Consequences: The Doctrine of Double Effect', *Philosophy and Public Affairs* 18, No.4 (1989), 334-251

Quong, J. 'Killing in Self-Defense', *Ethics* 119, No.2 (2009), 334-351

Quong, J. *Liberalism Without Perfection* (Oxford: Oxford University Press, 2011)

Quong, J. 'Liability to Defensive Harm', *Philosophy and Public Affairs* 40, No.1 (2012), 45-77

Rawls, J. *A Theory of Justice* (Cambridge: Harvard University Press, 1971)

Raz, *The Authority of Law: Essays on Law and Morality* (Oxford: Oxford University Press, 1979)

Raz, J. *The Morality of Freedom* (Oxford: Clarendon Press, 1986)

Raz, 'Facing Up: A Reply', *Southern California Law Review* 62 (1989), 1153-1235

Raz, J. *Between Authority and Interpretation*, (Oxford: Oxford University Press, 2009)

Raz, J. 'On Respect, Authority and Neutrality: A Response', *Ethics* 120, No.2 (2010), 279-301



Regan, R.J. *Just War: Principles and Cases*, (Washington DC: Catholic University of America Press, 1996)

Reichberg, G., Syse H. and Begby, E. (eds), *The Ethics of War: Classic and Contemporary Readings* (Oxford: Blackwell, 2006)

Reichberg, G. 'Just War and Regular War: Competing Paradigms' in David Rodin and Henry Shue (eds), *Just and Unjust Warriors: The Moral and Legal Status of Soldiers* (Oxford: Oxford University Press, 2008), 193-214

Reichberg, G. 'The Moral Equality of Combatants – A Doctrine in Classical Just War Theory? A Response to Graham Parsons', *Journal of Military Ethics* 12, No.2 (2013), 181-194

Ridge, M. 'Reasons for Action: Agent-Neutral Vs. Agent-Relative', in Edward N. Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition). Available at <http://plato.stanford.edu/entries/reasons-agent/>

Rodin, D. *War and Self-Defense* (Oxford: Oxford University Press, 2002)

Rodin D. and Shue, H. 'Introduction' in David Rodin and Henry Shue (eds), *Just and Unjust Warriors: The Moral and Legal Status of Soldiers* (Oxford: Oxford University Press, 2008), 1-18

Rodin, D. 'The Moral Inequality of Combatants: Why *jus in bello* Asymmetry is Half-Right' in David Rodin and Henry Shue (eds), *Just and Unjust Warriors: The Moral and Legal Status of Soldiers* (Oxford: Oxford University Press, 2008), 44-68

Rodin, D. 'The Morality and Law of War' in Huw Strachan and Sibylle Scheipers (eds), *The Changing Character of War* (Oxford: Oxford University Press, 2011), 446-463

Rodin, D. 'Justifying Harm', *Ethics* 122, No.1 (2011), 74-110.

Rodin, D. 'The Myth of National Defence', in Cecile Fabre and Seth Lazar (eds), *The Morality of Defensive War* (Oxford: Oxford University Press, 2014), 69-89

Russell, F. *The Just War in the Middle Ages* (Cambridge: Cambridge University Press, 1977)

Scheffler, S. *The Rejection of Consequentialism* (Oxford: Clarendon, 1982)

- Scheffler, S. 'Doing and Allowing', *Ethics* 114, No.2 (2004), 215-239
- Schwenkenbecher, A. 'Rethinking Legitimate Authority' in Fritz Alhoff, Nicholas G. Evans, and Adam Henschke (eds), *Routledge Handbook of Ethics and War* (Abingdon: Routledge), 161-170
- Shapiro, S. 'Authority' in Jules Coleman and Scott Shapiro (eds), *The Oxford Handbook of Jurisprudence and Philosophy of Law* (Oxford: Oxford University Press, 2002), 382-439
- Shaw, W. 'Utilitarianism and Recourse to War', *Utilitas* 23, No. 4 (2011), 380-401
- Shiffrin, S. 'Immoral, Conflicting and Redundant Promises' in R. Jay Wallace, Rahul Kumar, and Samuel Freeman (eds), *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon* (Oxford: Oxford University Press, 2011), 155-178
- Shue, H. 'Torture', *Philosophy and Public Affairs* 7, No.2 (1978), 124-143
- Shue, H. 'Do We Need a 'Morality of War' in David Rodin and Henry Shue (eds), *Just and Unjust Warriors: The Moral and Legal Status of Soldiers*, (Oxford: Oxford University Press, 2008), 87-111
- Simmons, A.J. 'External Justifications and Institutional Roles', *Journal of Philosophy* 93, No.1 (1996), 28-36
- Simmons, A.J. 'Philosophical Anarchism' in John T. Sanders and Jan Narveson (eds), *For and Against the State: New Philosophical Readings* (Boston: Rowman and Littlefield, 1996), 19-40
- Simmons, A.J. 'Political Obligation and Authority' in Robert L. Simon (ed) *The Blackwell Guide to Social and Political Philosophy* (Oxford: Blackwell Publishers, 2002), 17-37
- Smith, M.N. 'Political Obligation and the Self', *Philosophy and Phenomenological Research* 86, No.2 (2013), 347-375
- Sparrow, R. 'Hands Up Who Wants to Die: Primoratz on Responsibility and Civilian Immunity in War', *Ethical Theory and Moral Practice* 8, No.3 (2005), 299-319

Steinhoff, U. *The Ethics of War and Terrorism* (Oxford: Oxford University Press, 2007)

Steinhoff, 'Jeff McMahan on the Moral Equality of Combatants', *Journal of Political Philosophy* 16, No. 2 (2008), 220-226

Steinhoff, U. 'What is War? And Can a Lone Individual Wage One?', *International Journal of Applied Philosophy* 29, No.1 (2009), 133-150

Steinhoff, U. 'Rights, Liability and the Moral Equality of Combatants', *Journal of Ethics* 16, No.4 (2012), 339-366

Stilz, A. 'Authority, Self-Determination and Community in Cosmopolitan War', *Law and Philosophy*, (forthcoming)

Strawser, B.J. 'Walking the Tightrope of Just War', *Analysis* 71, No.3 (2011), 58-66

Swift, L.J. 'St. Ambrose on Violence and War', *Transactions and Proceedings of the American Philological Society* 101 (1970), 533-543

Syse, H. and Ingierd, H. 'What Constitutes a Legitimate Authority?', *Social Alternatives* 24, No.3 (2005), 11-16

Tadros, V. *The Ends of Harm: The Moral Foundations of Criminal Law*, (Oxford; Oxford University Press, 2011)

Tadros, V. 'Duty and Liability', *Utilitas* 24, No.2 (2012), 259-277

Taurek, J. 'Should the Numbers Count?', *Philosophy and Public Affairs* 6, No.4 (1977), 293-316

Temkin, L. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford: Oxford University Press, 2012)

Thompson, J. 'Terrorism and the Right to Wage War' in C.A.J. Coady and Michael O'Keefe (eds), *Terrorism and Justice: Moral Argument in a Threatened World* (Melbourne: Melbourne University Press, 2002), 87-96

Thompson, J. 'Terrorism, Morality and Right Authority' in Georg Meggle (ed), *Ethics of Terrorism and Counter-Terrorism* (Frankfurt: Ontos, 2005), 51-60

Thompson, M. 'Aquinas and Locke on Self-Defense', *University of Pittsburgh Law Review* 57 (1996), 677-684.

Thomson, J.J. 'Self-Defense', *Philosophy and Public Affairs* 20, No.4 (1991), 283-310

Thorburn, M. 'Justifications, Powers, and Authority', *Yale Law Journal* 117 (2008), 1070-1130

Toner, C. 'The Logical Structure of Just War Theory', *The Journal of Ethics* 14, No.2 (2010), 81-102

Tucker, A. 'The Limits of Razian Authority', *Res Publica* 18, No.3 (2012), 225-240

Twiss, S.B. and Chan, J. 'The Classical Confucian Position on the Legitimate Use of Military Force', *Journal of Religious Ethics* 40, No. 3 (2007), 447-472

Viehoff, D. 'Debate: Procedure and Outcome in the Justification of Authority', *Journal of Political Philosophy* 19, No.2 (2011), 248-259

Viehoff, D. *The Authority of Democracy* (manuscript)

Waldron, J. 'Self-Defense: Agent-Neutral and Agent-Relative Accounts', *California Law Review* 88 (2000), 711-750

Walzer, M. *Just and Unjust Wars: A Moral Argument With Historical Illustrations*, 4<sup>th</sup> Edition (New York: Basic Books, 2006)

Walzer, M. 'Response to McMahan's Paper', *Philosophia* 34, No.1 (2006), 43-45

Wilson, H. *The Use of Force by National Liberation Movements*, (Oxford: Oxford University Press, 1988)

Wolff, R.P. *In Defense of Anarchism*, (New York: Harper, 1970)

Zohar, N. 'Collective War and Individualist Ethics: Against the Conscriptio of Self-Defense', *Political Theory* 21, No.4 (1993), 606-622