

Instrumental Authority and Its Challenges: The Case of the Laws of War*

Jonathan Parry (Birmingham) and Daniel Viehoff (NYU)

Note: This is prepublication draft of a paper that is forthcoming in *Ethics* as part of a symposium on Adil Haque's *Law and Morality at War* (OUP, 2017). Please cite published version.

Abstract: *Law and Morality at War* offers a broadly instrumentalist defence of the authority of the laws of war: these laws serve combatants by helping them come closer to doing what they have independent moral reason to do. We argue that this form of justification (invoked by many legal and political theorists) sets too low a bar. An authority's directives are not binding, on instrumental grounds, if the subject could, within certain limits, adopt an alternative, and superior, means of conforming to morality's demands. It emerges that Haque's argument fails to vindicate the law's authority over all (or even most) combatants.

1. Introduction

At the heart of much contemporary writing on the ethics of war is a puzzle about the relation between the laws of war (or 'law of armed conflict', LOAC for short) and the moral rights and duties of combatants and non-combatants. The law famously treats those fighting just and unjust wars symmetrically: both sides are under the same legal duty not to target non-combatants, and neither is under a general legal duty not to target enemy combatants. By contrast, familiar moral principles governing killing and harming appear to attribute very different rights and duties to aggressors and defenders, and do not support a strict prohibition on targeting non-combatants.

For example, a police officer who uses lethal force to prevent an armed robbery does not thereby become liable to be defensively killed by the robbers, despite threatening their lives. By parity of reasoning, many just war theorists suggest, it is hard to see how combatants who fight in justified

*Versions of this paper were presented at a workshop on Adil Haque's *Law and Morality at War*, hosted by the Stockholm Centre for the Ethics of War and Peace, and at the Political Theory Research Group at The University of Birmingham. Thanks to the audiences at those events; to Helen Frowe and Adil Haque for their generous feedback, and for making this symposium possible; and to Juri Viehoff and two associate editors of this journal for their helpful comments.

wars could become liable to defensive force simply by resisting unjust aggressors.¹ So they retain their moral right not to be killed, and unjust combatants commit a grave moral wrong by killing them. Yet no corresponding duty on the unjust soldier is embodied in the laws of war.

Common-sense morality also accepts that non-threatening individuals can be morally liable to defensive killing. For example, if five friends hire a hitman to kill an innocent person, then it seems plausible that each may permissibly be killed if necessary to prevent the hit, even though none poses a direct threat to the victim. (If the only way the victim can avoid the hitman's bullets is by lethally using one of the five as a shield, then she would be morally permitted to do so.) By analogy, many contemporary just war theorists argue that some non-combatants render themselves liable in war, in virtue of their moral responsibility for unjust threats posed by combatants.² And yet the laws of war prohibit targeting non-combatants altogether.

Two responses to this apparent gap between the morality and law of war are common in the literature. One denies the 'reductivist' assumption that moral principles governing killing in non-martial contexts are applicable to war, and instead holds that war gives rise to special moral rights and duties. On this view, the supposed gap is illusory, because the laws of war correctly reflect war's *sui generis* morality.³ A different response accepts the reductivist assumption, and on that basis concludes that the laws of war misrepresent combatants' moral duties. Yet rather than argue for revisions to the LOAC that would more faithfully reflect those duties, many reductivists hold that the law is morally justified, despite its deviation from morality, because it minimizes the overall harmfulness of war as much as is practically feasible.⁴ For instance, while morality imposes a blanket prohibition on killing in unjust wars, a law to this effect would simply be ignored, and give belligerents no incentive to spare civilians, nor to surrender. By instead adopting laws that pair a general permission to target combatants

¹ See, most prominently, Jeff McMahan, *Killing in War* (Oxford: Oxford University Press, 2009).

² See, e.g., Helen Frowe, *Defensive Killing* (Oxford: Oxford University Press, 2014). Debate centres on what proportion of non-combatants are covered by the 'some'.

³ Michael Walzer, *Just and Unjust Wars*, 4th ed. (New York: Basic Books, 2006).

⁴ See, e.g. Jeff McMahan, "The Morality of War and the Law of War," in *Just and Unjust Warriors: The Moral and Legal Status of Soldiers*, eds. David Rodin and Henry Shue, 19-43 (Oxford: Oxford University Press, 2008).

with an exceptionless prohibition on targeting civilians, the law of war provides incentives that minimize death and suffering overall.⁵

In this essay we consider a third, and less commonly considered, response to the puzzle: The laws of war do indeed diverge from what morality requires. But this does not mean that they misrepresent what morality requires. For contrary to what most discussions of the puzzle assume, the laws of war do not claim to *represent* what morality requires in the first place. Rather, like all law, the LOAC claims to have authority over those to whom it applies, and thus to *change* what moral reasons they have. And if the LOAC in fact has the authority it claims to possess, then even though its norms diverge from the moral demands that ordinarily apply to agents engaged in harmful conduct, once the LOAC is in place, combatants have moral reason to follow these legally established norms.

A sophisticated version of this ‘authority response’ is central to Adil Haque’s *Law and Morality at War*, and much of our discussion will focus on Haque’s argument.⁶ Haque, in line with much of the existing literature on legal authority, offers a broadly *instrumentalist* argument for the law’s authority: the LOAC serves combatants by issuing directives that help those who (try to) follow them to do what they have independent moral reason to do; and it is this service that establishes the law’s practical authority over combatants. Crucially, even though Haque accepts that just and unjust soldiers are in very different moral positions, he believes that the laws of war nonetheless serve, and thus have authority for, both sides.

We argue that Haque’s proposal ultimately falls short. But its failure is instructive. It teaches us something important about the fundamental structure of instrumental justifications of authority, and about the relation between political and legal authority and the ethics of harm which, though central to much of our political life, has received relatively little philosophical attention.⁷ And once we

⁵ See, e.g., Janina Dill and Henry Shue, “Limiting the Killing in War: Military Necessity and the St Petersburg Assumption,” *Ethics and International Affairs* 26 (2013): 311-334.

⁶ Adil Ahmed Haque, *Law and Morality at War* (Oxford: Oxford University Press, 2017). Hereafter *LMW*. Unattributed page numbers in the text refer to this book.

⁷ For some exceptions, see David Estlund, “On Following Orders in an Unjust War,” *Journal of Political Philosophy* 15 (2007): 213-234; Massimo Renzo, “Political Authority and Unjust Wars,” *Philosophy and Phenomenological Research* (forthcoming in print, published online 2018): <https://DOI:10.1111/phpr.12487>; Jonathan Parry, “Authority and Harm,” in *Oxford Studies*

grasp which specific features of Haque’s account are problematic, we can also better understand what conceptual space there may still be for an authority-based solution to the puzzle of law and morality in war (though within the confines of this paper, we cannot do more than gesture at the structure of such a solution).

The essay proceeds as follows. After explaining Haque’s motivation for pursuing the authority response to the puzzle, we present his instrumental or ‘service’ argument for the authority of the laws of war (Section 2). Next we clarify the precise structure of Haque’s argument by distinguishing between instrumental justifications in general and what we call ‘remedial’ (instrumental) justifications, and showing that his argument (which depends on assumptions about morally problematic instances of ignorance or bias among combatants) falls into the latter category. ‘Remedial’ authority has only limited effect on the normative situation of its subjects. In particular, while (as Haque says) these subjects may have a permission “in the instrumental sense” (45) to obey authoritative directives even if this results in violating others’ rights, this permission does nothing to shield the subjects from blame for the violation, or from liability to pay compensation or suffer defensive harm. The permission instead merely bears on how the subjects ought to be guided in their deliberation by the authority’s directives (Section 3). In Section 4 we turn to the conditions under which the subjects indeed ought to be so guided. We argue that these are more demanding than is usually assumed. Even if obeying an authority improves a subject’s moral performance *compared to how she would otherwise act*, the authority’s directives are not morally binding if the subject could, within the limits set by her deficiencies that need correcting, adopt an alternative, and superior, means of improving her performance. In Section 5, we apply our more stringent justificatory test to the laws of war, and argue that Haque’s argument fails to vindicate the LOAC’s authority over all (or even most) combatants. In addition, the basic structure of Haque’s argument – with its focus on broadly epistemic considerations – makes it difficult to defend the symmetrical authority of the laws of war for just and unjust combatants alike. And yet this symmetry is a central feature of the laws of war, and their claim to authority. In Sections 6, we return to the relation between the law’s authority and its justification. We argue that, though Haque’s argument fails to establish the LOAC’s authority for unjust combatants, it may nonetheless be morally good for such combatants to treat the LOAC as if it had authority. More generally, the more

in Political Philosophy 3, eds. David Sobel, Peter Vallentyne, and Steven Wall, 252-277 (Oxford: Oxford University Press, 2017); Daniel Viehoff, “Legitimacy as a Right to Err,” *Nomos* LX (forthcoming).

demanding instrumental justification of authority we defend opens up conceptual space for rethinking the relation between an institution's claim to authority and its overall justification.

2. Haque on the Authority of the LOAC

Haque pursues the authority response because he deems implausible both of the alternative, and more prominent, responses to the puzzle with which we started this essay. Specifically, though he accepts the reductivist account of the fundamental moral rights and duties of combatants, he also holds that simply pointing out that the laws of war may minimize harm (as many reductivists do) is philosophically inadequate.⁸ As Haque emphasizes, the laws of war are, indeed, *laws*; and like all laws, they claim authority over those to whom they apply.⁹ And even if the harm-minimisation view could successfully justify the LOAC's *existence*, it would nonetheless fail to vindicate its claim to *authority* over combatants.

Haque here invokes a distinction familiar from debates within political and legal philosophy between an institution's *justification* and its *authority*.¹⁰ An institution is *justified* if its existence and activity is all-things-considered valuable. In virtue of its value, individuals may have moral reasons (or even duties) to create the institution, support it, refrain from undermining it, etc. An institution has *authority* over an agent S if it has the moral power to impose duties or binding reasons on S to ϕ simply by directing S to ϕ . Two points about authority are worth highlighting. First, authoritative directives are meant to have a privileged status in the subject's practical reasoning: they are meant to largely settle what action the subject is to undertake, even in the face of countervailing reasons. We will call such reasons 'binding'. Second, at least within certain limits, an authority's directives are intended to be binding even if the subject takes them to be mistaken, and even if the directives are in fact mistaken.

⁸ Haque also objects that the harm-minimisation view, in virtue of its focus on aggregative outcomes, falls foul of familiar objections to consequentialism (*LMW*, 38-43). We largely leave this point aside here.

⁹ Haque here follows a familiar line of argument in legal philosophy. See, e.g., John Gardner, "How Law Claims, What Law Claims," in his *Law as a Leap of Faith: Essays on Law in General*, 124-148 (Oxford: Oxford University Press, 2012).

¹⁰ See, e.g., A. John Simmons, "Justification and Legitimacy," *Ethics* 109 (1999): 739-771, for a version of this distinction (though Simmons' differs in some respects from the one we draw here).

Crucially for Haque's purposes, an institution may be justified without having authority: it may be good that it exists, and so S has reason to support it, and yet its directives do not impose duties on S to do as directed. This may be because the value of the institution's existence is in fact independent of its capacity to impose duties. (In the case of a legal system, for example, the benefit may rest solely on the additional incentives provided by the threat of punishment for breaking the law.) Alternatively, the value of the institution's existence, though not unrelated to its capacity to impose duties, may not suffice to establish that S *in particular* has a duty to obey. (Perhaps the law's directives contribute to the relevant good insofar as they guide others, but do not depend on S's obedience.)

So, Haque points out, even if the harm-minimization view is correct, and the laws of war produce overall good consequences by reducing net suffering, this does not establish that combatants are bound by the law's directives.¹¹ The harm-minimisation view thus fails to do justice to the laws of war, which qua law have as their constitutive aim to "guide human conduct through positive normative standards" (50). And, more urgently, it fails to address the problem of divergence that we started with. While it may resolve the tension between the law and morality of war from the perspective of the *legislator* (since it can explain why we ought to *put in place* laws whose content differs sharply from the applicable moral norms), it leaves the tension untouched from the perspective of the law's *subject*. Combatants are still confronted with diverging legal and moral norms. The fact that the law reduces *overall aggregate* harm does not give combatants reasons why *they specifically* should follow the law, rather than try to follow the demands of morality.

Haque thus sets out to offer an alternative account of the laws of war that vindicates the LOAC's claim to authority and establishes that combatants have moral reasons to follow the law's directives, even in cases where those directives depart from independent moral reasons. To do so, Haque adopts an instrumental approach, drawing on Joseph Raz's influential 'service conception' of authority. An instrumental account of authority starts from the observation that, though we have ultimate reason to do what morality requires of us, we are often unable to conform to these ultimate reasons by pursuing such conformity directly. We are limited in all sorts of ways that affect our abilities to access those reasons, assess their force, and act in light of them. Given our limitations, we have

¹¹ For a version of the harm-minimisation view which explicitly takes the latter to follow from former, see Henry Shue, "Do We Need a 'Morality of War'?" in *Just and Unjust Warriors: The Moral and Legal Status of Soldiers*, eds. David Rodin and Henry Shue, 87-111 (Oxford: Oxford University Press, 2008).

instrumental reasons to adopt means that will enable us to best approximate full conformity with the ultimate moral reasons.¹² In some contexts, the optimal means will involve acting on one's own assessment of the ultimate reasons at issue. But sometimes the best means will be to take a less direct route, such as letting the directives of another agent or institution settle how one acts. And in such cases, there is an instrumental reason for being bound by these directives, and acting accordingly.

On this view, authorities have the power to give binding directives to their subjects because, and to the extent that, this power helps the subjects better achieve what they have independent reason to do. As Raz formulates this central idea (which he terms 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.¹³

Authorities, Raz suggests, serve their subjects in four main ways. First, following a common authority can enable valuable coordination among subjects. Second, a policy of obedience can allow subjects to avoid the costs of complex deliberation. Third, an authority may possess greater expertise than the subject on some important matter. Fourth, deference to authority allows subjects to circumvent various deficiencies and vices that afflict their deliberations.

Service accounts aim to explain the central features of authority outlined above: that subjects are (in at least some cases) bound by the authority's directives even if those directives *fail* to reflect the balance of independent reasons (that is, the reasons the subject would have absent the authority's directives). This is because authorities need not be infallible in order to successfully serve their subjects. The standard for service is essentially comparative: which of the available means for conforming to the ultimate moral reasons is superior? Even if the authority makes occasional mistakes, a policy of following its directives may be better than one of acting on one's own assessment of the fundamental moral reasons if one's own mistakes are more severe. Similarly, even if one sometimes gets the matter right where the authority goes wrong, a general policy of not second-guessing the

¹² We say more about the relation between instrumental and moral reasons in Section 3 below.

¹³ Raz, *The Morality of Freedom*, 53.

authority may be superior, if one cannot reliably distinguish between cases where disagreement is due to the authority's mistake or one's own. Service accounts thus explain in a satisfyingly unified way how, when law and morality conflict, one can nonetheless have binding reason to follow the law. When an authority serves its subject, obedience is an indirect, but more effective, way for the subject to respond to the applicable moral reasons. As Haque puts it:

...often we are more likely to satisfy the moral standards that apply to us, not by applying those standards, but instead by following a well-designed decision procedure. In such cases, we have decisive reason to adopt that decision-procedure rather than exercise our own (more fallible) moral judgment on a case-by-case basis. We may say that an act that is permissible under a rule or decision procedure that we have decisive reason to adopt is permissible in the *instrumental* sense. (45)

The first step in Haque argument holds that the LOAC has authority to the extent that it satisfies the normal justification thesis:

The LOAC claims that combatants should follow the law rather than their own moral judgment when the two conflict. According to the service conception, this claim is vindicated only to the extent that combatants will better conform to the moral reasons that apply to them by following the law than by trying to act directly on those reasons. If combatants will make morally better decisions by ignoring the law and relying on their own judgment, then they have decisive reason to do just that. (45-46)

Haque's second step is to argue that core aspects of the LOAC do satisfy this justificatory test.¹⁴ He focuses, in particular, on the law's prohibition on targeting civilians and its non-prohibition on targeting combatants. On Haque's view, these aspects of the law *do* provide a service to combatants, who are therefore bound to follow the LOAC's rules.

Though it is common to defend legal authority by appeal to its role in facilitating valuable coordination among many agents, Haque rejects this approach when it comes to the LOAC. This is

¹⁴ Haque's proposal also has important reformist aspects: "We should defend, *interpret*, and *develop* the law of war by reference to the service view." (LMW, 21) (emphasis added)

because, he explains, “in war we compete with our adversary, and our moral reasons do not depend on their behaviour. We can avoid targeting civilians, and have moral reason to do so, irrespective of how our adversary behaves.” (42) Haque instead adopts a broadly epistemic strategy to ground the LOAC’s authority: the law of war serves combatants by compensating for (i) their poor ability to make individualised judgments of combatant and non-combatant liability, and (ii) the susceptibility of those judgements to ‘non-rational’ forces, which include emotions (such as fear and hatred) as well as biases (such as underestimating one’s own fallibility and overestimating the wrongdoing of outgroup members). Given these epistemic deficiencies, Haque maintains that both just and unjust combatants will improve their moral performance by fighting in accordance with the laws of war, rather than acting on their own assessment of the moral reasons that apply.¹⁵

Just combatants, by following a general rule that restricts intentional attacks to combatants, act wrongly less often than they would without the law’s guidance. There may well be cases in which the law permits targeting a non-liable soldier, or prohibits targeting a liable civilian. In these instances, following the law may lead just combatants away from what they have ultimate moral reason to do. But, Haque argues, given the relative infrequency of non-liable combatants and liable civilians, and the difficulty of identifying them, just combatants avoid wrongdoing in more cases by deferring to the law, compared to acting on their own assessment of individuals’ liability.

The case of unjust combatants is less straightforward, since Haque accepts the reductivist claim that such combatants have no legitimate targets. Whether they kill combatants or civilians, unjust warriors will typically violate individuals’ rights not to be killed. However, Haque argues, though all of unjust combatants’ targeting options (short of ceasing to fight) involve grave moral wrongdoing, the laws of war may nonetheless provide an important service to them. Though unjust combatants violate basic rights whomever they kill, some violations are worse than others. Following Seth Lazar, Haque maintains that, other things being equal, it is morally worse to target innocent civilians than to target innocent combatants. This putative asymmetry is grounded in, for example: civilians’ greater

¹⁵ Haque focuses on the role authority plays in minimizing wrongdoing, rather than in improving conformity to reasons more generally, as Raz does. This may allow Haque to escape some worries that have been raised against Raz’s view. (See, e.g., Jonathan Quong, *Liberalism Without Perfection* (Oxford: Oxford University Press, 2011), Ch.4.) While we also limit ourselves to improving conformity to *moral* reasons, we do not assume that the authority must improve conformity with moral *duties*, for reasons that emerge in Section 3.

vulnerability to harm, their lesser capacity to defend themselves, and that combatants chose to make themselves lawful targets, often (at least partly) in order to draw fire away from civilians.¹⁶ This differential wrongness explains how the law can serve even unjust combatants. Combatants fighting unjust wars will indeed commit nothing but wrongs; yet they will commit less grave wrongs if they fight according to the laws of war, and thus concentrate their attacks on combatants rather than civilians, than they would if they instead relied on their own moral judgment to identify permissible targets.

Haque's overall project is thus to establish that:

...by following the LOAC (suitably interpreted and developed) all combatants can ensure that they will fight, if not permissibly, then less wrongfully than they would otherwise. In other words, I hope to vindicate the LOAC's claim to legitimate authority over just and unjust combatants alike. (49)

If Haque's 'Service View' of the LOAC is correct, then the puzzle about the relation between the law and morality of war can be resolved: the law of war deviates from the underlying morality, but it does so in order to improve combatant's capacity to act on what morality requires. Combatants thus have reason to abide by the law, even if it deviates from the ultimate moral requirements governing killing and harming in war.

In what follows, we criticise each step in Haque's argument. We argue that the instrumental justificatory test invoked by Haque (and by many other legal and political theorists) sets too low a bar for authority. Once we apply the correct, more demanding test, it is much harder to vindicate the law's authority in terms of the broadly epistemic service it provides to combatants. First, however, we clarify further the precise structure of instrumental accounts in general, and Haque's argument in particular, by introducing, and responding to, two objections that may be raised to it.

3. Instrumental Reasons to Obey and Moral Duties Not to Harm

¹⁶ Seth Lazar, *Sparing Civilians* (Oxford: Oxford University Press, 2015).

On Haque's view, combatants are bound by the law's directives for instrumental reasons: if a combatant follows the law's directives, he will better conform to the moral reasons that govern killing and injuring. One objection to this proposal concerns the relation between an agent's instrumental reasons and her fundamental moral reasons.

An initial worry is that instrumental reasons are not genuinely *moral* reasons, and so cannot do the normative work required of them. But this concern can be addressed. The distinctions between moral and non-moral reasons, and between instrumental and non-instrumental ones, cut across one another. We have an instrumental reason to ϕ when ϕ -ing facilitates another action (ψ) that is independently worthwhile.¹⁷ When we have a *moral* reason or duty to ψ , we very often have derivative reason to undertake another, facilitative action (ϕ) that enables us to conform to that reason or duty. The instrumental or facilitative reason we have then inherits its normative force from the ultimate moral reason or duty. So the reason to undertake the facilitative action is simultaneously instrumental and moral.

But even with this clarification in place, one might still object that instrumental (moral) reasons are not of the right type to change combatants' normative situation in the way Haque's account seems to require. Consider the following example:

Mistaken Soldier. Private and General believe their side is fighting a just war. In fact, it is seriously unjust. General's view of which enemy soldiers it is necessary to kill to bring about victory is more fine-grained than Private's. If Private obeys General's directives about which enemy soldiers to kill, he will kill some enemy soldiers, but fewer than if he tries to pursue victory by acting on his own judgment.

On Haque's account, Private is under a duty to obey General: Private, by following General's orders, will commit fewer wrongful acts than he would by trying to act on the underlying moral reasons directly. And yet that seems to have the counter-intuitive (indeed, contradictory) result that Private has binding moral reasons to kill some enemy soldiers, as well as a moral permission (albeit "in the

¹⁷ Cf. Joseph Raz, "The Myth of Instrumental Rationality," *Journal of Ethics and Social Philosophy* 1 (2005): 1-28.

instrumental sense” (45)) to kill them, even though he is clearly under a moral duty not to kill any (and would intuitively become liable to blame, compensation, and defensive harm if he obeyed).¹⁸

While we think this is a serious worry, we offer a diagnosis of the problem and a response on Haque’s behalf. First, we suggest that the problem with cases like *Mistaken Soldier* arises, not from the instrumental nature of the authority relation per se, but from its specifically *remedial* character. Second, we show that in remedial cases, the ‘instrumental’ (or ‘remedial’) permission must be understood very narrowly, bearing only on what considerations should guide Private’s deliberation, but not on Private’s ultimate moral duties (nor his liability to defensive harm, compensation, or blame).

Instrumental reasons for action depend, by their very nature, on certain inabilities or deficiencies (in a very broad sense) of the agent. Omnipotent, omniscient, and non-akratic agents

¹⁸ This worry focuses on the moral character of Private’s reasons (and the corresponding moral valence of his action). Someone may also push back against the idea that Private’s obedience would be justified on instrumental grounds. Private has instrumental reason to obey because obedience facilitates conformity to his moral duty not to kill innocent people. But, the objection goes, Private could also adopt another option that is both clearly superior to disobedience and incompatible with it: he can simply not kill any innocent people. Yet if that alternative option is available to him, then how could his obedience be instrumentally justified? The problem with this objection, however, is that it elides the distinction we introduce below between two types of capacities: those that bear on an agent’s fundamental moral reasons or duties, and those that bear on his instrumental reasons. We cannot here develop an account of which capacities fall on either side. But an example should help show that it would be very counter-intuitive to deny that someone in Private’s situation has instrumental reasons that flow from his ignorance of what he has moral reason to do: Imagine Sam promises his mother to turn off the light in her garage at 6pm sharp. But now it is 5.55pm, and while Sam remembers that he promised his mother *something*, he cannot remember what it is. He walks around the house trying to remember, and finds himself in the garage at 5.59. There is no question that he *could* turn off the light, in one morally relevant sense: turning the light switch would kill the light, he has the physical strength to turn it, etc. But there is also no question that, given his current ignorance of what he has a duty to do, he has good instrumental reason to walk back to the kitchen and call his mother to find out what he had promised her; and he has reason to do this even if, as a result, he doesn’t turn the switch at 6pm, but only at 6.05. That he has an instrumental reason is shown, e.g., in how we might criticize him: not only for failing to turn off the light at 6pm sharp (which he promised), but also for standing in the garage pondering what to do when he could instead go to the kitchen and call his mother, thereby minimizing how far he falls short of fulfilling his promissory duty.

would have no instrumental reason to adopt facilitative actions, since they can comply with their ultimate reasons directly. Only imperfect agents like us have instrumental reasons. But not all of our deficiencies are morally on a par. Some deficiencies or inabilities are part of the background against which our moral duties are determined in the first place: if we could run as fast as a race-car, or perform calculations as quickly as a super-computer, or knew the entire human genome, we might have duties to save others (because doing so would be feasible and relatively costless for us) that we currently lack. Yet other deficiencies are not part of the background, and do not stand in the way of our having certain moral duties. Instead they simply prevent us from doing what we in fact have a moral duty to do. Call the latter ‘moral deficiencies’. Where an instrumental action serves to remedy moral deficiencies, we will call the subject’s reasons for action ‘morally remedial’, or ‘remedial’ for short.

Our concern here is not to develop an account of which deficiencies fall into either category, but merely to flag that, insofar as one accepts that there are different categories, this has important implications for the normative effects of instrumentally justified authority.¹⁹ To show this, we consider first a non-remedial instance of authority. In non-remedial cases, an instrumentally grounded

¹⁹ There is an extensive literature on the relation between (in)capacities and moral reasons, and the associated question what it takes for moral considerations to be ‘available’ or ‘accessible’ to an agent so as to affect what reasons she has. We take the limited claims we need for our argument here to be compatible with a wide variety of positions within this literature, and so largely abstain from discussing it in detail. First, while we assume (because we find most plausible) that what moral reasons we have is partly dependent on (some of) our capacities, this assumption is not in fact necessary for our central argumentative purposes. (For a recent defence of the assumption’s plausibility, see Errol Lord, “Acting for the Right Reasons, Abilities, and Obligation,” in *Oxford Studies in Metaethics* 10, ed. Russ Shafer-Landau, 26-52 (Oxford: Oxford University Press, 2015).) The crucial contrast we wish to draw, between how fundamental moral reasons and instrumental reasons are respectively affected by deficiencies, would stand even if moral reasons were independent of our capacities. Second, while we assume that we can (and often do) have moral duties that we are ignorant of, even those who believe that ignorance and similar epistemic deficiencies about our moral reasons can limit what moral reasons or duties we have should not wish to deny that some moral duties apply to us even if we are currently ignorant of them. For a general discussion of the link between capacities and reasons, see Bart Streumer, “Reasons and Abilities,” in *The Oxford Handbook of Reasons and Normativity*, ed. Daniel Star, 233-254 (Oxford: Oxford University Press, 2018).

permission need not be in tension with our moral duties, even when the moral stakes are high. For at least sometimes the instrumentally beneficial intervention of an authority properly changes what moral reasons or duties we have, by changing the very considerations on which our moral duties depend. Consider the following example:

Beach Savior: Ben goes for a stroll on the beach when he spots three people drowning. Swimmers 1 and 2 are some distance from the shore. The beach is known for its treacherous undercurrents. Though there is a hidden underwater sandbank that Ben could safely take to reach the two, he is non-culpably ignorant of where that sandbank is. As a result, saving Swimmers 1 and 2 would be quite dangerous. Given the risk, Ben is under no moral duty to save them. Swimmer 3, however, is very close to the beach, and Ben can easily save him. So Ben has a moral duty to save Swimmer 3.

But before Ben saves Swimmer 3, Old Rick appears on the beach. He is in a wheelchair, and so cannot himself engage in any rescue. But he is an expert on local currents and sandbanks. If Ben follows Old Rick's directives, he can walk along the hidden sandbank to reach Swimmers 1 and 2, and save them in an essentially riskless fashion. However, if he follows Old Rick's directives and saves Swimmers 1 and 2, he won't have time to save Swimmer 3, who will drown.

In this case, Old Rick plausibly acquires instrumentally justified authority over Ben, since following his directives helps Ben act on certain moral reasons (though not duties) he has anyway (to save lives). By issuing directives, Old Rick changes the underlying distribution of benefits and burdens associated with Ben's options (by making saving the greater number relatively costless), and thereby also changes Ben's moral duties and permissions. As a result of Old Rick's authority, Ben acquires a moral duty to save Swimmers 1 and 2, and loses the duty he had to save Swimmer 3. The intuitive problem we encountered in the case of *Mistaken Soldier* is absent here.

Contrast this with a variation on the case, in which obedience to authority is remedial:

Beach Bum: The situation is as in *Beach Savior*, except that the potential rescuer is Bella, the lifeguard on duty. If Bella knew the location of the hidden sandbank, she would have enough time to first save Swimmers 1 and 2, and then rescue Swimmer 3. But Bella doesn't know

where the sandbank is. The location was part of her mandatory lifeguard training, but she was too lazy to show up to that particular training session. If she follows Old Rick's directives, she will be able to walk along the sandbank (though more slowly than if she had learned about its location herself) and save Swimmers 1 and 2. By the time she has saved them, however, Swimmer 3 will have drowned.

Both Bella and Ben have the same basic deficiency: ignorance of the location of the sandbank. And both are morally bound to obey Old Rick's directives, because obedience enables them to compensate for that deficiency and rescue Swimmers 1 and 2, though this requires letting Swimmer 3 drown. However, there is an important moral difference between the cases. *Unlike* Ben, Bella still intuitively violates a duty she had to Swimmer 3. This is reflected in our judgment that she is blameworthy for his death, and that she is liable to pay compensation (and potentially suffer defensive harm, if it were possible to protect Swimmer 3 by harming Bella). Insisting that Bella had a duty to protect Swimmer 3 is nonetheless compatible with saying that Bella is bound by Old Rick's directives, and would do something wrong if she failed to be guided by them – even though being so guided will effectively make it impossible for her to save Swimmer 3. Given Bella's ignorance of where the sandbank is, the best she can do is to rescue the two swimmers while letting the third drown.

In light of such examples we suggest that the 'instrumental permission' that Haque invokes may come in two varieties. In non-remedial cases, it may indeed change the subject's fundamental moral duties. But in remedial cases, like Bella's, it plays a much narrower normative role. While those subject to remedial authority are permitted to obey authoritative directives even if this results in violating others' rights, such permissions do nothing to extinguish the duties correlative to these rights, nor shield the subject from blame, or liability to pay compensation or suffer defensive harm, for the violation. The relevant permission instead merely bears on how the subject ought to be guided in her deliberation by the directives.

What about cases like *Mistaken Soldier*? Private, we assume, fails to fully respect the innocent enemy's rights because his "beliefs are unreliable, the available evidence is incomplete or misleading, or [his] moral reasoning is distorted by emotion or bias" (45). Given these deficiencies, the best that Private can do is to be guided by General's directives. When Private is so guided, he ought to treat the option of killing those he is ordered to target as permissible, even though they in fact have a right against being killed. He ought to do so because unless he does, he cannot effectively be guided by General's directives, and – the assumption is – will end up killing more innocent people. But this, we

have seen, need not be incompatible with Private's violating the rights of his opponents (and being liable to be blamed, defensively harmed, and forced to pay compensation), as long as we recognize that General's authority is remedial only.

The intuitively problematic character of *Mistaken Soldier* thus rests, not on the instrumental justification of General's authority, but on its remedial quality. Such remedial authority must be understood to have only very modest normative implications. These implications are, however, sufficient for Haque's specific purposes; for what he seeks to establish is that soldiers have moral reason to be guided by the LOAC's directives when determining whom to target in war.

4. Authority, Deficiencies, and Options

We now turn to Haque's discussion of the conditions under which authority is justified, and so has the normative implications identified in the previous section. At the heart of Haque's view is the idea that subjects effectively face a choice between two different means for conforming to morality's demands. One means (M1) is to consider the underlying moral reasons, form a corresponding judgment about what course of action they require, and act on it. Another (M2) is to consider an authority's directives, form a judgment about what the directive requires, and act according to that (even if, by the subject's own independent assessment of the underlying moral reasons, another action seems superior). The subject ought to adopt M2, and thus be guided by the authority's directives, if, by trying to follow the authority's directives, the subject will do better than she would if she instead adopted M1, and tried to act directly on the underlying moral reasons.

This section has two aims. The first is to clarify the background conditions that enter into making the appropriate comparison between M1 and M2. The second is to argue that the comparison between M1 and M2 is not sufficient to establish that the subject ought to be guided by the authority's directives on instrumental grounds.

When making the comparison between M1 and M2, we must consider how well the agent would do when *trying* to follow the authority's directives or the underlying moral reasons respectively. And how well the subject will do when *trying* to follow them will depend on various *actual* capacities or deficiencies that she has: if she is ignorant of her moral duties, or tends to give inadequate weight to some of them, then this will shape how effectively trying to follow them will translate into actually conforming to them. But, similarly, how good the subject is at understanding the authority's directives, or figuring out what course of action they require of her, will determine how effectively the subject's

being guided by the authority translates into the subject's conforming to the underlying moral duties. Thus the instrumental argument must treat as a baseline certain deficiencies; and an argument for being guided by the authority must depend on the thought that the authority's guidance is effective *despite the subject's deficiencies* (most obviously – though not exclusively – the deficiencies that make simply being guided by the underlying moral reasons an ineffective strategy for achieving actual conformity to these reasons). Any instrumental justification of authority must thus be *deficiency-compatible*. As Jeremy Waldron points out, in the context of war the law must be “administrable” at the hands of imperfect, morally deficient combatants if it is to actually help these combatants to overcome or compensate for their shortcomings.²⁰

We want to emphasize again here that the deficiencies that determine whether the subject has instrumental reason to be guided by the authority must be distinguished from those deficiencies that determine what *fundamental* moral duties or reasons the subject has. Bella's culpable ignorance of how to save all three swimmers, for example, did not undo her fundamental moral duty to save all three, or her corresponding liability to be blamed, suffer defensive harm, etc. But it did bear on what instrumental reason she had in light of this fundamental duty: in this case, a reason to follow another's directives so as to (at least partially) make up for her own (culpable) deficiency. Moreover, Bella's deficiency not only gave her reason to adopt indirect means, but also reason to adopt indirect means that are in fact compatible with her deficiency, and thus actually useful for her.

Let us turn to the second point. Haque, we saw, seeks to establish authority by comparing M1 and M2. In this he follows Raz's influential formulation of the normal justification thesis, which compares how the subject acts “if he accepts the directives ... as authoritatively binding and tries to follow them” to how he would act if he tried “to follow the reasons which apply to him directly.”²¹ Haque, in the same vein, writes that the law has legitimate authority if “individuals who conform to the law ... perform fewer or less wrongful objectively impermissible acts than they would by relying on their own moral judgment.” (48)

²⁰ Jeremy Waldron, “Deep Morality and the Laws of War,” in *The Oxford Handbook of Ethics of War*, eds. Seth Lazar and Helen Frowe, 80-98 (Oxford: Oxford University Press, 2018). For a more general discussion of this point, and the role it plays in expertise-based justifications of political authority, see Daniel Viehoff, “Authority and Expertise,” *Journal of Political Philosophy* 24 (2016): 406-426.

²¹ Raz, *The Morality of Freedom*, 53.

We suggest that this comparison is insufficient to establish the instrumental justification of authority that Haque is after. Such justification also requires a comparison between how the subject would do if she were guided by the authority's directives (M2), and how she would do if she adopted *another compensatory strategy* (M3) that is compatible with her deficiencies.

Call the view according to which authority is justified as long as M2 is superior to M1 the 'narrow' instrumental justification, and the alternative view (requiring comparison between M1, M2, and M3) the 'wide' version. The narrow version, we suggest, has implausible results. For it implies that a subject who could improve her conformity with reason *even more* by ignoring an authority that satisfies the narrow version, and instead following some other strategy that even more effectively compensates for her deficiencies, yet is incompatible with obedience to the authority, would still be bound by the authority's directives. It is hard to see how the binding force of the authority's directives would be instrumentally justified in this case.²² To illustrate, consider:

Ambulance Driver: Dean the Ambulance Driver is good, but not perfect, at driving fast on busy roads. He gets excited when driving fast, which negatively impacts his concentration and judgement. If he drives according to his own judgement, he poses $risk_n$ of killing pedestrians. But if he obeys the directives of Vicky the Radio Controller, who directs him along less busy streets, he will pose $risk_m$ (where $risk_m < risk_n$). However, if he listens to soothing music on his headphones, he will be unable to hear Vicky but will be less prone to excitement. If Dean drives while using his headphones he will pose $risk_p$ (where $risk_p < risk_m$).

According to the narrow version of the instrumental justification, Dean is morally bound to obey Vicky's directives. But this is clearly the wrong verdict. We should compare all of Dean's available options for compensating for his risk-imposing excitement, and conclude that he should ignore Vicky and put on his headphones. Indeed, Raz's more recent discussions of authority indicate that he recognizes the need for the wider justification. For as Raz points out, if an agent is confronted with two would-be authorities, each of whom would improve her reliability, but not both of whom can be

²² On this point, see Patrick Durning, "Joseph Raz and the Instrumental Justification of the Duty to Obey the Law," *Law and Philosophy* 22 (2003): 597-620; Matthew Coakley, "On the Value of Political Legitimacy," *Philosophy, Politics, and Economics* 10 (2010): 345-369.

followed simultaneously, then the agent is bound only by the directives of the better of the two.²³ This point generalizes to authority-independent means: if, instead of acting on my own judgment or following the authority, I could listen to soothing music (or follow some other rule, etc.), and doing *that* would lead to a superior outcome, then that is what I have instrumental reason to do.

It should be clear that whether a subject has a relevant option M3 that is superior to either M2 or M1 will also depend on whether trying to act in accordance with M3 (say, following another rule than that promulgated by the authority) will be effective in improving the subject's moral conformity despite his deficiencies. In other words, M3 too must be deficiency-compatible. This significantly restricts the range of alternative means that are admissible into the authority-determining comparison.

At the same time, it is important to recognize that the deficiency-compatibility restriction does not entail that we should only compare what *actually* happens if the subject is not guided by the authority with what happens if he tried to be so guided (and conclude, on the basis of the latter's superiority, that the authority is instrumentally justified). Imagine, for example, that Dean is aware of the compensatory benefits of using his headphones, but can't be bothered to do so. He is, however, willing to obey Vicky's directives. (Call this variation *Lazy Ambulance Driver*.) If Dean is not guided by authority, what will actually happen is that he drives according to his own judgement. But despite the fact that obeying Vicky is superior compared to what Dean will actually do, the accessibility of a superior option (using his headphones) undermines his being instrumentally bound to obey her. More generally, the subject has instrumental reason to adopt M3 instead of M2 as long as, *if he tried* to follow that alternative rule, he would in fact do better than *if he tried* to follow the authority's directives instead. If *given his deficiencies* the subject would do better, instrumentally, if he tried to follow the alternative strategy, rather than to follow the authority's directives, then it would be instrumentally rational to follow that strategy.

5. The Options of Combatants

Haque's attempt to vindicate the law's authority over just and unjust combatants assumes the narrow reading of the instrumental justification (which compares combatants' moral performance under only two conditions: if they follow the LOAC and if they act on their own judgement). We argue next that once we apply the correct, and more demanding, wide reading of the instrumental justification

²³ Joseph Raz, *Between Authority and Interpretation* (Oxford: Oxford University Press, 2011), 143.

introduced in the previous section, Haque's defence for the LOAC's authority becomes much less plausible. In particular, it becomes possible that just combatants would not best be guided by the LOAC; it becomes doubtful that unjust combatants will best be guided by the LOAC on the grounds that Haque adduces; and it becomes unlikely that just and unjust combatants will each best be guided by rules with the same content.

We start with the case of just combatants. In line with the narrow justification, Haque's argument regarding just combatants compares how many wrongful killings they will perform if they adopt a policy of obeying the law's strict prohibition on targeting non-combatants, or if they try to use their own judgement to assess the liability of potential civilian targets in each choice scenario. Haque argues that combatants will perform fewer wrongful killings if they follow the former strategy than the latter, and that in virtue of this superiority the LOAC has authority. This is true even though, by following the law, combatants will have to forsake opportunities to advance just goals (and thereby protect innocent people) by means of targeting liable civilians.

But once we introduce the wider version of the instrumental argument, we must also compare other strategies that just combatants could use to compensate for their deficiencies (which, for Haque, are their epistemic deficiencies with respect to identifying liable civilians). If such strategies are available, then the laws of war may lack authority over them after all.

Moreover, there are reasons to think that such strategies are available. When it comes to targeting civilians, following one's own assessment of the permissibility of targeting civilians in every case is not the only alternative to following the law's strict prohibition. Another option for just combatants may be to adopt a policy of following a prohibition in most cases, but not in cases where the liability of a civilian is relatively *clear*. Instrumental accounts of authority often include such an exception²⁴, and Haque accepts that they obtain with respect to other aspects of the LOAC. (For example, he argues that just combatants should ignore the law's non-prohibition on targeting combatants when it is obvious that a combatant performs a non-combat role.²⁵) It doesn't seem implausible that by following a limited prohibition of this type, just combatants will strike a morally superior balance between avoiding targeting non-liable people and protecting innocents from unjust

²⁴ See, e.g., Raz, *The Morality of Freedom*, 61-62.

²⁵ *LMW*, 45 and 86.

attack.²⁶ If so, then the law's exceptionless prohibition lacks authority for just combatants, because it fails to satisfy the wide version of the instrumental justification, even if it satisfies Haque's narrow version.

Haque may reply here that the moral risk of targeting civilians is sufficiently high that following an exceptionless prohibition is superior to any available alternative strategy, and the law's prohibition thus in fact satisfies the wide justificatory test as well. But the extent of this risk varies greatly across different theories of civilian liability. If the correct theory holds that liability is distributed quite widely among certain categories of relatively identifiable civilians (for example, those who work in war-related heavy industries), then the moral risk of targeting does not seem high enough to justify the law's prohibition.²⁷ Haque himself proposes a complex account of civilian liability that is highly restrictive. His view places a lot of weight on the distinction between posing a direct unjust threat oneself and contributing to unjust threats posed by others, and argues that the latter is rarely a ground of liability to defensive killing.²⁸ Though ingenious, Haque's view is controversial and open to important objections.²⁹ To the extent that civilian liability is more widespread than Haque allows, the law's service to just combatants is correspondingly weakened.

It is also worth flagging a second problem with Haque's argument that emerges from reflection on the case of just combatants. Haque's overarching aim is to do justice to the LOAC *as law*; and as he explains, law has as its constitutive aim to "guide human conduct through positive normative standards" (50). Justifying a law's authority thus not only requires establishing that its content makes a moral difference to what its putative subject ought to do. It also requires showing that that fact that this content is *posited or promulgated by an authority* makes an essential contribution to that change in the subject's normative situation.

²⁶ For further discussion of this idea, in the context of non-international armed conflicts, see Christopher Finlay's contribution to this symposium.

²⁷ See Frowe, *Defensive Killing*, Chs.6-8. For a different kind of argument in support of broad civilian liability, see Victor Tadros, "Orwell's Battle with Brittain: Vicarious Liability for Unjust Aggression," *Philosophy and Public Affairs* 42 (2014): 42-77.

²⁸ *LMW*, Ch.3 and Appendix.

²⁹ For trenchant criticism, see Helen Frowe's contribution to this symposium.

Despite his stated ambitions, Haque's own account of how the law changes what combatants (and perhaps especially just combatants) ought to do makes remarkably little reference to *positive* normative standards. Haque's argument does not show that just combatants should follow the law, even if it occasionally misleads them. Instead it shows only they should follow *a rule with a specific content*; and his argument for this seems completely detached from whether the law, or any other positive norm, promulgates that rule. Even if there were no law, or other positive social norm, requiring just combatants not to target civilians, each of them would – on Haque's account – have reason to adopt a rule of his own to this effect, given that this would minimize the risk of wrongdoing. The fact that the prohibition on targeting civilians has the status of law seems to add nothing to the instrumental justification of its bindingness.

Now there may be ways in which Haque's account could accommodate this concern. Perhaps the fact that the rule has been suitably posited (say, via social or legal conventions or agreements) makes it especially likely that the rule is one worth adopting, or makes following the rule especially easy. If so, then the fact that there are positive norms prohibiting targeting civilians may be relevant after all; and so if these positive norms had articulated another rule, combatants would have reason to follow that rule instead (at least within certain limits). But nothing Haque says provides support for this line of thought.

Indeed, Haque's actual argument for the instrumental value of following a rule against targeting civilians threatens to effectively undermine any such argument for the bindingness of the rule *qua* positive normative standard. His account of the benefit of following the rule rests solely on the rule's content, and not on its being posited. Furthermore, recognition of the rule's usefulness (at least as Haque presents it) requires *only* the observation that many civilians are non-liable, and the further thought that we have stringent reason to minimize the risk of targeting non-liable civilians by mistake. But then, if just combatants are offered a choice between following posited legal or social rules governing the targeting of civilians, and instead adopting the particular rule of not targeting civilians at all, it is unclear what reason they could have to do the former rather than the latter. In particular, since posited rules can go wrong, simply adopting the relevant substantive rule directly may be presumptively superior. But then Haque's argument, even if it establishes the bindingness of a rule such as that posited by the laws of war, does not establish, and in fact undermines, the bindingness of the laws of war *as positive norms*.

Next consider the case of unjust combatants. Haque's argument is, remember, as follows: Unjust combatants typically commit a serious moral wrong by fighting, since both the enemy

combatants and the enemy civilians are morally innocent. Still, it is a greater wrong to kill civilians than to kill even innocent armed enemy soldiers. So the unjust combatant will “perform fewer or less wrongful objectively impermissible acts” by obeying the LOAC and targeting only enemy combatants “than they would by relying on their own moral judgment” of whom to target (48).

But as the previous section showed, comparing these options isn’t enough to establish the authority of the LOAC’s prohibition on targeting civilians. Following the LOAC must also be superior to alternative deficiency-compatible strategies the subject could adopt. To determine whether such a strategy may be available, we must get clearer on how the LOAC’s directives differ from unjust combatants’ fundamental moral reasons, and thus what moral downsides there could be to following the LOAC as an unjust combatant.

The most obvious way in which the LOAC’s rules deviate from an unjust soldier’s moral obligations is that the LOAC explicitly prohibits targeting civilians, but does not equally prohibit targeting enemy combatants. By contrast, as a matter of morality, the unjust soldier is under a duty not to kill either. One might think that this divergence is no reason for concern: as Haque emphasizes, the LOAC, though it prohibits killing civilians, does not create a moral permission, let alone a moral duty, to kill enemy combatants.³⁰ (In this regard, the LOAC’s authority may seem much less problematic than General’s authority in *Mistaken Soldier*.) And since the prohibition on killing civilians simply reiterates an existing moral duty unjust soldiers have, the law, one may think, cannot make things morally worse.

But this is too quick. True, the LOAC does not change that the unjust soldier has a duty not to kill either enemy combatants or civilians. But it does change the *comparative significance* that these considerations have in his deliberation. Even if the unjust soldier has a duty to target neither enemy combatants nor civilians, the question arises how, *if he will nonetheless target someone*, he ought to choose his target. It is here that the LOAC crucially deviates from the underlying moral reasons. The LOAC categorically prohibits targeting civilians, but does not similarly prohibit killing soldiers; and so an unjust soldier who is guided by the LOAC will, when choosing between targeting soldiers and targeting civilians, always do the former.

Compare this to what one would expect morality to advise. Haque holds that it is morally worse to kill enemy civilians than enemy combatants if both are innocent; and so the unjust soldier’s

³⁰ “The law of armed conflict (LOAC) does not tell combatants what they may do, only what they may not do.” *LMW*,

targeting decisions should reflect this. Like Haque's highly restrictive account of civilian liability, this is a controversial view. Some simply deny that considerations like vulnerability, defenceless, choice of risk exposure, etc. have independent moral significance. Others, more subtly, argue that the significance of these considerations applies only to just combatants: while it may be true that it is harder to justify targeting innocent civilians rather than innocent combatants, it is *not worse* to *unjustifiably* target innocent civilians rather than innocent combatants.³¹ But even if we follow Haque in accepting that killing innocent civilians is generally morally worse than killing innocent combatants, it is not plausibly *so much worse* that we must target the latter instead of the former *no matter their respective number*.³² So there will be a point at which an unjust combatant will commit a less severe wrong by targeting a small number of innocent civilians instead of many innocent soldiers.³³

This observation, in turn, suggests that there may in principle be another rule that would provide better guidance to unjust combatants than the LOAC. To minimize the wrongfulness of one's targeting decisions, a rule allowing for certain trade-offs between combatants and civilians might be superior.³⁴ So to explain why unjust combatants should follow the LOAC's directives rather than such an alternative rule (or at least to explain this within the structure of his own theory), Haque would have to argue that unjust combatants cannot reliably be guided by such an alternative rule given their ignorance, biases, and similar deficiencies.

But it is unclear why such a rule would be deficiency-incompatible. Consider, for instance, a rule that prohibits targeting civilians except if the alternative is targeting a much larger number of enemy combatants. Like the LOAC, such a rule would not require the combatants to rely on their own moral judgment of the moral liability of different people. Instead it would (again like the LOAC)

³¹ Victor Tadros, "The Moral Distinction Between Combatants and Non-Combatants: Vulnerable and Defenceless," *Law and Philosophy* 37 (2018): 289-312. See also Helen Frowe's contribution to this symposium.

³² The numbers-sensitivity of the putative asymmetry is emphasized in Yitzhak Benbaji, "The Lesser-Evil Dilemma for Sparing Civilians," *Law and Philosophy* 37 (2018): 243-267.

³³ Interestingly, Haque makes just such an observation about the harm-minimization view (*LMW*, 40). But, as we argue, his own response – that killing civilians is intrinsically morally worse than killing even innocent combatants – fails to establish that the laws of war should "strictly limit[] the more wrongful killings" (41).

³⁴ This is compatible with Haque's claim that the asymmetrical worseness of killing innocent combatants and civilians renders unjustified a targeting rule that requires combatants to simply *minimise* the number of innocents they kill.

merely require them to distinguish between combatants and civilians, as well as the relative number of each who would be targeted. Such a rule would be effective only if combatants could be motivated to draw the relevant distinction, and act on it. But the same is true with regard to the LOAC. The alternative rule would require a judgment of proportionality that is absent where the combatant simply follows the LOAC's prohibition on targeting civilians. But what reason is there for thinking that soldiers will generally go awry here, and do so specifically in ways that lead to more wrongful action? Indeed, this alternative rule seem no more demanding than Haque's interpretation of the standard *in bello* proportionality rule (which requires combatants to compare the amount of harm they will collaterally inflict on enemy civilians with the anticipated reduction in harm to combatants and civilians on their own side³⁵).

Perhaps the problem is not that unjust combatants could not apply the rule if they tried to follow it, but that even trying to follow it is incompatible with what may seem like their most fundamental deficiency: their mistaken belief in the justice of their own cause. Given this deficiency, it might be proposed, unjust combatants cannot help themselves to the benefits of rules that provide otherwise optimal guidance to them if those rules deviate from those for just combatants. And so pointing to the in-principle superiority of rules such as the ones we discussed does not show that the LOAC lacks authority for unjust combatants, because these alternative rules fail to satisfy deficiency-compatibility.

Yet this pleasingly simple objection rests on a mistaken assumption. As Haque himself helpfully observes, combatants may have a wide variety of views regarding the justice of their cause: "They may doubt that they fight for a just cause, or that war was necessary or proportionate. They may suspect that justice favors their adversary, or that justice condemns both parties. Nevertheless, they may choose to fight in order to protect their friends, families, and communities from the consequences of their government's decision to go to war. The LOAC must speak to these combatants as well, by showing them that, even in an unjust war, there are rules worth following." (48) But, by the argument sketched earlier, those combatants who in fact recognize the injustice of their cause (or even those who reasonably suspect that their cause may be unjust) will not minimize

35

their expected wrongdoing by following the LOAC rather than adopting another set of rules that does not strictly prohibit targeting civilians.³⁶

We have argued that both just and unjust combatants are less likely to have instrumental reasons to follow the LOAC's rules on targeting civilians than Haque's argument suggests. But just as crucial is a third, and final, point: it is highly doubtful that just and unjust combatants will be under the authority of the same rules if, as Haque assumes, the function of these rules is to correct combatants' epistemic deficiencies and help them minimize the wrongfulness of their targeting choices. The discussion of just and unjust soldiers has already suggested as much: while just combatants might do best following the LOAC, or perhaps to adopt a rule requiring careful discrimination between liable and non-liable civilians, unjust combatants would do best adopting a rule that requires, not discrimination among civilians, but targeting civilians only if doing so is the alternative to targeting a much greater number of enemy combatants. But in fact the problem emerges from the basic structure of Haque's argument, with its focus on overcoming combatants' epistemic deficiencies and biases. Because their respective moral situation is fundamentally different, it would seem enormously surprising if just and unjust combatants each had reason to be guided by the same rule. We are not denying that it is conceivable that the best rule that unjust combatants can be guided by, given their particular deficiencies and moral duties, and the best rule that just combatants can be guided by, given their own deficiencies and moral duties, just happen to be the same. But if this is so, it seems highly contingent given the structure of Haque's argument – much more contingent than one might have expected given how central the symmetry between just and unjust soldiers is to the LOAC.

Importantly, the difficulty of accommodating the law's symmetry is not internal to instrumental justifications as such. Instead it follows from Haque's specific articulation of the mechanism by which the LOAC improves combatants' conformity to moral reasons.

The problem could, for example, be avoided if the LOAC's role were not to guide combatants as to their respective moral duties, but rather to make combatants' choices interdependent in ways that enable both just and unjust soldiers to achieve morally superior outcomes. Haque rejects this possibility, recall, on the grounds that war is a competitive rather than cooperative endeavour,

³⁶ Note that our argument is not restricted to those combatants who are aware (or suspect) that their *overall war* is unjustified. It also applies to those who are aware (or suspect) that they are participating in an unjustified *phase* or *component* of a war.

and so one side's "moral reasons do not depend on [the other side's] behaviour." (42) But the inference from competition to moral independence is premature. It is true that, qua competitors, the warring parties are not pursuing some joint end around which they can orientate their deliberation. But cooperation so understood is not the only form of interdependence that matters. For even if competitors are not pursuing a joint end, they may be able to shape, through their actions, their opponent's choices for the better.

For instance, my adversary may (for good or bad reasons) be more likely to abstain from attacking my civilians if I abstain from attacking his, at least if my abstention is easily publicly recognizable. A scenario where I abstain from attacking his civilians, and he abstains from attacking mine, is superior to one in which we each attack the other's civilians. But this does not mean that we share an end, and act cooperatively. For this symmetrical restraint is in principle morally inferior to one in which the just side attacks liable civilians (if necessary to prevent sufficiently great harm to innocents), while the unjust side attacks no civilians. If I take my adversary to be fighting an unjust war, then my end is to prevent the killing of what I consider innocent people (primarily, but not exclusively, on my side). Yet I also think that some enemy civilians are not innocent, but liable to suffer harm. So my choice whether to kill or spare those civilians is ultimately a matter of which means most effectively advances my end. The converse applies to my opponent. So we do not share ends (my ends, as I understand them, and my enemy's ends, as he understands them, are in principle in competition). And yet our respective ends are best pursued by making our choices interdependent, such that each side contributes to the other's end as a means for realizing its own: I abstain from killing what I think are his liable civilians, so as to motivate him not to kill my civilians (whom I deem innocent); and he abstains from killing what he thinks are my liable civilians, so as to motivate me not to kill his (whom he considers innocent).

Where interdependence of this sort exists, public binding rules such as the LOAC may serve the parties, not by correcting their epistemic shortcomings or biases, but by extending their options. In particular a public binding rule may make it possible to effectively signal one's willingness to take off the table the choice to target enemy civilians, and thus help make it more likely that the other party will do the same. Our point is not here to defend such a view (though we find it quite attractive). It is simply to flag that such a view could both ground the LOAC's authority in instrumental considerations, and easily account for the symmetrical character of the LOAC's rules, in

ways that Haque's account, with its focus on the individualized epistemic deficiencies of very differently situated parties, cannot.³⁷

6. Conclusion: Authority and Justification Revisited

We have argued that an instrumental justification of the sort proposed by Haque can establish the law's authority only if it satisfies more demanding conditions than Haque recognizes; and that Haque's argument consequently fails to vindicate the LOAC's claim to authority over all (or even most) combatants in war. Given the availability of superior strategies for improving their moral performance, many combatants are not genuinely served by the LOAC's authority, at least on the grounds Haque adduces.

If one accepts this conclusion regarding the law's authority, must one also conclude that the current laws of war should be done away with, or at least condemned? No. One could defend the LOAC's authority on other grounds (such as the ones very briefly sketched at the end of Section 5). But let's assume instead that no such alternative vindication of the LOAC's authority succeeds. Even then, the laws of war we currently have may be worth keeping, and partly for reasons that flow from Haque's own argument. As we discussed in Section 2, which laws ought to be supported is distinct from the question which laws have authority, and for whom. An institution may lack authority (or, more relevant for our purposes, lack authority over at least some of its purported subjects) and yet be justified. Whether it is depends (among other things) on the institution's effects on the behaviour of its subjects.

One possibility is that an institution has genuine authority over some subjects, in virtue of optimally improving their moral performance, but its effect on the remaining subjects is sufficiently suboptimal that its authority is all-things-considered disvaluable. This disvalue gives us moral reasons

³⁷ There are obvious similarities between the view we sketch and the harm minimization view that Haque finds wanting. But Haque's objection to that view is that, *by itself*, it does not establish combatants' duty to follow the laws, since it does not consider combatants' particular moral reasons. We grant this. Our point is that we can add to this view an account of the moral reasons that individual combatants have (and take themselves to have), such that combatants have reason to follow rules that help minimize the overall harm to civilians, both on their and the other side.

(or even duties) to remove or reform the institution, despite its legitimate authority. Call this normative property *Unjustified Legitimate Authority*.

To illustrate, consider:

Double Ambulance Driver. Vicky the Radio Controller is tasked with directing two ambulance drivers to the scene of an accident, using the single emergency radio frequency that all ambulances use. The fastest route is through a busy part of downtown packed with pedestrians. Sarah the Ambulance Driver is very skilled, though not perfect, at driving fast on busy roads. If Vicky issues directives that carefully guide Sarah through the downtown streets, Sarah can make it to the accident while posing an acceptably small risk of killing a pedestrian. Jim the Ambulance Driver has average driving skills, but he is mistakenly convinced that his skills rival Sarah's. If Jim hears Vicky's directives to Sarah over the radio, he will also be guided by them and take the downtown route. If he does so, he is very likely to kill a pedestrian. Vicky directs Sarah to take the downtown route.

In this case, Sarah is morally bound to obey Vicky's directives, since they enable her to improve her conformity with her moral duties. But Vicky would not be justified in issuing those directives, since doing so would overall impose an unacceptable risk on pedestrians via their effect on Jim.

Are the laws of war unjustified in this sense? This depends on the costs of their existence. Our argument in the previous section suggests that these costs could be very significant: The law could stand in the way of just and unjust combatants following strategies (such as a limited trade-off rule) that would reduce their wrongful killing to a greater extent than they would by following the law's guidance.

But it would be too quick to conclude that the laws of war lack justification. These potential costs only count against a law to the extent that its subjects are likely to *in fact* adopt a superior instrumental strategy in the law's absence. If, however, combatants would *not* adopt a superior strategy in those circumstances, and would instead commit more serious wrongs than they would by following the law (as Haque predicts), then the existence of the law does not prevent valuable conformity. To illustrate this idea, consider again the *Lazy Ambulance Driver* case we briefly introduced at the end of Section 4: Dean the Ambulance driver can choose between driving without assistance, according to his own judgment; following the directives of Vicky the Radio Controller; or listening to soothing music on his headphones. These three options are mutually incompatible, and listed in descending

order of the risk they pose to pedestrians. Though Dean is aware of the benefits of listening to soothing music, he can't be bothered to put on his headphones. But he is willing to obey Vicky's directives. And if he does not receive her directives, Dean will drive without assistance.

As we argued earlier, the fact that Dean has available a superior strategy for conforming with his duties (using his headphones) precludes Vicky's directives having authority for him. This is true even though he *will not* adopt this strategy. However, the relevant point right now is that Vicky's lack of authority over Dean is compatible with Vicky having a straightforward justification for issuing directives to Dean. By guiding Dean, she brings about the best possible outcome relative to *her* option-set, though not relative to his.³⁸

It seems plausible that the same story could be true of the laws of war. On this picture, the LOAC provides combatants with what we earlier called a *narrow* and not a *wide* service: it improves their compliance with their moral duties relative to how they will *in fact* act, but not relative to how they *could* act given their alternative available means of achieving compliance. Though the narrow service Haque identifies fails to vindicate the law's authority, it may well still provide a service of sufficient value to justify the law's existence.

Interestingly, cases like this suggest that the relationship between justification and authority may be more complex than often thought. They show that sometimes the law lacks the moral power to impose duties that it claims, and is yet justified (at least in part) because it claims such power and (at least some of) its subjects treat its directives *as if* they imposed duties. Another way of putting this is that an institution may be morally justified, in virtue of possessing what is sometimes called *mere de facto* authority, and yet lack justified authority. De facto authority is usually taken to exist where a person or institution either claims or is acknowledged by others to have justified authority.³⁹ *Justified* de facto authority is simply de facto authority with regard to which such claims or acknowledgements are justified. It is natural to assume that such claims or acknowledgments are justified if and only if they are *correct*. Since they are correct if and only if the authority is in fact justified, it would then follow that justified de facto authority just is justified authority. Our discussion here, however, points to a

³⁸ This reflects a point made earlier: questions of law's *justification* adopt the perspective of the *legislator*, whereas questions of law's *authority* are framed from the perspective of the law's subject.

³⁹ Cf. Raz, *The Morality of Freedom*, 46.

distinct notion of justified de facto authority – where those claims are mostly false, and yet worth making – which we should include in a full moral evaluation of institutions:

Justified De Facto Authority: An individual or institution claims to possess the moral power to place others under duties by issuing commands, and (some of) those purportedly subject to its authority acknowledge it as having this power, *and it is all-things-considered valuable that the individual or institution makes those claims. In virtue of its value, individuals may have moral reasons (or even duties) to create the institution and see to it that it makes those claims, support it in making those claims, refrain from undermining its ability to make those claims, etc.*