

*Part I*

A Theory of Moral Psychology  
and International Norms

## 1 Introduction

### The Laws of War and the Puzzle of Norm Emergence

---

On January 23, 1368, Zhu Yuanzhang, a man of humble origins who rose to become the leader of a successful peasant revolt, proclaimed himself emperor of China in the eastern city of Nanjing. To consolidate his power against the remnants of the Yuan, a foreign-led dynasty that had been established by the Mongolian conqueror Kublai Khan, Zhu ordered his generals to undertake an armed expedition in the north in the first weeks of August in 1368. Depicted in later histories as “having observed with revulsion the senseless destruction of warfare and as having attempted ... to impose troop discipline in order to win the goodwill of conquered populations,”<sup>1</sup> Zhu is reported to have ordered his generals to avoid slaughtering civilians:

The people of North China have long suffered the oppression of the armed bands. Scattered far and wide, they long for one another. It was for this reason that I commanded my generals to campaign in the North in order to rescue the people from the flood and flame. The founders of the Yuan ruling house won merit in their service to others. Their descendants, however, were not moved to compassion by the suffering of the people. Heaven had quite enough of them and cast them out. The rulers were at fault... When former dynasties were overthrown, it was because they had recklessly indulged in slaughter, defied Heaven and oppressed the people. I simply cannot abide these things. When my generals subjugate cities, they shall not wantonly burn and plunder, not abandon themselves to slaughter. To the Yuan imperial clansmen and consort families, let full protection be given, in order that all may be preserved. Above, respond to the will of Heaven and below, comfort the people in their needs. This, in turn, will further my purpose of supplanting evil-doers and giving peace to the people. Those who fail to respect (my) decree shall be removed from office without hope of forgiveness.<sup>2</sup>

Having thereby established and consolidated the Ming dynasty, Zhu was portrayed by later Ming historians as a “warrior hero” who exemplified

<sup>1</sup> Mote (1988, 49).      <sup>2</sup> Quoted in Taylor (1975, 59–60).

4 A Theory of Moral Psychology and International Norms

the Mandate of Heaven – to see “as the people see,” hear “as the people hear.”<sup>3</sup> He ruled China until his death on June 24, 1398.

Ming historians almost certainly exaggerated Zhu’s benevolence and compassion. Indeed, he is reputed to have been a fairly capricious ruler and to have been quite “brutal” “in dealing with those he considered disloyal.”<sup>4</sup> Nevertheless, some modern historians have noted that early Ming depictions of him likely bear some relation to the truth. In the first years of his reign, for example, “he strove to create the image of a wise future ruler, granting tax remissions to war-ravaged regions, punishing looters among his own troops, and rewarding loyal, altruistic service to the Yuan as well as among his own followers.”<sup>5</sup> Accurate or not, Zhu’s purported order to refrain from killing civilians reflected a long-standing tradition of strategic and moral restraint in Chinese theorizing about the use of military force.<sup>6</sup> In *The Art of War*, the ancient Chinese general Sun Tzu advised military leaders to avoid destroying enemy cities, and instead to win without fighting – a perspective on war that is usually interpreted as advocating the use of deceptive stratagems, secrecy, and psychological warfare tactics. Yet, Sun Tzu’s emphasis on moral restraints is undeniable, especially in light of the fact that he viewed “wisdom” and “benevolence” as crucial traits of a successful military leader.<sup>7</sup> Other ancient Chinese treatises are even more direct in their emphasis on the ethics of restraint toward civilians. For example, the *Wei Liao-Tzu* stipulated that

In general, [when employing] the military do not attack cities that have not committed transgressions or slay men who have not committed offenses. Whoever kills people’s fathers and elder brothers, whoever profits himself with the riches and goods of other men; whoever makes slaves of the sons and daughters of other men is in all cases a brigand.<sup>8</sup>

Although some scholars suggest that the *reason* for emphasizing compassion and benevolence was strategic<sup>9</sup> – to get the people to willingly submit – the evidence also shows that political and military theorists from the earliest years of ancient Chinese history believed that it is wrong to kill civilians.

Modern international law reflects very similar concerns. Indeed, since the late nineteenth century, states have constructed an increasingly elaborate system of international law – one that is designed specifically to protect civilians in war and that takes inspiration from secular interpretations of just war theory and pre-Enlightenment theorists such as Hugo Grotius. Despite the fact that compliance with the laws of war remains

<sup>3</sup> Taylor (1975, 1); Mote (1988, 50).

<sup>4</sup> Langlois (1988, 156).

<sup>5</sup> Mote (1988, 50).

<sup>6</sup> Lo (2012a).

<sup>7</sup> Sawyer (1993, 157).

<sup>8</sup> Sawyer (1993, 254–255).

<sup>9</sup> Johnston (1995).

imperfect, at no earlier point in human history have innocent civilians enjoyed the kinds of legal protections that they do in contemporary international law. Indeed, prior to the development of the modern laws of warfare, villages and towns were systematically pillaged, burned, starved out, or decimated, and their inhabitants were beaten, raped, or summarily executed, often with full legal impunity. While civilian fatalities are still a regrettable feature of modern war, contemporary international law clearly defines these practices as war crimes, and it has developed an increasingly authoritative regime of criminal law institutions for administering international humanitarian law (IHL) in practice.<sup>10</sup>

Across vast stretches of time, space, and culture, the norms of war in early China and modern international law are surprisingly similar. In Part II, I show that something similar can be said about Islamic law, which likewise contains rules against targeting the innocent. Taking cultural variation as a given, how can we explain why human societies as disparate as ancient China, medieval Islam, and the modern West nevertheless converged on very similar rules for regulating the use of military force? Drawing on recent scholarship in moral psychology and cognitive neuroscience, in this book I argue that the laws of war – specifically those rules devoted to the protection of noncombatants – are grounded in universal moral sentiments that shape how people think about the ethics of killing in war. I show that civilian protection norms are not simply a figment of the modern West, but that they have also emerged in civilizations as culturally diverse as ancient China and the early Islamic empire. The laws of armed conflict do reflect a *growing* revulsion toward civilian deaths, but I show that this sense of revulsion was embryonic in earlier civilizations, in the East as well as in the West.

But despite their ubiquity, civilian protection norms are inherently fragile. Indeed, civilians continue to die at an alarming rate – not only in ethnic cleansing campaigns and mass killings, but also as unintended side effects of otherwise lawful military operations. In what follows, I show that their fragility lies not *just* in failures of compliance, but also in the moral beliefs and emotions that shaped the creation of the laws of war. Although the laws of war forbid states from killing civilians intentionally, they permit states to use strategies, tactics, and weapons systems that expose civilians to high risks. The laws of war are *asymmetric* in how they address intentional versus unintentional civilian casualties: they place a much higher valence on intentional and foreseen killing than they do on unintentional deaths. I argue that the moral beliefs and sentiments that shaped the creation of the laws of war paradoxically enable states to evade accountability for incidental collateral

<sup>10</sup> Jo and Simmons (2016); Sikkink (2011).

6 A Theory of Moral Psychology and International Norms

damage. While the main focus of this book is on developing a theory of how societies converge on similar norms of war, my analysis also provides critical insights into the failures and imperfections of the laws of war. In the conclusion, I argue that in order to improve the laws of war, policymakers and human rights activists, not to mention average citizens, need to exercise their capacities for empathy and perspective-taking *to modify how we think* about the ethics of killing in war. More specifically, I argue that IHL needs to be improved so that states are required to more equitably distribute the risks between their own armed forces and civilians on the ground.

Thus, I stake out two key claims in this book. First, I argue that the laws of war are rooted in universal moral sentiments that govern how people think about the ethics of killing. Most human beings have innate abilities for perspective-taking and empathic concern, and when these abilities are engaged, for example, via persuasion, deliberation, or emotionally salient events, societies converge on similar norms for protecting noncombatants in war. When political actors design norms of war for protecting civilians, they use a grammar of moral rules for how to categorize individual actions, a grammar of moral rules that reflects systematic asymmetries in how humans process information about the ethics of killing in war: intentionally shooting one civilian is a war crime, yet accidentally killing a much larger number is seen as acceptable as long as the expected military gains outweigh the civilian losses. Second, I argue that the laws of war have both *restrictive* and *permissive* effects. Although the laws of war do generate significant protections for civilians (their restrictive effects), states can exploit their permissive effects, thereby evading accountability for unintended casualties. By shaping how states define the meaning of compliance with IHL, our moral beliefs and emotions simultaneously restrict and permit the use of violence against civilians in armed conflict. To set the stage for the rest of this book, in the following section, I briefly sketch out my theory of international norms, and I explain how it advances debates on norm emergence in International Relations (IR) scholarship. I then outline the plan of the book and discuss the normative implications of the theory and historical case studies.

### **International Relations Theory and the Puzzle of Norm Emergence**

How did civilian protection rules emerge,<sup>11</sup> and why do they have the content that they do? In existing IR theory, there are three traditional

<sup>11</sup> I examine four laws of war principles: distinction, proportionality, precaution in attacks, and unnecessary suffering. The principle of distinction holds that states are required to distinguish military from civilian targets. Deliberate attacks against civilians or civilian

approaches to explaining the evolution and design of international institutions and norms. Realist and rational choice theorists argue that states create institutions to advance their own interests. While realists argue that powerful states create international institutions to secure their political and military interests, rational choice theorists argue that they design institutions to overcome collective action problems and to enforce mutually beneficial agreements.<sup>12</sup> Still others argue that states design institutions to uphold international order through strategic restraint and to signal benign intentions to other states.<sup>13</sup> Lastly, constructivists argue that international law reflects the moral values and practices of states (and certain non-state actors) – values and practices that are unique to specific historical time periods.<sup>14</sup> *In The Image Before the Weapon*, for example, Helen Kinsella writes that the “laws of war ... are consistently described as arising from a specifically European (Western) culture, and from a highly Christianized one as well.”<sup>15</sup>

While these theories offer useful insights, they have three principal limitations. First, they do not explain why societies or civilizations that are otherwise very different nevertheless converge on similar norms of warfare. Drawing on a range of historical evidence, I show that moral and legal institutions for safeguarding civilians in war are far more prevalent in human societies than many IR theorists recognize. Although the conventional wisdom holds that civilian protection norms are unique to the modern West, I show that norms for protecting the innocent in war have also emerged in non-Western societies, specifically Warring States China and early Islam. The existing paradigms in IR provide important insights into the processes whereby human societies converge on specific norms, but I argue that in order to explain their *similarity in content*, the traditional IR paradigms need to be integrated with a neuroscientific framework of moral cognition, emotion, and discourse.

Second, existing IR scholarship does not adequately account for the *content* of the laws of warfare. As Kathryn Sikkink acknowledges in *The Justice Cascade*, “the single most difficult issue to explain is why certain

objects are war crimes. The principle of proportionality holds that incidental civilian losses must not be excessive in relation to expected military gains, and the principle of precaution in attacks holds that states must take feasible precautions to minimize civilian suffering. Finally, the principle of unnecessary suffering holds that soldiers are not allowed to use weapons or tactics that cause unnecessary suffering (Kalshoven and Zegveld 2001, 22).

<sup>12</sup> Mearsheimer (1994/1995); Carr (1939); Krasner (1999); Keohane (1984); Koremenos, Lipson, and Snidal (2001); Morrow (2014).

<sup>13</sup> Ikenberry (2001); Thompson (2010); Savarese and Witt (2017).

<sup>14</sup> Wendt (1999); Bull (1977); Finnemore (1999); Reus-Smit (2004).

<sup>15</sup> Kinsella (2011, 17).

8 A Theory of Moral Psychology and International Norms

ideas at certain moments in certain places resonate, grab attention, and become possible.”<sup>16</sup> In this book, I argue that in order to fully explain why “certain ideas at certain moments” resonate and “grab attention,” IR scholars need to revise much of what they think about the nature of human cognition, emotions, and discourse. In particular, although many constructivists believe that cultural values shape moral beliefs *all the way down*,<sup>17</sup> a more compelling account of norms should leave space for evolved moral psychology. Interestingly, Sikkink explicitly mentions that research in cognitive science provides a potentially fruitful avenue for explaining the content of international norms. However, she points out that it would be “heresy” for a political scientist to argue that some moral principles are substantially innate.<sup>18</sup> In what follows, I take up this challenge by showing how shared moral beliefs and emotions shape the content of the law and ethics of war.

Third, existing scholarship does not fully explain how the laws of war constrain and enable violence against civilians. With some important exceptions,<sup>19</sup> most scholars have focused on how the laws of war shape (or fail to shape) political and military decision-making in war. In short, they have focused on the problem of compliance.<sup>20</sup> While the compliance literature has led to important insights, it overlooks some of the paradoxical effects that the law can have on military practices. In particular, the laws of armed conflict prohibit military personnel from *deliberately* killing civilians, yet they permit civilians to be killed as an *unintended side effect* of otherwise justifiable strategies, provided, of course, that civilian casualties are not excessive in relation to the direct and concrete military gains anticipated. But while prohibiting intentional attacks on civilians is important, it still allows for quite a lot of bloodshed. For example, during the US war in Vietnam, Operation Rolling Thunder – a three-year bombing campaign that was directed against North Vietnam – caused about 52,000 civilian deaths.<sup>21</sup> Although the United States did attack dual-use structures – that is, those that have both military and civilian uses – high-level officials did not order the deliberate slaughter of civilians.<sup>22</sup> Since the 52,000 civilian deaths were incidental to the Johnson administration’s overall objectives, and since the administration did not order intentional attacks on civilians, some scholars argue that

<sup>16</sup> Sikkink (2011, 261). <sup>17</sup> Wendt (1999, 113–135). <sup>18</sup> Sikkink (2011, 261).

<sup>19</sup> Sanders (2018); Hurd (2017); Zehfuss (2010); Cronin (2013); Jochnick and Normand (1994).

<sup>20</sup> Downes (2008); Valentino, Huth, and Croco (2006); Morrow (2007, 2014); Finnemore (1999).

<sup>21</sup> Pape (1996, 190).

<sup>22</sup> Bellamy (2012, 173–182); Crawford (2014b); Parks (1982); Clodfelter (1989); Traven and Holmes (2021).

Rolling Thunder was broadly compliant with the laws of war.<sup>23</sup> Had Rolling Thunder intentionally targeted even 5,000 civilians (or fewer), the United States would have thereby committed an unequivocal war crime. Similar comments could be made about more recent US operations in Iraq and Afghanistan. Examples like this show that far from simply prohibiting direct attacks on civilians, IHL sometimes allows for what Neta Crawford calls *systemic military atrocities*, collateral losses that are unintended.<sup>24</sup>

But apart from the *direct* side effects of military operations, armed conflict also has *indirect effects* on the population. The epidemiologist Paul Wise argues that when we factor in the indirect effects that armed conflict has on human societies, the unintended effects of war vastly overshadow their directly intended effects, not to mention their immediate, if still incidental, side effects. As he puts it, “most civilian casualties in war are not the result of direct exposure to bombs and bullets; they are due to the destruction of the essentials of daily living, including food, water, shelter, and health care.”<sup>25</sup> According to one report, “the indirect health consequences of civil wars between 1991 and 1997 throughout the world were twice that associated with direct, combat-related effects.”<sup>26</sup> In Wise’s view, numbers like this show that “war generates significant elevations in indirect mortality and disability above prior baselines.”<sup>27</sup> Though IHL requires states to avoid attacks on objects that are indispensable to the livelihood of civilians, the emphasis that it places on intentions means that it provides fewer legal safeguards for holding states accountable for the indirect effects of warfare.

Constructivists and critical theorists are well aware of the permissive effects of international law.<sup>28</sup> However, I argue that they have overlooked a key source of some of these permissive effects. While several IR scholars have called attention to the strategic and discursive sources of permissive international norms,<sup>29</sup> here I argue that the permissive effects of IHL are rooted in shared cognitive and emotional structures that lead people to see intentional killings as a bigger moral problem than unintentional side-effect killings. While the main goal of this book is to examine how moral beliefs and emotions shaped the *design* of the laws of war, in doing

<sup>23</sup> See Parks (1982), who suggests that the Johnson administration could have lawfully bombed a more extensive array of targets.

<sup>24</sup> Crawford (2013). <sup>25</sup> Wise (2017, 139). <sup>26</sup> Wise (2017, 143).

<sup>27</sup> Wise (2017, 144).

<sup>28</sup> Tannenwald (2007); Sanders (2018); Zehfuss (2010); Owens (2003); Buzas (2017); Jochnick and Normand (1994).

<sup>29</sup> For a strategic account, see Dixon (2017). For a more discursive account, see Owens (2003) and Tannenwald (2007).



10 A Theory of Moral Psychology and International Norms

this it also provides insights into how IHL constrains and enables violence against civilians. Using the hard case of Allied bombing policy during the Second World War, I argue in Chapter 6 that moral beliefs and emotions have an indirect effect on state policy by shaping the restrictive and permissive norms that states use to justify their decisions. I show that not only did moral beliefs about intended and unintended killings ensure that aerial bombing practices were not fully outlawed in the negotiations over The Hague Conventions of 1899 and 1907, but that states used the distinction between intended/unintended killings to rhetorically justify strategic bombing policies. In short, I argue that by shaping how states define law compliance, our shared moral belief systems can sometimes facilitate unintentional violence against noncombatants.

To resolve these problems, I defend three key claims. First, I argue that the laws of armed conflict are grounded in the cognitive–emotional wiring of the brain. Drawing on a range of recent research in cognitive neuroscience, social psychology, evolutionary psychology, and moral theory, I argue that most humans have cognitive and emotional capacities that lead them to believe that it is wrong to intentionally kill innocent people in certain circumstances. These moral intuitions lead people to create social norms that set out what is permissible and impermissible in armed conflicts. When people think about the ethics of killing in war, they use a grammar of moral rules<sup>30</sup> for how to categorize individual behaviors, and this grammar of moral rules shapes how diplomats design the rules of IHL. Social psychological processes, such as empathy, shape the emergence of civilian protection rules: critical events such as major wars or humanitarian crises induce empathy for war victims, and as a result, people are more likely to support the creation of stronger laws for protecting innocent people. While several prominent scholars have started to question the value of empathy,<sup>31</sup> I contend that as long as perspective-taking and empathy are properly targeted, they can encourage people to take the moral interests of others far more seriously than they otherwise would. Empathy is not a panacea, but it can help to expand the horizons of moral concern beyond their current limits.

Second, I argue that most people have cognitive–emotional biases that shape how they think about the ethics of killing in war. As a result, these cognitive–emotional biases influence the creation and design of IHL. One such bias is what I refer to as the *intention/side-effect* distinction. The *intention/side-effect* distinction holds that intended killings are morally

<sup>30</sup> Mikhail (2002, 2007, 2011); Dwyer (1999, 2006, 2009); Chomsky (2009).

<sup>31</sup> Bloom (2016); Prinz (2011).

worse than unintended, side-effect killings. In the chapters that follow, I explain how this distinction has significantly affected the development and application of the laws of armed conflict – in ways that not only generate restrictive effects for how states can treat civilians in war, but that also generate permissive effects that allow states to expose civilians to mass death and destruction. Because these psychological mechanisms are broadly universal rather than culture specific, they help explain how societies that are materially and culturally diverse can nevertheless converge upon moral and legal norms that are very similar. Furthermore, my theory helps explain *the content* of the laws of armed conflict by examining what makes people distinguish between intentional and unintentional harms, legitimate and illegitimate targets, and restrained and unrestrained acts of violence. My contention is that in order to explain how human societies converge on civilian protection norms, IR theorists need to build upon the more *naturalistic* theory of moral cognition and emotion that I develop here.

Finally, I argue that my theory not only helps us understand how the laws of war emerged, but it also helps us explain why they fail to adequately protect. As James Morrow has persuasively shown,<sup>32</sup> the laws of war do influence how states make decisions in war. Yet, as I mentioned earlier, they still allow for quite a bit of innocent bloodshed. Recent incidents in Syria, Iraq, and Myanmar have shown that deliberate attacks on civilians are an ongoing concern in international politics. Yet, unintended side-effect killings are a significant problem as well. Indeed, when we incorporate the indirect effects of war on healthcare resources, infant mortality rates, and so forth,<sup>33</sup> the unintended effects of armed conflict can be quite extensive. Although the principle of proportionality prohibits attacks that are *expected* to generate *excessive* civilian suffering, this still allows for high levels of incidental harm. Furthermore, the notion of *excessiveness* is inherently vague, and as a result, it is open to interpretation. As such, compliance with IHL still enables states to cause significant harm.

The main goals of this book are theoretical: to explain how civilian protection norms arise in human societies and to understand how these norms constrain and enable civilian victimization. However, my argument also has normative implications for the laws of war. In particular, I argue that when it comes to the law and ethics of killing in war, our moral sentiments are imperfect. Even though they encourage people to create international norms that protect civilians against deliberate

<sup>32</sup> Morrow (2014).      <sup>33</sup> Wise (2017).