2016

The Combatant's Stance: Autonomous Weapons on the Battlefield

Jens David Ohlin
Cornell Law School, jdo43@cornell.edu

INTERNATIONAL LAW STUDIES
Published Since 1895
U.S. NAVAL WAR COLLEGE

The Combatant's Stance: Autonomous Weapons on the Battlefield

Jens David Ohlin

92 INT'L L. STUD. 1 (2016)

Volume 92 2016

Published by the Stockton Center for the Study of International Law

The Combatant's Stance: Autonomous Weapons on the Battlefield

Jens David Ohlin[*]

CONTENTS

I. INTRODUCTION

Effective international criminal regulation of autonomous weapon systems (AWS) faces two conceptual obstacles.[1] The first is deeply philosophi-

* Associate Dean for Academic Affairs & Professor of Law, Cornell Law School; jdo43@cornell.edu.

The thoughts and opinions expressed are those of the author and not necessarily of the U.S. government, the U.S. Department of the Navy or the U.S. Naval War College.

1. For an introduction to the basic challenges raised by autonomous weapon systems (AWS), see generally Michael N. Schmitt & Jeffrey S. Thurnher, *"Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict*, 4 HARVARD NATIONAL SECURITY JOURNAL 231, 233 (2013) ("whereas some conceivable autonomous weapon systems might be prohibited as a matter of law, the use of others will be unlawful only when employed in a manner that runs contrary to the law of armed conflict's prescriptive norms governing the 'conduct of hostilities'"); Michael A. Newton, *Back to the Future: Reflections on the Advent of*

cal: when might an autonomous system be sufficiently self-aware that it could be held criminally liable as a morally responsible agent?[2] The second issue is far more urgent: When could a military commander be held criminally responsible for violations of international humanitarian law (IHL) perpetrated by an autonomous weapon system? I argue in this article that the answer to the second question is logically independent from the first.

Although the literature on AWS is new,[3] the basic framework for AWS liability was outlined, in nascent form, at Nuremberg and its aftermath. Although this sounds unlikely, a close reading of international criminal law's infancy shows that modes of liability were designed for convicting those who indirectly perpetrate war crimes through a "machine" or organized "apparatus" of power.[4] Indeed, the whole narrative of post-

---

Autonomous Weapons Systems, 47 CASE WESTERN RESERVE JOURNAL OF INTERNATIONAL LAW 5, 8 (2015).

2. In asking this question, I am referring to a fully autonomous weapon system. However, many of these issues are also raised with regard to semi-autonomous systems. Regarding the distinction between different levels of autonomy in a weapon system, see U.S. Department of Defense, DoD Directive 3000.09, Autonomy in Weapon Systems (Nov. 2, 2012), http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf. For analysis of this directive, see U.S. Department of Defense, *U.S. Department of Defense Directive on Autonomous Weapons Systems*, 107 AMERICAN JOURNAL OF INTERNATIONAL LAW 681 (2013).

3. In particular, some human rights organizations have strongly criticized the development of AWS on the basis that these systems will increase—rather than decrease—the number of violations of IHL. *See* HUMAN RIGHTS WATCH, LOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS (2012), https://www.hrw.org/report/2012/11/19/ losing-humanity/case-against-killer-robots. Although this is an important debate, my argument in this article does not depend on either the acceptance or rejection of this critique. *See also* Mary Ellen O'Connell, *Banning Autonomous Killing: The Legal and Ethical Requirement That Humans Make Near-Time Lethal Decisions*, *in* THE AMERICAN WAY OF BOMBING: CHANGING ETHICAL AND LEGAL NORMS, FROM FLYING FORTRESSES TO DRONES 224 (Matthew Evangelista & Henry Shue eds., 2014); Peter Asaro, *On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making*, 94 INTERNATIONAL REVIEW OF THE RED CROSS 687 (2012); Noel E. Sharkey, *The Evitability of Autonomous Robot Warfare*, 94 INTERNATIONAL REVIEW OF THE RED CROSS 787 (2012); WILLIAM H. BOOTHBY, CONFLICT LAW: THE INFLUENCE OF NEW WEAPONS TECHNOLOGY, HUMAN RIGHTS AND EMERGING ACTORS 148 (2014).

4. Criminal liability is not the only mode of responsibility that could be applied to AWS. Some scholars have argued that private remedies in tort law may be especially helpful in regulating AWS. *See, e.g.*, Rebecca Crootof, *War, Responsibility, and Killer Robots*, 40 NORTH CAROLINA JOURNAL OF INTERNATIONAL LAW & COMMERCIAL REGULATION 909, 932 (2015) (arguing that "for example, who will be held liable for a war crime committed by an autonomous weapon system? The weapon system itself? Its deployer? His or her commander? Its programmer? Its manufacturer? Will this be a question of international

Nuremberg international criminal law involves individuals who were mere "cogs in a machine."[5] Although this language was once deployed as a metaphor to refer to human "machine-like" organizations, the technical requirements for liability under these legal doctrines map on surprisingly well to the indirect perpetration of war crimes through an AWS. In particular, international criminal law, following several domestic systems, abandoned the "innocent instrumentality rule," thus paving the way for prosecuting individuals for indirectly perpetrating an international crime even if the "instrument" of their criminality is also a morally culpable agent.[6] This is an important development because I also argue in this article that combatants on the battlefield would be required by the demands of behavior interpretation to approach a sophisticated AWS with what I call the "combatant's stance"—the *ascription* of mental states required to understand the system's strategic behavior on the battlefield. However, military commanders can be held responsible for perpetrating war crimes through an AWS *regardless* of the moral status of the AWS as a culpable or non-culpable agent. In other words, a military commander can be liable for the acts of the machine independent of what conclusions we draw from the fact that combatants—even artificial ones—must approach each other with the combatant's stance.

However, there is still one jurisprudential area where international criminal law is ill-suited to prosecuting AWS cases, and that involves the mental state of recklessness. Many AWS cases will involve commanders who are reckless in deploying an AWS that launches attacks that violate the core prohibitions of IHL. Unfortunately, international criminal law's treatment of crimes of recklessness remains wholly inadequate, mostly because there is no international equivalent to manslaughter or a similar crime that meets any reasonable standard of fair labeling. These problems will need to be solved before international criminal law can generate a responsible legal regime for holding military commanders responsible for recklessly perpetrating a war crime through an autonomous weapon system.

---

criminal law, or rather some form of international products liability tort?"); Rebecca Crootof, *War Torts: Accountability for Autonomous Weapons*, UNIVERSITY OF PENNSYLVANIA LAW REVIEW (forthcoming) (proposing that war torts would fill the gap left over from criminal regulation because international criminal law is often unwilling to impose liability based on mere negligence). However, the present article will focus on forms of criminal responsibility.

5. *See infra* notes 23–38 and related text.

6. *See infra* notes 43–53 and related text.

II. "COGS IN THE MACHINE" AND OTHER METAPHORS

During the Nuremberg era, it was not uncommon for lawyers to talk of a machine-like organization that collectively perpetrated an international crime. Consider the *Borkum Island* case, which involved a U.S. warplane (Flying Fortress) that was shot down over the German island of Borkum in 1944.[7] The crewmembers that survived the crash were taken prisoner by the German army and were forced to march through local streets as they suffered various beatings and other forms of mistreatment.[8] Members of the German military perpetrated some of the beatings, but local civilians perpetrated other beatings after their mayor urged his constituents to attack the prisoners. The entire incident ended with the prisoners being shot and killed at the local city hall.[9]

Several of the perpetrators were prosecuted for their involvement in the crime before a U.S. military court that was convened right after the war. In a description that would later become influential for the subsequent development of international criminal law, the prosecutor in the case referred to the perpetrators as "cogs in the wheel of common design, all equally important, each cog doing the part assigned to it. And the wheel of wholesale murder could not turn without all the cogs."[10] At least part of the impetus for using the machine metaphor was that the prosecutor was unable to assign individualized blame for particular results, since so many individuals had participated in the beatings and mistreatment that eventually resulted in the final murders of the airmen.[11] However, in addition to this evidentiary issue, there was also a greater inspiration: the idea of collective action,

---

7. *See* United States v. Goebell et al. (*Borkum Island*), U.S. Military Commission, Case No 12–489 (1946) (microformed on 1–6 Records of United States Army War Crimes Trials, M1103 Rolls 1–7, National Archives Microfilm Publications (1980)).

8. For an extensive discussion of the case, see Robert Charles Clarke, *Return to Borkum Island: Extended Joint Criminal Enterprise Responsibility in the Wake of World War II*, 9 JOURNAL OF INTERNATIONAL CRIMINAL JUSTICE 839 (2011).

9. *Id.*

10. This quote from *Borkum Island* was cited with approval in Prosecutor v. Tadić, Case No. IT-94-1-A, Appeals Chamber Judgment, ¶ 210 (Int'l Crim. Trib. for the former Yugoslavia July 15, 1999) [hereinafter *Tadić* Appeal Judgment].

11. *Id. See also id.*, ¶ 208, discussing the same point with reference to the *Essen Lynching Case* (Trial of Erich Heyer and six others) before the British Military Court for the Trial of War Criminals.

i.e. that multiple individuals might combine together to form a group agent capable of collectively carrying out truly horrific crimes.[12]

Indeed, this phenomenon was part of the entire psychology of the Holocaust. While some members of the German government (including the High Command) were directly involved in all aspects of the Holocaust, many of the individual perpetrators were only involved in discrete and isolated parts of the enterprise.[13] Perhaps this was essential to their cooperation in the endeavor. Although they knew that what they were doing was wrong, they rationalized their behavior with the belief that they were only responsible for the small contribution that they were making to the effort, such as being a guard in a concentration camp or perpetrating some other persecution against Jews. However, the aggregation of these individual acts of criminality combined together to produce a massive criminal enterprise that collectively perpetrated the Holocaust. Moreover, the collective endeavor was no act of mere parallelism. It was coordinated to devastating effect by higher command officials who designed the larger "system" so as to maximize efficiency through a division of labor.

It is precisely for these reasons that the Holocaust and other international crimes are often referred to with the language of a system or "machine."[14] The metaphor shows up in other cases as well. For example, in the *Ponzano* case, the military prosecutor noted that the defendant was responsible under common law principles for setting in motion a chain of events that were the natural and probable consequences of his actions.[15] Specifically, the argument was that "if these men . . . set the machinery in

12. For a discussion of this issue, see ANTONIO CASSESE, INTERNATIONAL CRIMINAL LAW 202 (2d ed. 2008) (discussing the need to "protect society against persons who (i) band together to take part in criminal enterprises; and (ii) while not sharing the criminal intent of those participants who intend to commit more serious crimes outside the common enterprise, nevertheless are aware that such crimes may be committed; and (iii) do not oppose or prevent them").

13. On division of labor in criminal endeavors, see generally MICHAEL BOHLANDER, PRINCIPLES OF GERMAN CRIMINAL LAW 163 (2009); ROBERT CRYER, HÅKAN FRIMAN, DARRYL ROBINSON, & ELIZABETH WILMSHURST, AN INTRODUCTION TO INTERNATIONAL CRIMINAL LAW AND PROCEDURE (3d ed. 2014).

14. Mark Osiel, *The Banality of Good: Aligning Incentives Against Mass Atrocity*, 105 COLUMBIA LAW REVIEW 1751, 1831 (2005).

15. *See Tadić* Appeal Judgment, *supra* note 10, ¶ 199 (citing *Ponzano* as evidence of common design liability under the new doctrine of Joint Criminal Enterprise, although the language of the prosecutor's statement is equally consistent with holding the defendant responsible as an accomplice under the natural and probable consequences doctrine).

motion by which the four men were shot, then they are guilty of the crime of killing these men."[16]

The International Criminal Tribunal for the former Yugoslavia (ICTY) viewed these precedents as precursors to a doctrine that it dubbed Joint Criminal Enterprise (JCE).[17] The basic idea behind the doctrine was that defendants could be convicted for their participation in a machine-like endeavor, even if the defendant was not the physical perpetrator of the criminal act.[18] The machine-like nature of the criminal syndicate—cogs working together to produce a machine-like result—was most evident in the second form of the doctrine known as JCE II.[19] In that version of the doctrine, a defendant could be convicted for any crimes committed in a concentration camp, based on the premise that the camp guard (or other soldier or employee) willingly participated in an organized system of ill treatment. The machine metaphor is ubiquitous in JCE II cases, with the courts consistently rejecting the suggestion that concentration camp guards are mere cogs in the machine.[20] Rather, the judges of the ICTY concluded that cogs in the machine were just as culpable as anyone else for the criminality produced by the machine.[21]

III. INDIRECT PERPETRATION THROUGH MACHINE-LIKE ORGANIZATIONS

The notion of machine liability was brought to its natural conclusion in the *Eichmann* case.[22] The Israeli court rejected the notion that Eichmann was

16. *Id.*

17. *Id.*, ¶ 220.

18. *See* CASSESE, *supra* note 12, at 192–93.

19. *See Tadić* Appeal Judgment, *supra* note 10, ¶ 202; CASSESE, *supra* note 12, at 195–99.

20. *See, e.g.*, M. CHERIF BASSIOUNI, INTRODUCTION TO INTERNATIONAL CRIMINAL LAW 400 (2d rev. ed. 2013) ("loose talk about such inferiors as gears or cogs always seemed to imply that the law must treat them as 'blameless instrument of an alien will'").

21. *See* Prosecutor v. Vasiljevic, Case No. IT-98-32-T, Trial Chamber Judgment (Int'l Crim. Trib. for the former Yugoslavia Nov. 29, 2002), ¶ 67 (holding that "[i]f the agreed crime is committed by one or other of the participants in a joint criminal enterprise such as has already been discussed, all of the participants in that enterprise are equally guilty of the crime regardless of the part played by each in its commission").

22. Adolf Eichmann was detained by Israeli agents in Argentina and transported to Israel where he was placed on trial in an Israeli court for being the architect of the implementation of the so-called Final Solution—the Nazi policy of killing all remaining Jews as quickly as possible. Officially the criminal allegations were charged as war crimes, crimes

himself a "cog in the machine," and concluded instead that Eichmann was the driver of the entire enterprise who had deployed and "propelled" the machine in order to perpetrate the crimes.[23] The large bureaucracy at Eichmann's disposal became a machine-like enterprise that he deployed in order to implement the final solution; as a result, he was convicted of crimes against humanity.[24] At the time of the *Eichmann* trial, the notion of deploying a social machine in service of mass atrocity was mostly a metaphor, but it would soon move from metaphor into reality.

The *Eichmann* trial proved a major inspiration for Claus Roxin, who saw in Eichmann's crimes a general structure for the commission of mass atrocity.[25] Roxin was particularity influenced by Eichmann's deployment of massive bureaucratic machinery under his control. This fit quite nicely with Roxin's Control Theory, which asserts that the perpetrator with ultimate control over the crime should be labeled as the principal perpetrator (as opposed to a mere accomplice).[26] If the perpetrator shares control over the crime with other perpetrators, they may be labeled as co-perpetrators by virtue of their shared or joint control over the crime.[27] In that case, the multiple individuals enjoy a form of joint hegemony over the act because the crime cannot be committed without their joint cooperation in a mutual endeavor.[28] This is classic co-perpetration.

Similarly, the Control Theory also entails that individuals can be deemed responsible as indirect perpetrators if they use an instrument under

---

against humanity and crimes against the Jewish people. After conviction, he was executed in 1962 as the only death sentence carried out by the Israeli judicial system. Attorney-General of the Government of Israel v. Adolf Eichmann, *reprinted in* 36 INTERNATIONAL LAW REPORTS 5, 331 (1968) [hereinafter *Eichmann*].

23. *See id.*. For an extensive discussion, see *Tadić* Appeal Judgment, *supra* note 12, ¶ 265.

24. *Eichmann*, *supra* note 22.

25. *See* HECTOR OLASOLO, THE CRIMINAL RESPONSIBILITY OF SENIOR POLITICAL AND MILITARY LEADERS AS PRINCIPALS TO INTERNATIONAL CRIMES 20 (2010).

26. In German criminal law, unlike in U.S. criminal law, the decision to apply the principal and accomplice labels is very significant, since accomplices are entitled by statute to mitigation in punishment in the form of a percentage reduction compared with the sentence for a principal perpetrator. This contrasts markedly with U.S. criminal law, where most jurisdictions allow judges the discretion to give principals and accomplices the same sentence.

27. *See generally* CLAUS ROXIN, TÄTERSCHAFT UND TATHERRSCHAFT (8th ed. 2006) (concluding that the perpetrator behind the scenes has ultimate hegemony over the act and should be classified as the principal perpetrator).

28. *See* GEORGE P. FLETCHER, RETHINKING CRIMINAL LAW 655 (2d ed. 2000).

their disposal to commit the crime. In these cases, although the perpetrator does not perform the criminal act (or *actus reus*), he is nonetheless responsible because he *controls* the individual or individuals who perform the criminal act. In that sense, the perpetrator "uses" the physical perpetrator as an instrument under his direction.[29] If two co-perpetrators both "use" an instrument in this manner, they would be responsible as indirect co-perpetrators for perpetrating the crime through an instrument.[30] The key issue here is that the instrument is another person, although the doctrine urges that we assign responsibility in just the same way as if the instrument were a physical thing, like a gun or a knife.

After the *Eichmann* trial, Roxin was motivated to amend his Control Theory to include a new variant of indirect perpetration referred to as *Organisationsherrschaft*, or indirect perpetration through an organized apparatus of power.[31] The idea was a specific mode of liability for cases involving the deployment of an organization, or machine-like entity, as the instrument of criminality (as opposed to a single individual, such as in garden-variety cases of indirect perpetration). The reason for a separate mode of liability is that the organization has *parts*, each of which must work together in order to accomplish the particular result, with the entire entity working under the direction of the *Hintermann*, the mastermind behind the scenes.[32]

Although this sounds far afield from a commander's liability for deploying an AWS, the requirements of the *Organisationsherrschaft* doctrine map on well to the case we are considering. The similarity stems from the fact that the organization is viewed as a bureaucratic machine. Specifically, Roxin's doctrine required that the defendant's order be carried out by the organization as a matter of course. In other words, the organization is the

---

29. *See* BOHLANDER, *supra* note 13, at 156 (cataloguing the various scenarios under which a perpetrator can commit an offense through another).

30. This mode of liability is critically evaluated in Jens David Ohlin, *Second-Order Linking Principles: Combining Vertical and Horizontal Modes of Liability*, 25 LEIDEN JOURNAL OF INTERNATIONAL LAW 771, 777–84 (2012).

31. *See* Claus Roxin, *Crimes as Part of Organized Power Structures*, 9 JOURNAL OF INTERNATIONAL CRIMINAL JUSTICE 193 (B. Cooper trans., 2011) (translating GOLTDAMMER'S ARCHIV FÜR STRAFRECHT 193 (1963)). For recent analysis, see also NEHA JAIN, PERPETRATORS AND ACCESSORIES IN INTERNATIONAL CRIMINAL LAW: INDIVIDUAL MODES OF RESPONSIBILITY FOR COLLECTIVE CRIMES 125–140 (2014).

32. BOHLANDER, *supra* note 13, at 159; MICHAEL BOHLANDER, THE GERMAN CRIMINAL CODE: A MODERN ENGLISH TRANSLATION 90 (2008); Kai Ambos, *Article 25 Individual Criminal Responsibility*, *in* COMMENTARY ON THE ROME STATUTE OF THE INTERNATIONAL CRIMINAL COURT 743, 752 n.68 (Otto Triffterer ed., 1999).

"through element" by which the defendant carries out the crime, via a form of automaticity.[33] Second, the doctrine also requires that the individual persons of the organization are essentiality fungible (*Fungibilität*) and would be automatically replaced if they were unable or unwilling to perform their function within the organization.[34] These requirements were necessary to view the organization as a complete subordinate to the *Hintermann*, as opposed to a collaborator working *alongside* the defendant. The idea here is that the *Hintermann* has complete control over the crime and uses the organization to physically perpetrate the crime, thus making it legitimate to view the *Hintermann* as the sole principal perpetrator of the crimes.

Although the target of this doctrine is a human, bureaucratic machine, it could very well describe an actual machine. The only change required to reorient the doctrine in this way is to shift the metaphorical language of machine to a literal case of machine liability. The doctrinal requirements would be the same.[35] The military commander would be responsible for the machine's actions if the machine were to carry out the orders of the commander as a matter of course. Furthermore, the machine's pieces are, by definition, mere fungible parts, capable of replacement and having no independent existence. In other words, the weapon system as a whole works under the control of the military commander. In a sense, this move is unsurprising since the normative argument for the original criminal law doctrine was based on the symmetry between physical instruments and organizational instruments; all that is required now is to reverse the original comparison back to physical machines again.

One might object that an AWS is more than just a machine; in theory it could operate independently, make its own decisions, and engage in target selection on its own.[36] Indeed, this is precisely the anxiety about regulating

---

33. Jens David Ohlin, Elies van Sliedregt & Thomas Weigend, *Assessing the Control Theory*, 26 LEIDEN JOURNAL OF INTERNATIONAL LAW 725, 738 (2013).

34. Jain, *supra* note 31, at 130. *See also* CARL ERIK HERLITZ, PARTIES TO A CRIME AND THE NOTION OF A COMPLICITY OBJECT 278 (1992).

35. For a good discussion of individual criminal responsibility for acts committed by an AWS, see Markus Wagner, *The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems*, 47 VANDERBILT JOURNAL OF TRANSNATIONAL LAW 1371, 1399 (2014).

36. For a discussion of these issues, see Kenneth Anderson & Matthew Waxman, *Law and Ethics for Robot Soldiers*, POLICY REVIEW (Dec. 1, 2012), http://www.hoover.org/ research/law-and-ethics-robot-soldiers.

AWS of the future.[37] This might sound like a problem for the doctrine since it requires automatic execution of the demands of the *Hintermann*. In AWS, the system is capable of exercising its own independent judgment.[38]

This is entirely consistent with *Organisationsherrschaft*, and indeed the doctrine even *requires* this. The idea in the doctrine is that the organization is a machine-like entity beholden to the whim of the *Hintermann*, although the organization usually enjoys substantial *discretion* over how to carry out its tasks.[39] Indeed, that is the whole point of deploying an organization to commit an international crime; one individual would be incapable of exercising that level of micro-control. Rather, the level of control involved in *Organisationsherrschaft* is far more macro—it is the ability to decide whether the crime will occur, not necessarily decide every aspect of *how* the crime will occur.[40] Each part of the bureaucracy will carry out its appointed task with the requisite level of discretion.

The advent of AWS creates the possibility that an AWS would have precisely the same level of discretion as the human organization under the power of the *Hintermann*. Indeed, the AWS might have the discretion to find appropriate targets and decide for itself whether the target meets certain operational, strategic and legal criteria for engagement.[41] For example,

---

37. At least some scholars believe that automaticity itself does not entail that a system will necessarily act in ways contrary to IHL. *See, e.g.*, Rebecca Crootof, *The Killer Robots Are Here: Legal and Policy Implications*, 36 CARDOZO LAW REVIEW 1837, 1892 (2015) ("There is nothing intrinsic to autonomous weapon systems that would cause superfluous injury or unnecessary suffering; a bullet fired by an autonomous sentry robot causes the same amount of injury as one fired by a human sentry.").

38. *Id.* at 1894.

39. On this point, see KAI AMBOS, TREATISE ON INTERNATIONAL CRIMINAL LAW, VOLUME 1: FOUNDATIONS AND GENERAL PART 115 (2013) (discussing, in the context of prosecutions of the Argentinian junta, the discretion that was given to the organizational apparatus of power).

40. It is unclear if this is a coherent distinction because it presupposes a definition of the "crime" that spans possible words. In other words, one can imagine all sorts of ways in which a crime might have been committed in a different way and it is mostly arbitrary to say that at some point those changes requires us to label it as a different crime.

41. Corn correctly notes that even regular soldiers have this autonomy. *See* Geoffrey S. Corn, *Autonomous Weapons Systems: Managing the Inevitability of "Taking the Man out of the Loop,"* 11 (June 14, 2014) (unpublished manuscript), http://papers.ssrn.com/sol3/ papers.cfm?abstract_id=2450640 (observing that "[t]his re-conception should begin by acknowledging an undeniable reality: that soldiers are themselves autonomous weapon systems. The soldier, like the predicted autonomous weapon system, is capable of exercising cognitive reasoning. This is obviously inherent in any human being. But the soldier is not an 'autonomous actor' in the sense that she may exercise judgment with no parameters.

the AWS might be told to look for enemy military assets and destroy them when they are located, assuming that firing on them would be an efficient use of resources (a strategic consideration) and also assuming that firing on them would be consistent with basic prohibitions of IHL (a legal and ethical consideration). This level of discretion brings the AWS closer to, rather than further away from, the operation of a human organization deployed under the command of a single perpetrator.[42]

IV. THE INNOCENT INSTRUMENTALITY RULE

Before Roxin's *Organisationsherrschaft* doctrine could be used as an effective tool in mass atrocity prosecutions, a substantial change in the doctrine was necessary. Originally, many domestic jurisdictions required that an instrument be "non-culpable" or "innocent" in order for someone else to be convicted as an indirect perpetrator of the actions performed by the instrument.[43] For example, indirect perpetration was allowed in cases where the defendant manipulated an insane individual into committing a crime at the behest of the defendant. Normally, common law lawyers would evaluate the defendant as an accomplice since he did not perform the physical act of the killing. This result was not particularly problematic since U.S. law allows accomplices to be treated as severely as principals.[44] However, German criminal lawyers, with their characteristic mitigation for accomplices, would chafe at this particular result, so the preferred result in German criminal law was to view the defendant as the indirect perpetrator of the act, with the insane individual as a mere instrument at his disposal. Likewise, a defendant could be convicted as an indirect perpetrator for manipulating a child into committing a crime for the benefit of the defendant; the

---

Instead, the soldier operates as an agent of responsible command, and in that capacity must frame her decision-making process within the parameters established by superior command. How the soldier is developed and prepared to exercise this inherently autonomous cognitive capacity without becoming an autonomous actor therefore provides a logical template for the 'preparation' of a weapon system with autonomous cognitive capacity. The goal must ultimately be to ensure the autonomous weapon functions in a manner that, like the soldier, is subordinated to the will and parameters imposed by responsible command. The ability to employ combat power consistent with LOAC obligations is inherent in that superior/subordinate relationship.").

42. *Id.*

43. *See* Marcelo Ferrante, *Argentina*, *in* THE HANDBOOK OF COMPARATIVE CRIMINAL LAW 12, 32 (Kevin Jon Heller & Markus Dubber eds., 2011).

44. *See* 18 U.S.C. § 2(a).

child, under the control of the defendant, was the mere instrument of criminality.[45]

The limiting principle for this doctrine was that the instrument had to be non-culpable—in other words, functionally equivalent to a *real* instrument. Just as one would hold a defendant responsible as a principal for using a gun to kill a victim, so too one should hold the defendant responsible as a principal for using an *individual,* just as long as that individual was not a free and responsible agent. If the agent was laboring under some defect of reason so as to make him non-culpable, then and only then could the indirect perpetration theory be applied against the defendant. However, the defendant need not be completely incapacitated, but only non-culpable insofar as the crime is concerned. It would, therefore, be appropriate to use the doctrine if the defendant manipulated a *sane* individual so completely that he or she was laboring under a mistake of fact and non-culpable for that reason.[46]

However, applying the innocent instrumentality rule would be a major barrier to applying indirect perpetration in the *Eichmann* case. While Eichmann was certainly the most culpable individual within his bureaucratic department, he was not the *only* culpable individual. Those who served underneath him were culpable in their own way for carrying out individual acts of murder, mistreatment and persecution against Jews. If convicting Eichmann under the doctrine of indirect perpetration required *denying* the culpability of those who executed his commands, then that would be too high a jurisprudential price. The soldiers on the ground were neither insane, nor acting under duress, nor mistake of fact. They were still culpable, though less culpable than Eichmann himself.

Over time, the innocent instrumentality rule was dropped in some jurisdictions.[47] This was not so much a feature of the *Eichmann* case, but rather a general anxiety over the requirement in all cases.[48] Scholars and

---

45. *See* BOHLANDER, *supra* note 13, at 156.

46. *See* ELIES VAN SLIEDREGT, INDIVIDUAL CRIMINAL RESPONSIBILITY IN INTERNATIONAL LAW 95 (2012).

47. For example, the rule was abandoned in Germany in the *Katzenkönig Case*, 4 StR 352/88, BGHSt. 35, 347 (German Federal Court of Justice, 1988). Obviously, in jurisdictions that retain the rule, the doctrine of indirect perpetration has a much narrower application.

48. For a discussion, see AMBOS, *supra* note 39, at 160. *See also* AVITUS A. AGBOR, INSTIGATION TO CRIMES AGAINST HUMANITY 11 n.1 (2013); Kai Ambos, *The* Fujimori *Judgment: A President's Responsibility for Crimes Against Humanity as Indirect Perpetrator by Virtue*

courts both agreed that a defendant could exercise sufficient control over an individual or organization—as an instrument of criminality—even if the physical perpetrators were still culpable for their actions. Of course, this required a delicate dance. If the defendant truly *controlled* the crime, then it would seem safe to assume that the physical perpetrator was not in control. If the physical perpetrator was not in control, then he could not be responsible for his actions as a free moral agent. The answer to this quandary is that "control" for purposes of the Control Theory is not the type of control that crowds out all other forms of responsibility, which can be based on lower degrees of authority with regard to the physical perpetrators.[49]

When indirect perpetration was incorporated into the law of the International Criminal Court (ICC), where it has been used to prosecute defendants such as Thomas Lubanga of the Democratic Republic of the Congo,[50] and underlies the indictment against President al-Bashir of Sudan,[51] the innocent instrumentality rule had already been excised from the doctrine. This was essential for its widespread adoption as the preeminent mode of liability for collective criminality.[52] International courts are generally tasked with prosecuting higher-level politicians and military planners who are considered most responsible for wartime misconduct. That being said, international prosecutions must not "crowd out" alternative mechanisms for accountability at the local level, at least some of which might target street-level perpetrators who commit crimes as instruments of the *Hintermann*. Indeed, the doctrine of positive complementarity assumes (and even celebrates) local systems of justice that carry out such prosecutions, even in a world with a functioning International Criminal Court.[53] If the

of an Organized Power Apparatus, 9 JOURNAL OF INTERNATIONAL CRIMINAL JUSTICE 137, 147 (2011).

49. For a discussion, see HANS VEST, VÖLKERRECHTSVERBRECHER 375 (2011).

50. *See* Prosecutor v. Thomas Lubanga Dyilo, Case No. ICC-01/04-01/06, Decision on the Confirmation of the Charges, ¶ 340 (Jan. 29, 2007).

51. *See* Prosecutor v. Omar Hassan Ahmad Al Bashir, Decision on Prosecution's Application for a Warrant of Arrest against Omar Hassan Ahmad Al Bashir, Case No. ICC-02/05-01/09, ¶ 216 (Mar. 4, 2009). For analysis, see Florian Jessberger & Julia Geneuss, *On the Application of a Theory of Indirect Perpetration in Al Bashir: German Doctrine at The Hague?*, 6 JOURNAL OF INTERNATIONAL CRIMINAL JUSTICE 853 (2008).

52. This decision was codified in Article 25(3)(a) of the Rome Statute, which states that a person shall be convicted if he "commits such a crime, whether as an individual, jointly with another of through another person, *regardless of whether that other person is criminally responsible*." Rome Statute of the International Criminal Court, art. 25, July 17, 1998, 2187 U.N.T.S. 90 [hereinafter Rome Statute] (emphasis added).

53. *See* SARAH M. H. NOUWEN, COMPLEMENTARITY IN THE LINE OF FIRE 109 (2013).

doctrines applied by the ICC were to imply that the street-level perpetrators were not responsible for their actions, this would be a deeply troubling and counter-intuitive result for international justice.

To summarize the current state of affairs: early on, international criminal justice used the metaphor of the machine to describe mass criminality in wartime. To operationalize that insight into doctrinal reality, contemporary international courts now prosecute higher-level political leaders as indirect perpetrators when they deploy an organized and hierarchical apparatus of power as an instrument of criminality. Furthermore, these prosecutions are not barred by the fact that the individuals who make up the relevant organizations are themselves culpable for their participation in the criminality. I conclude that this doctrine provides the basic framework for prosecuting a military commander or other political leader for deploying an AWS that perpetrates a war crime.

V. THE COMBATANT'S STANCE

There are, of course, grave philosophical questions over the status of AWS, which could be technologically simple or phenomenologically quite complex, depending on the nature of the system. Various proposals abound in the scholarly literature for evaluating the status of artificial entities as rational agents. The Turing Test is one obvious example, although it assumes the existence of a conversational agent, which may not be applicable here.[54] In "Computing Machinery and Intelligence," Turing imagined an evaluator having a conversation, via written messages, with two interlocutors.[55] One of the interlocutors was a programmed computer while the other interlocutor was a human being. Turing suggested that if, after a period of conversation, the evaluator was unable to determine which interlocutor was the computer and which was the real human being, then we should conclude that the computer was a "thinking" machine, i.e. an artificial form of consciousness.[56]

The test is philosophically controversial; it seems to equate behavior with mental states, or at the very least to treat sophisticated behavior as a crude proxy for deeper mental states.[57] While it is one thing to treat com-

54. *See* DAVID BRADDON-MITCHELL & FRANK JACKSON, PHILOSOPHY OF MIND AND COGNITION 109 (1996).

55. *See* Alan Turing, *Computing Machinery and Intelligence*, 59 MIND 433–60 (1950).

56. *Id.*

57. For criticism, see JOHN SEARLE, MINDS, BRAINS AND SCIENCE (1984).

plex behavior as a proxy for cognition, it is quite another to treat it as a *reliable* proxy for consciousness. It would seem that not every machine capable of satisfying the Turing Test would have something approaching consciousness.

In any event, the Turing Test by its own terms is inappropriate for evaluating an AWS, since it presumes the existence of a putative conversational agent.[58] An AWS probably would not be configured in this way. But this objection can be quickly dispensed with. One could modify the Turing Test in non-conversational terms, such that the artificial agent would qualify for personhood if its behavior *simpliciter* (as opposed to linguistic behavior) were virtually indistinguishable from the behavior of a natural human being. One could then further modify the Turing Test to apply it to AWS. The original Turing Test was based on the idea that the artificial computer could engage in a game—whether a linguistic game or a chess game—in such a complex way that its behavior was indistinguishable from that of a natural person.

This was, in a sense, a version of pragmatism: if the artificial being is truly indistinguishable from a natural person, how could one do anything but treat it as a natural person and approach it with Dennett's intentional stance or Pettit's conversational stance?[59] Indeed, the whole point of behavior interpretation is that another conversational agent would be required to attribute beliefs and desires to the agent in order to understand its behavior in a meaningful way.[60] If that situation comes to pass, then the conversational agent has, out of necessity, made the decision already. By the necessity of the situation, the artificial agent must be treated as a ration-

58. On the notion of a conversational agent, see PHILIP PETTIT, THE COMMON MIND 354 (1996) (discussing the conversational stance, an extension of Dennett's intentional stance).

59. *See* DANIEL DENNETT, BRAINSTORMS: PHILOSOPHICAL ESSAYS ON MIND AND PSYCHOLOGY 6 (1981) ("Lingering doubts about whether the chess-playing computer *really* has beliefs and desires are misplaced; for the definition of intentional systems I have given does not say that intentional systems *really* have beliefs and desires, but that one can explain and predict their behavior by *ascribing* beliefs and desires to them, and whether one calls what one ascribes to the computer beliefs or belief-analogues or information complexes or intentional whatnots makes no difference to the nature of the calculation one makes on the basis of the ascriptions.").

60. One way of putting the point is that the ascription of beliefs and desires to the artificial system would be part and parcel with Davidson's Principle of Charity, though perhaps not strictly required by it. *See* Donald Davidson, *Radical Interpretation*, 27 DIALECTICA 314–28 (1973); SIMON EVNINE, DONALD DAVIDSON 108 (1991).

al being (by ascribing beliefs and desires to it) because there is no other way of interacting with the agent and making sense of its behavior.

The same argument could be applied to an AWS in battle. The idea of a Turing Test for Combatancy would be that the non-linguistic behavior of the AWS—*acting at a distance*—would be functionally indistinguishable from any other combatant engaged in an armed conflict.[61] This would mean that the AWS does everything that any other combatant does: engage enemy targets, attempt to destroy them, attempt as best as possible to comply with the core demands of IHL (if it is programmed to obey them) and most likely prioritize force protection over enemy civilians. In this situation, an enemy combatant would be unable to distinguish the AWS from a natural human combatant. More importantly, the enemy combatant would be forced to *interact* with the AWS in just the same way that it engages with natural enemy combatants: it would have to destroy it, avoid it, or surrender to it and hope that it complies with the prohibition against killing surrendered combatants.[62] This would raise the same type of anxiety as the Turing Test: one could imagine an elaborately programmed device that was merely copying intelligent human behavior without actually *engaging* in intelligent behavior. The *appearance* of independent thought might not be the same thing as independent thought. In this situation, though, would any *deeper* question regarding the intentional states of the AWS really matter? As a matter of practical rationality, nothing would change.

Consequently, the standard for rational belligerency is whether an opposing combatant views the AWS as virtually indistinguishable from any other combatant, not in the sense of being physically indistinguishable (which is absurd), but rather *functionally* indistinguishable in the sense that the combatant is required to attribute beliefs and desires and other intentional states to the AWS in order to understand the entity and interact with it—not so much as a conversational agent but to interact with the AWS as an enemy combatant. This is the *Combatant's Stance*.

---

61. The term "combatant'" in this discussion is meant to refer to the common-sense use of the word, meaning a soldier in combat, and is not meant to carry the significance of the legal term of art.

62. *See, e.g.*, Robert Sparrow, *Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender Robert Sparrow*, 91 INTERNATIONAL LAW STUDIES 699 (2015) (recognizing the importance of surrender recognition and discussing various technological and legal solutions that might allow combatants to effectively surrender to autonomous weapon systems).

What does it mean to interact with an AWS as an enemy combatant, to take the combatant's stance in one's interaction with it? Indeed, it might not be obvious that enemy combatants *interact* with each other at all, since they are trying to kill each other. But this would be a hasty conclusion. Combatants are, in fact, confederates in some deeper sense, members of a global fraternity of professional soldiers who operate according to a different code of conduct that finds expression in rules of engagement, norms of ethical conduct, and legal prohibitions.[63] Professional soldiers hold themselves and each other to specific standards: they wear uniforms and distinguish themselves from the civilian population, they limit their direct attacks to military targets, and their aim is to kill the enemy without avarice or sadism, but simply to achieve the war aim and end the conflict as quickly as possible.[64] These ethical and legal norms bind together professional soldiers, even those fighting on opposite sides of a battlefield.

With these shared norms in mind, it is at least possible to imagine an AWS whose *behavior* was such that it could only be understood in this light, as operating under those constraints and being subject to them as well.[65] If the AWS were limiting its attacks in this way, and demanding of others with whom it operated a similar level of constraint, it would be hard to interact with the AWS in any other way. Understanding its behavior would be impossible unless one attributed to it certain normative commitments, i.e. to pursue warfare within the above constraints. One would need to engage with the AWS as an *enemy combatant*, as opposed to engaging with the AWS as a *weapon*.[66] This would signal that the AWS is, in fact, pragmatically

63. *See* Mary Ellen O'Connell, *Legal Development and Historical Basis*, *in* THE HANDBOOK OF INTERNATIONAL HUMANITARIAN LAW 1, 18 (Dieter Fleck ed., 3d ed. 2013).

64. Killing simply to inflict suffering on the enemy has no place in the actions of a professional soldier. *See* U.S. Department of War, Instructions for the Government of Armies of the United States in the Field, General Orders No. 100, arts. 14–16, Apr. 24, 1863 (Lieber Code).

65. *But see* Benjamin Kastan, *Autonomous Weapons Systems: A Coming Legal "Singularity"?*, UNIVERSITY OF ILLINOIS JOURNAL OF LAW, TECHNOLOGY, & POLICY 45, 50 (2013) ("There are substantial debates in the robotics community regarding the likelihood of highly intelligent systems ever being developed.").

66. In order to engage in this process of behavior interpretation, it would not be necessary to understand the intricacies of the program's mechanics, but simply understand the goals that the system is trying to rationally obtain. *See* Kenneth Anderson, Daniel Reisner & Matthew Waxman, *Adapting the Law of Armed Conflict to Autonomous Weapon Systems*, 90 INTERNATIONAL LAW STUDIES 386, 394 (2014) [hereinafter Anderson, Reisner & Waxman, *Adapting the Law*] ("Perhaps foremost among these is the fact that as machine-learning and artificial intelligence technologies develop, it is becoming increasingly clear

equal to an intentional agent in warfare. This would mean adopting the combatant's stance in one's interaction with the AWS because this would provide one with the best avenue for understanding its behavior. This approach would be justified by pure pragmatism. And asking deeper questions about the underlying cognitive states of the AWS—as John Searle would encourage us to do with his Chinese Room thought experiment[67]—would be largely irrelevant for how we interact with the AWS. Adopting the combatant's stance means that treating the AWS as an enemy combatant—and ascribing to it certain beliefs and desires regarding its ultimate goal of winning the war and what belligerent actions it takes in order to achieve that goal—would be the only way of understanding, predicting and ultimately defeating its behavior.

One might object that deeper questions of cognition might be important for determining whether the AWS is a moral agent and whether the rights and duties of combatants can be legitimately ascribed to them.[68] If the AWS is merely *copying* rational behavior, then arguably it is not a moral agent; in that case, it would make no sense to say that the AWS has a moral right not to be killed or tortured after capture. By extension, if the AWS is not a moral agent, it would make little sense to hold it accountable for its violations of international criminal law through some form of machine punishment. These are, of course, legitimate constraints on the moral pro-

---

that human beings may not necessarily always be able to understand how (and possibly why) autonomous systems make decisions. In some cases, this is a question of the complexity of the system, making it realistically impossible to fully predict in advance how the system will behave, to understand the reasoning behind the system's decision making as it happens or even to reconstruct after the fact how it did behave.").

67. *See* John Searle, *Minds, Brains and Programs*, 3 BEHAVIORAL & BRAIN SCIENCES 417–57 (1980) (describing a thought experiment where an individual is in a locked room with a manual for mechanically transcribing a set of Chinese characters into responses that are then given back to a person on the outside). Searle concludes that although the responses of the person in the room are identical to a person who speaks and understands Chinese, the person in the room does not "understand" Chinese in any meaningful sense. He concludes, on the basis of this intuition, that computers capable of engaging in behavior that is functionally indistinguishable from the behavior of human beings are not, *pace* Turing, engaged in meaningful thought or artificial intelligence. In other words, "whatever purely formal principles you put into the computer, they will not be sufficient for understanding, since a human will be able to follow the formal principles without understanding anything." *Id.*

68. *Cf.* ARMIN KRISHNAN, KILLER ROBOTS: LEGALITY AND ETHICALITY OF AUTONOMOUS WEAPONS 140 (2009) (concluding that the use of robots in combat raises multiple ethical and moral issues even if robots are not capable of emotion or meaningful intelligence).

cesses of holding other rational agents accountable for their behavior and decisions. Furthermore, adopting the combatant's stance in no way logically presumes, or even assumes, an answer regarding the moral status of the other combatant.

However, IHL is not reducible to individual criminal accountability or even moral responsibility. Even in situations where particular agents are not culpable—and not valid subjects of *ex post* criminal punishment—the rest of IHL is still applicable. To see why this is so, consider an army composed entirely of children. The children might be so young that they are not fully responsible agents, and in any event as a matter of law the ICC would not have jurisdiction to prosecute them for committing violations of the Rome Statute (because it excludes jurisdiction over children).[69] In this situation, the conflict with this child army would still be governed by the core principles of IHL, and both sides of the conflict would be duty bound by *jus in bello* to comply with its restrictions. The same thing could be said regarding an armed conflict against an army composed of psychotic aggressors.[70]

Now apply this insight to a conflict that includes an AWS. If the AWS is indeed functionally indistinguishable at a distance from a natural combatant, then all combatants would be forced to treat, by logical necessity, the AWS as they would any other combatants, both in order to understand its behavior but also to decide on whatever counter-move is appropriate. It being impossible to see inside the AWS and learn its programming from the inside, the best way of predicting or understanding its behavior would be to posit the very same mental states (including beliefs and desires) that one posits when confronted with a natural combatant and attempt to come to terms with its behavior and why it is doing one thing rather than another. There would be no other practical alternative for how one engages with an AWS; one would have to adopt the combatant's stance with it.

One might object that in battle a combatant would simply blow up an AWS and would not "engage" with the AWS on any level—intentional, conversational, combatant, or otherwise. This objection misses the point. In some circumstances, blowing up the AWS would constitute "engaging" with it as a natural enemy combatant, since often that is what one does to enemy combatants in battle. But that simple interaction hides a complex set

69. *See* Rome Statute, *supra* note 52, art. 26.

70. The issue of psychotic aggressors was first addressed by George Fletcher in his essay George Fletcher, *Proportionality and the Psychotic Aggressor: A Vignette in Comparative Criminal Theory*, 8 ISRAEL LAW REVIEW 367 (1973).

of assessments that structure the interaction: the assumption that the other combatant constitutes an enemy, that the other combatant would destroy you if it had the opportunity, that its movements on the battlefield are evasive in nature and designed to avoid detection or destruction, that engaging the other combatant is consistent with one's rules of engagement (if one is committed to their observance) and that destroying the agent is constitutive of the war aim and would hasten, in some small way, a victorious conclusion to the conflict. These are all complex assessments that are internal to the relationship of belligerency between the parties. The fact that one of the parties to the engagement is an artificial agent in no way liberates the other combatants from having to engage with the AWS in this way. In other words, victory over the AWS would require adopting the combatant's stance towards it and treating it as an enemy combatant.

Again, the fact that the AWS is not a responsible agent for purposes of criminal law is irrelevant to determining whether the AWS should be viewed and interpreted with the combatant's stance. Moral agency for purposes of determining rights and responsibilities is one thing; treating the AWS as an enemy combatant is quite another. One might object here that many of the principles of IHL are based on rights that flow from the moral dignity of the combatants, such as the right to be free from torture or the right not to be summarily executed after capture. This much is true but it is beside the point. Even if one assumes that the AWS is entitled to *none* of the IHL protections that flow from moral agency (and could be summarily destroyed after capture, for example), it would *still* be the case that engaging the AWS in battle, and understanding its behavior, would require thinking of the AWS as an enemy combatant pursuing particular objectives in accordance with its own rules of engagement. Whether those objectives and rules of engagement are the result of humanlike obedience towards command influence or, in contrast, the result of its internal *programming*, would make no difference for how the combatant engages with, treats and thinks about the AWS. It would be, in other words, an enemy combatant *simpliciter*. Any deeper investigation of intentional states would be irrelevant for inquiry and pragmatic interaction.[71]

Having now defended an account for evaluating the status of an AWS as a rational agent and combatant, I want to emphasize that this inquiry is

71. Obviously, part of the inspiration here is Richard Rorty's suggestion that if something makes no difference to inquiry it should make no difference to philosophy—a motto for pragmatism if you will. *See* RICHARD RORTY, TRUTH AND PROGRESS: PHILOSOPHICAL PAPERS 19 (1998).

logically independent of the analysis regarding the commander's liability for deploying the AWS. As I tried to outline in the previous section, the commander is liable for the crimes committed by the AWS regardless of whether the AWS is a responsible moral agent or not and regardless of whether the AWS satisfies the Turing Test for Combatancy as I have outlined it above or whether an enemy combatant would adopt the combatant's stance in his or her interaction with it. That conclusion follows logically from the rejection of the innocent instrumentality rule—a doctrinal move that was made for solid normative reasons. Even if the "instrument" of one's criminality is a morally culpable agent, there are still some situations when the commander ought to be viewed as a principal perpetrator who controls the crime indirectly. The participation of the AWS would be the "through element" by which the crime is completed, analogous to the hierarchical organization that is deployed in *Organisationsherrschaft* cases.

VI. THE PROBLEM OF RECKLESSNESS

Unfortunately, there are other areas of its doctrine where international criminal law is *not* well suited to prosecuting commanders for committing an international crime through an AWS. The uncontroversial (and probably rare) cases involve a commander who purposefully deploys an AWS to commit an international crime such as deliberately killing civilians or engaging in genocide. In that instance the current doctrine functions adequately. But these are unlikely (though important) circumstances. The far more frequent occurrence is one where the commander deploys the AWS for military operations and the AWS violates a core prohibition of IHL: distinction, necessity, or proportionality.[72] Furthermore, if the commander is

72. Several scholars have argued that a commander who deploys an AWS that commits an international crime should be liable under basic principles of command responsibility. *See, e.g.*, Christopher P. Toscano, *"Friend of Humans": An Argument for Developing Autonomous Weapons Systems*, 8 JOURNAL OF NATIONAL SECURITY LAW & POLICY 189, 236 (2015) ("Therefore, military commanders are assigned responsibility even if they do not control the outcome, because they are accountable for the creating conditions [sic] under which their subordinates act. Although an AWS is not a subordinate, *stricto senso*, because it is a weapon and not a sentient being, a commander nevertheless retains effective control over this weapon and subsequent employment on the battlefield no differently from any other military equipment."). Designers and programmers could be liable for failure to program the AWS with sufficient concern for proportionality or other IHL norms. *See* BOOTHBY, *supra* note 3, at 146; CHRISTIAN TOMUSCHAT, HUMAN RIGHTS: BETWEEN IDEALISM AND REALISM 346–47 (3d ed. 2014) (suggesting that fully autonomous systems

aware of the strong possibility of this outcome, the commander is reckless as to his deployment of the AWS.[73]

Although this sounds like bread and butter application of IHL principles, international criminal law has a serious blind spot where crimes of recklessness are concerned, and no one is clear how they fit into a general theory of international criminality.[74] Take, for example, Article 30 of the Rome Statute, which lists the default mental requirements for international crimes as intent and knowledge "unless otherwise provided."[75] Unfortunately, it is unsettled whether Article 30 permits a conviction for crimes of recklessness or what civil law trained criminal lawyers call *dolus eventualis* (liability for the risk of a future event and a concrete decision to move forward even in light of the negative outcome). Some scholars have argued that "intent" in Article 30 includes *dolus eventualis*, since criminal lawyers in civil law jurisdictions are trained to think of it as a species of "intent" (*dolus*) in the broadest sense of the word. In contrast, common law trained criminal lawyers are inclined to view intent as encompassing purpose and knowledge, but not recklessness, which is a less culpable mental state than "intent." The other possible route for reading *dolus eventualis* or recklessness into Article 30 of the Rome Statute is through the "unless otherwise provided" prong, which envisions that specific international crimes might allow lower mental states to qualify for a conviction, provided that this departure from the Article 30 default mental state requirement is actually codified somewhere.[76]

---

will fail to respect the principles of distinction and proportionality but "it will be hard, almost impossible, to establish criminal responsibility").

73. For a discussion of the difficulties inherent in establishing *mens rea* for military commanders deploying an AWS, see Wagner, *supra* note 35, at 1406 (discussing situations where a "commanding officer sends AWS into a situation for which they were not designed"). *See also* Allyson Hauptman, *Autonomous Weapons and the Law of Armed Conflict*, 218 MILITARY LAW REVIEW 170, 194–95 (2013) ("As the concept of command responsibility had to develop, its application to autonomous weapons systems will as well.").

74. Other scholars have noted that the criminal law may have problems regulating negligent or reckless behavior with regard to AWS. *See* Kastan, *supra* note 65, at 78 ("The largest gap in applying current civil law to AWSs is in the area of operational negligence. It is not clear how courts would approach who may be properly held negligent in the case of deploying AWSs.").

75. *See* Rome Statute, *supra* note 52, art. 30.

76. *But see* Gerhard Werle & Florian Jessberger, *"Unless Otherwise Provided": Article 30 of the ICC Statute and the Mental Elements of Crimes Under International Criminal Law*, 3 JOURNAL OF INTERNATIONAL CRIMINAL JUSTICE 35, 41 (2005) (arguing that lower mental states

This is relevant because some IHL lawyers believe that a number of war crimes are governed by a lower mental state such as recklessness or *dolus eventualis*.[77] For example, the ICRC Commentary to the Geneva Convention concludes that the appropriate mental state for the war crime of indiscriminately attacking a civilian population—a crime that would certainly be relevant for a rogue AWS—is recklessness.[78] In other words, a commander who orders an attack that is indiscriminate would violate the relevant IHL prohibition if he is simply reckless as to this result. One could very well imagine this scenario playing out with the AWS as the relevant weapons platform executing the indiscriminate attack.[79]

The issue of recklessness also comes into play, albeit controversially, when specific modes of liability are applied to the military commander. At least some interpretations of command responsibility allow for a military commander to be prosecuted for crimes committed by subordinates if the military commander fails to adequately punish or investigate the crimes.[80] This has nothing to do with intent or purpose with regard to the resulting crimes, but rather is a crime of recklessness—failure to supervise or punish troops under one's command is reckless because the commander realizes that there is a risk that lack of discipline might lead to the commission of war crimes—but the commander refuses to supervise his troops anyway.[81] Furthermore, the third variant of the Joint Criminal Enterprise doctrine (JCE III) also involves a crime of recklessness or *dolus eventualis*, because the defendant is convicted for criminal acts performed by other members of the JCE, even if those acts fall outside of the scope of the criminal design.

---

from *customary* international law are incorporated into Article 30 through the "unless otherwise provided" phrase).

77. Prosecutor v. Galic, Case No. IT-98-29-T, Judgment, ¶ 25–32 (Int'l Crim. Trib. for the former Yugoslavia Dec. 5, 2003).

78. COMMENTARY ON THE ADDITIONAL PROTOCOLS OF 8 JUNE 1977 TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949, ¶ 3474 (Yves Sandoz, Christophe Swinarski & Bruno Zimmermann eds., 1987) (defining the war crime of willfully attacking civilians as "the accused must have acted consciously and with intent, i.e., with his mind on the act and its consequences, and willing them ('criminal intent' or 'malice aforethought'); this encompasses the concepts of 'wrongful intent' or 'recklessness,' viz., the attitude of an agent who, without being certain of a particular result, accepts the possibility of it happening; on the other hand, ordinary negligence or lack of foresight is not covered, i.e., when a man acts without having his mind on the act or its consequences").

79. *See* Wagner, *supra* note 35, at 1407.

80. GUÉNAEL METTRAUX, THE LAW OF COMMAND RESPONSIBILITY 225 (2009).

81. For a discussion of the mental state, see generally MOHAMED ELEWA BADAR, THE CONCEPT OF *MENS REA* IN INTERNATIONAL CRIMINAL LAW (2013).

The only doctrinal constraint on JCE III is that the defendants reasonably foresee the possibility that confederates would stray from the agreed upon plan and the defendants willingly took the risk anyway. This is recklessness.[82]

The problem here—and it is a major problem—is that the doctrine has no way of encoding this lower level mental state in the grading of the offense. In contrast, domestic criminal law usually distinguishes between homicides committed with intent and those committed with recklessness or negligence. A typical legislative grading scheme in, say, the United States, classifies first-degree murder as intentional homicide, while second-degree murder penalizes homicides committed with depraved indifference to human life. Other homicides performed recklessly, but which do not meet the depraved indifference threshold, are classified as manslaughter. The key point here is not the specifics of the classification scheme, but rather the bare fact that the definition of offenses reflects key differences in culpability. The most culpable killers are convicted of one offense, while less culpable killers are convicted of a lesser offense—whatever that may be.

This legislative grading of culpability is not only pragmatically desirable and smart policy—it also constitutes a moral imperative. Defendants and society both have a reasonable expectation that the criminal categories that are attached to their conduct will adequately reflect the state of their criminality and will not unreasonably conflate significant differences in culpability. This notion is often referred to in the literature as "fair labeling."[83] Although it would be an exaggeration to claim that defendants in the United States have a constitutional right to fair labeling, it is nonetheless a fundamental principle of criminal law and arguably one that is implicit in the principle of culpability. Indeed, the principle of culpability means that defendants should never be convicted of offenses that are defined in such a way that they make no attempt to calibrate convictions for particular offenses and individual judgments regarding culpability.

The counter-point to this argument is that judges can usually make relative determinations of culpability at the sentencing phase, thus ensuring that defendants never receive a punishment that exceeds their level of culpability. While this is an attractive point, it ignores the entire impulse behind the special part of the criminal law, which is to require legislative and pre-announced definitions of the offenses, which make morally significant

82. *See* ILIAS BANTEKAS & SUSAN NASH, INTERNATIONAL CRIMINAL LAW 33 (2009).

83. *See* Andrew Ashworth, *The Elasticity of* Mens Rea, *in* CRIME, PROOF, AND PUNISHMENT: ESSAYS IN MEMORY OF SIR RUPERT CROSS 45 (C.F.H. Tapper ed., 1981).

distinctions regarding culpability. To suggest that a judge can make all of these distinctions at sentencing is to open up an impermissible level of discretion towards the sentencing phase and to reduce the guilt phase to the simple question of whether the defendant engaged in *some* culpable act, rather than determining his level of culpability based on the specific legal prohibition that he or she has violated.

International criminal lawyers are usually sensitive to this point when it comes to the distinction between principals and accessories.[84] Since the distinction is imperative to the deep structure of several domestic systems of criminal law, lawyers from those jurisdictions are inclined to argue that international criminal law *must* distinguish between principals and accomplices in its doctrine, as opposed to simply handing down different sentences without labeling one as principal and another as accomplice. To a German criminal lawyer, for example, a unitary approach to perpetration violates the principle of fair labeling—the idea that the criminal law ought to attach the right labels to the conduct.[85] For this reason, unitary theories of perpetration (which ignore the distinction between principals and accomplices) have garnered little traction in international criminal law recently.[86] Strangely, though, the same point is not always recognized when crimes of recklessness are understood. Domestic criminal law contains specific doctrinal labels for defendants who have engaged in crimes of mere recklessness: they are punished for different offenses and not confused with murderers who commit intentional killings. This is also demanded by the principle of fair labeling.

Unfortunately, though, international criminal law has no analogue to manslaughter, a specific offense that separates out reckless defendants from those who commit intentional crimes. The core international crimes are aggression, genocide, crimes against humanity and war crimes. True, each so-called "chapeau" offense must be linked with a more specific predicate offense such as killing, persecution, torture, etc., from a defined list of predicate offenses that are different for each chapeau offense. What would be needed is a different chapeau offense that distinguishes between, say, an

84. *See* Gerhard Werle & Boris Burghardt, *Establishing Degrees of Responsibility: Modes of Participation in Article 25 of the ICC Statute*, *in* PLURALISM IN INTERNATIONAL CRIMINAL LAW 301 (Elies van Sliedregt & Sergey Vasiliev eds., 2014).

85. *Id.*

86. For one recent and cogently argued example, see James G. Stewart, *The End of "Modes of Liability" For International Crimes?*, 25 LEIDEN JOURNAL OF INTERNATIONAL LAW 165 (2012).

intentional war criminal from a reckless war criminal, or distinguishing someone whose conduct carried the risk of being a crime against humanity from someone who intentionally carried out such a crime.

For some of these crimes, recklessness or *dolus eventualis* has generally been excluded. For example, the crime of genocide requires that the defendant act with a *dolus specialis*, i.e. genocidal intent. Most scholars, even Antonio Cassese, concluded that the requirement of genocidal intent is, by definition, inconsistent with the application of recklessness or *dolus eventualis*.[87] But, as for other crimes, most especially war crimes (and to a lesser extent crimes against humanity), the question of recklessness and *dolus eventualis* remains a hotly debated subject, with some scholars concluding that it is imperative for policy reasons that proving reckless conduct should satisfy the prosecutorial burden for a conviction for many international crimes.[88] And often the recklessness comes in through the teeth of the mode of liability, further obfuscating the distinctions of culpability that the criminal law ought to be highlighting and magnifying, not obscuring.[89] The question remains, though, why do we police so thoroughly levels of culpability with regard to principals and accessories, but the similarly important distinction between intentional and reckless conduct gets a free pass in international criminal law? If we are committed to the principle of fair labeling in the doctrine itself (as opposed to letting sentencing taking care of it) then we should demand clarity in this area of the doctrine as well.

This is a huge problem for the prosecution of a commander who has deployed an AWS that commits a war crime. As I said above, the most likely scenario is that the military commander was reckless in his deployment of the AWS, in the sense that he was aware of the risk that the AWS would violate a core prohibition in IHL, but the military commander decided to deploy the system anyway.[90] But is it really appropriate to convict this mili-

87. Antonio Cassese, *The Proper Limits of Individual Responsibility under the Doctrine of Joint Criminal Enterprise*, 5 JOURNAL OF INTERNATIONAL CRIMINAL JUSTICE 109 (2007).

88. *See, e.g.*, Steffen Wirth, *Co-Perpetration in the Lubanga Trial Judgment*, 10 JOURNAL OF INTERNATIONAL CRIMINAL JUSTICE 971, 990 (2012).

89. *See* Stewart, *supra* note 86, at 175.

90. For a similar conclusion, see Tim McFarland & Tim McCormack, *Mind the Gap: Can Developers of Autonomous Weapons Systems Be Liable for War Crimes?*, 90 INTERNATIONAL LAW STUDIES. 361, 384–85 (2014) ("In situations of legally significant levels of autonomy, the operator who deploys the weapons system may well be excused of individual liability for lack of effective control over the behavior of the weapons system. However, the commander who has called for the deployment of the weapons system will be on notice

tary commander for the exact same crime as the war criminal who *intentionally* targets civilians or intentionally executes a fallen soldier who is *hors de combat* or a prisoner of war? Surely, the war criminal who commits these intentional crimes (whether by his own hands or indirectly through the deployment of an AWS) is much more culpable than the commander who has recklessly deployed an AWS that fails to abide by the restrictions imposed by IHL. As it stands now, there is no convincing way to codify these distinctions. Although international criminal law has largely skirted this issue, the prosecution of a commander for indirectly perpetrating a crime through an AWS will place this problem on the front burner.

Are there any solutions to this problem? One solution is to fully exonerate a military commander for recklessly perpetrating a war crime indirectly through an AWS, but this seems unsatisfactory.[91] While this hypothetical military commander is less culpable than one who deploys the AWS with the intent of committing the crime, it would be an exaggeration to say that the defendant is not culpable at all. What possibilities exist for representing the commander's culpability?[92]

One possibility is to take off the shelf—and re-purpose—the distinction between principals and accessories and use it to represent the distinction between crimes of intent and crimes of recklessness. In other words, if a military commander is merely reckless as to the violations of IHL com-

---

once the system behaves in such a manner as to have resulted in a serious violation of the law of armed conflict.").

91. Commentators who have written about international responsibility for crimes committed by AWS have generally not discussed the problem of adequately distinguishing between crimes of negligence and crimes of intent. *See, e.g.*, George R. Lucas, Jr., *Legal and Ethical Precepts Governing Emerging Military Technologies: Research and Use*, 2013 UTAH LAW REVIEW 1271, 1280 (2013) ("By contrast, R&D, design, or manufacturing of systems undertaken through culpable ignorance or in deliberate or willful disregard of these precepts (including failure to perform or attempts to falsify the results, tests regarding safety, reliability of operation, or compliance with applicable law and ROEs, especially in the aftermath of malfunctions as noted above) shall be subject to designation as war crimes under international law, or as reckless endangerment or criminally negligent behavior under the terms of applicable international or domestic law.").

92. Criminal prosecutions might be subject to other obstacles as well. *See* McFarland & McCormack, *supra* note 90, at 384 ("Attempting to identify an individual most responsible for subsequent behavior of a deployed weapons system that constitutes a war crime may simply be too difficult for the purposes of initiating trial proceedings."). Also, some of the responsible perpetrators may be corporations, who are currently not prosecutable before most international tribunals. For more discussion, see Jack M. Beard, *Autonomous Weapons and Human Responsibilities*, 45 GEORGETOWN JOURNAL OF INTERNATIONAL LAW 617, 648 (2014).

mitted by the AWS that he deploys, then perhaps the military commander should be convicted as an *accomplice*, despite the fact that the commander meets all other doctrinal requirements for prosecution as a principal, i.e. he indirectly perpetrated the crime through the AWS as an instrument of his criminality.[93] Labeling a military commander as an accomplice in this context would capture his relative degree of culpability as compared to commanders who commit crimes of intent. One potential downside to this doctrinal solution is that it *requires*—unlike the discussion above—that the AWS be viewed as a morally responsible agent. *Why*? As a matter of logic and structure, accomplice liability is derivative and requires the existence of some principal perpetrator—somewhere—whose endeavor is then supported or assisted by the accomplice. But if the AWS is neither autonomous nor responsible, then there is no principal perpetrator in the scenario to whom the military commander can direct his assistance. Accomplice liability would require the very finding that I earlier suggested that international criminal law could safely avoid.

The other possibility, which avoids this conundrum, is to codify a new criminal offense for recklessly perpetrating an international crime (under some circumstances).[94] The military commander who recklessly deploys an AWS that violates IHL would then be guilty of a specific criminal offense, whose criminal definition would make clear that the crime is less culpable

93. Sassóli concludes that a military commander should be liable for actions committed by an AWS, though strictly speaking not under the command responsibility doctrine. *See* Marco Sassóli, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to Be Clarified*, 90 INTERNATIONAL LAW STUDIES 308, 324 (2014) ("As for the first options, it is as fair to hold a commander of a robot accountable as it would be to hold accountable a commander who instructs a pilot to bomb a target he describes as a military headquarters, but which turns out to be a kindergarten."). Sassóli concludes that the commander's responsibility in these cases is one of direct responsibility, not derivative responsibility, because the AWS is a mere instrument like any other weapon. *Id.*

94. I should note that this article focuses on formal legal regulation through criminal responsibility. Of course, there are also avenues of *informal* legal regulation that I have not considered here. *See* Crootof, *supra* note 37, at 1901 ("The international legal system has myriad alternative sources of guidance and governance, many of which can be extremely effective in channeling state action, notwithstanding their lack of formal international legal status."). One reason to focus on criminal regulation is that the basic *institutional* resources are already in place. *See* Lucas, *supra* note 91, at 1275 ("Instead, a regulatory and criminal regime, respecting relative legal jurisdictions, already exists to hold accountable individuals and organizations that might engage in reckless or criminally negligent behavior in the design, manufacture, and ultimate use of unmanned systems of any sort.").

than the other core crimes of international law. The virtue of this approach is that it would signal to the rest of the world the defendant's real level of culpability, and it would not confuse matters by hijacking the distinction between principals and accessories. The "new substantive offense" approach has been used in domestic criminal law before. For example, some jurisdictions solve the purpose-knowledge debate for complicity by crafting a new substantive offense for criminal facilitation—a less serious offense that only requires acting with knowledge.[95] Similarly, German criminal lawyers solve the problem of voluntary intoxication (should deliberately intoxicated defendants be excused?) by crafting a special lesser offense for the commission of a crime while in a state of voluntary intoxication.

The downside, however, is that such legislative solutions are unlikely in a system controlled by the Assembly of State Parties at the ICC—an institution unlikely to consider substantial revisions to the Special Part of the Rome Statute. That being said, some legislative amendment is surely required if and when AWS (and their associated war crimes) become a reality. The amendments proposed above represent relatively modest changes that should garner support.

VII. CONCLUSION

The basic structure of international criminal law is already well-suited to prosecuting military commanders for deployment of an AWS that commits a war crime.[96] The metaphor of deploying an organized "machine" to perpetrate an international crime is deeply embedded in the founding DNA of international criminal law. In its more modern instantiations, the doctrine of indirect perpetration would allow an international court to prosecute a military commander for using an AWS as an instrument of criminality, in analogous fashion to the deployment of a hierarchical apparatus of power for the same purpose. Moreover, the doctrine applies regardless of whether the physical perpetrator of the criminal act is culpable or not, thus making

95. *See, e.g.*, N.Y. PENAL LAW § 115 (McKinney 1967) (substantive offense of criminal facilitation).

96. Indeed, the appropriate question is how to produce a system of regulation that is "coherent and practical." *See* Anderson, Reisner & Waxman, *Adapting the Law*, *supra* note 66. *See also* Sassóli, *supra* note 93, at 338 ("I assume even autonomous weapon systems with artificial intelligence, though capable of learning, cannot do what the human beings who created them do not want them to do—or that it is at least possible to limit their autonomy in this regard. Such must be the case because they are not addressees of the law.").

it largely irrelevant whether the AWS qualifies as an autonomous agent with moral responsibility. This is an important development because I have stressed that an AWS might become so sophisticated that opposing combatants will have to take the "combatant's stance" in order to understand it, predict its next move and defeat it on the battlefield. But the removal of the innocent instrumentality rule would still permit the prosecution of commanders as indirect perpetrators. Unfortunately, the one area where international criminal law is still ill-suited for the task is its inadequate treatment of crimes of recklessness. There are multiple reasons to solve the problem of recklessness, many of which have nothing to do with AWS, but we should not need another motivation to get the doctrine right. In cases where a military commander is reckless in his or her deployment of an AWS, the law should be carefully calibrated to express his or her exact level of culpability—not more and not less.