Chapter III

The “Accountability Gap” Problem
Who is to Blame for Autonomous Weapons Systems’ Misdoings?


1. Introduction

In the ever-growing scholarly and diplomatic debate on the ethical and legal implications of autonomous weapons systems (AWS), a prominent place is given to responsibility issues. To be discussed, in particular, is who (or what legal entity) would be to blame in case an AWS makes a targeting decision that, were it taken by a human agent, would trigger State responsibility or, in the most serious cases, individual criminal responsibility. Indeed, even the most vocal opponents of a ban on AWS are ready to concede that their deployment on the battlefield will not result in a spotless application of relevant international norms: in their predictions, which are in some cases overly optimistic, AWS will maybe outperform human soldiers, but will never be able to ensure a 100% degree of abidance to the law.

Academic discussion on this topic was kicked off in 2007 by Rob Sparrow, who pondered the position of the various candidates for responsibility in case of damages (first and foremost, civilian casualties)

1 See above Chapter II, Section 6.
caused by an AWS – in particular, the programmer, the commanding officer and the machine itself – reaching the conclusion that increasing autonomy in weapons systems would have made it impossible to hold someone responsible, if not at the cost of blaming a human agent for actions lying beyond his/her effective control. Sparrow’s insights set the stage for the subsequent debate on what has become known as the “responsibility (or accountability) gap” problem.

The Group of Governmental Experts (GGE) on Lethal Autonomous Weapons Systems has shown awareness of this problem. The “Guiding Principles” adopted by the GGE in 2018, and slightly modified in 2019, deal with the issue of accountability at Principle (d), whereby:

Accountability for developing, deploying and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control.

Notwithstanding the floods of ink devoted to this topic, it is fair to affirm that the entire discussion is polarised around just two antagonistic views. At one end of the spectrum, we have those who scramble to identify the responsibility-bearers for AWS’ misdemeanours, be them one or more designated individual(s) in the decision-making chain and/or legal entities (i.e., States or corporations). At the other end, there are critics of AWS who argue that every attempt to fill the accountability gap created by AWS is doomed to fail.

Our inquiry will be therefore developed along these lines. After illustrating the “structural problems” which make the ascription of responsibility for AWS’ activities particularly difficult (Section 2), the alternative routes proposed to solve the accountability gap problem will be assessed. The analysis will focus, in the first place, on the international criminal responsibility of the individuals who somehow involved in the process of production, deployment and activa-

4 A quick clarification on terminology is in order here. In the present Chapter, “accountability” will be employed as an umbrella term capturing the various forms of responsibility coming into play in the AWS debate, including individual criminal responsibility, State responsibility and corporate liability.
tion of AWS (Section 3). The possibility to hold the deploying State accountable for AWS’ wrongdoings will then be gauged (Section 4). Subsequently, attention will be paid to the responsibility of the corporations manufacturing and/or programming AWS (Section 5). It will be observed that existing law may contribute, with minor adaptations, to solve some accountability issues more effectively than the critics of AWS are ready to admit. At the same time, it will be submitted that the more we proceed along the path of autonomy in weapons systems, the more radical reforms – such as the expansion of the scope of superior responsibility and the adoption of a no-fault liability regime for States and corporations – will be needed to ensure that someone (individual, State or corporation) is held to account for a weapons system objectively acting in contrast with international legal prescriptions. Section 6 will conclude by questioning the normative desirability of such a paradigm shift in the law governing responsibility issues in the context of armed conflicts.

2. Structural Problems with Responsibility Ascription for AWS’ Misdoings

Quite unsurprisingly, the discussion on the accountability gaps potentially provoked by autonomous machines is not peculiar to military technology. Rather, it builds upon the wider debate among ethicists on the way Artificial Intelligence (AI) and, more generally, computer technology may affect responsibility ascription for harmful events. This debate has brought two factors with the potential to function as “structural” sources of accountability gaps into the limelight, namely i) the inherent unpredictability of autonomous machines and ii) the “many hands” problem, which will be examined in turn in the following sub-paragraphs.

2.1. The Unpredictability of Autonomous Machines

In a farsighted 2004 article, Andreas Matthias pinpointed the inevitability of “responsibility gaps” as a consequence of the development and use of autonomous, learning machines.7 He set out from the observation that machines are nowadays entrusted with increasingly complex tasks, whose accomplishment is ever more difficult to

ensure through traditional programming, whereby all possible situations are coded into long and articulated sequences of behavioural rules based on “if-then-else” constructs.\(^8\) Against this background, optimal performances can be ensured only by endowing machines with some learning capabilities, so that they may develop their own rules to face situations that were not (and could not have been) foreseen during the code production process. To this end, different programming techniques can be resorted to, ranging from more conventional symbolic programming systems (whose decision-making process is still humanly understandable – at least \textit{ex post facto}) to artificial neural networks and genetically evolved software, which produce decisions whose rationale may remain obscure to humans.\(^9\) What all these different techniques have in common, however, is that they turn programming from mere “coding” to “creating” software organisms, which no longer behave predictably on the basis of fixed rules, but shape their actions in response to the operational environment.\(^10\) In this way, the relationship of control between the machine and the human agent (be it the programmer or the final user) is thus severed, with the consequence that the latter cannot be blamed for the former’s harmful activities.\(^11\)

While weapons systems are mentioned by Matthias only cursorily, they constitute a promising field of application for his theory. As evidence thereof, the aforementioned analysis by Rob Sparrow follows a very similar argumentative path.\(^12\) In his view, if an AWS “has sufficient autonomy that it learns from its experience and surroundings then it may make decisions which reflect these as much, or more than, its initial programming”. This inevitably affects the predictability of its choices, which will lie at some point beyond the control of its programmers (or users), so breaking the connection “between the programmers/designers [as well as the users] and the results of the system, which would ground the attribution of responsibility”.\(^13\)

The work by Matthias (and, relatedly, that by Sparrow) stirred a

\(^8\) In computer science, “if-then-else” constructs are a kind of control flow (i.e. a sequence of instructions), whereby the machine is programmed to execute a certain action, only “if” certain conditions are met, otherwise (“else”) different instructions apply.


\(^10\) \textit{Ibidem}, p. 182.


\(^12\) Sparrow, \textit{cit. supra} note 3.

\(^13\) \textit{Ibidem}, p. 70.
lively debate,\textsuperscript{14} within which the inevitability of the creation of responsibility gaps as a consequence of the development of autonomous machines has been put into question from a two-fold perspective.

On the one hand, the deterministic vision they propounded is contested, whereby technology may only develop in such a way as to rule out, sooner or later, human control and understanding of machines’ behaviour. In fact, it is argued, technological development proceeds in a non-linear way, along a path that is drawn by many actors (scientists, corporations, States, NGOs, the public, and so on) pushing in different directions. True, “[a]rtificial agents can be designed so that no human can understand or control what they do”, but they can also be designed in a fully transparent way or, still, so as to keep human control only “on certain aspects or levels of the machine behaviour”.\textsuperscript{15} The direction that will be taken will ultimately constitute the outcome of the interactions among the actors involved in shaping the design practices relating to autonomous machines.

On the other hand, it is denied that a just ascription of responsibility must always be based on a “control requirement”, whereby “a person is responsible for $x$ only if the person has control over $x$”.\textsuperscript{16} A control requirement is perhaps needed in order to establish moral blameworthiness, but this does not necessarily entail that the application of other forms of responsibility – disconnected from moral blame – contrasts with our sense of justice.\textsuperscript{17} It suffices to recall, in this respect, the strict liability regimes introduced in different legal orders with a view to allocating responsibilities in case of damages resulting from dangerous activities, whose application could be extended – mutatis mutandis – to the harmful events caused by autonomous machines.

In the light of the above, it should be ascertained whether, and to what extent, these objections to Matthias’ thesis are relevant also to the AWS debate. In other words, one should verify whether i) AWS can actually be designed in a way to ensure the predictability and the understandability of their behaviours so as to rule out responsibility gaps; ii) the enactment of a responsibility regime that misaligns liability and moral blame is acceptable with regard to the harmful activities of AWS.

\textsuperscript{14} For an overview, see D.G. Johnson, Technology with No Human Responsibility? in J. of Business Ethics, 2014, p. 707 f., pp. 709-710.

\textsuperscript{15} Ibidem, p. 714.

\textsuperscript{16} D. Marino and G. Tamburrini, Learning Robots and Human Responsibility, in Int. Rev. of Information Ethics, 2006, p. 47 f., p. 49.

\textsuperscript{17} Ibidem, pp. 49-50.
Supporters of autonomy in weapons systems would answer the first question in the affirmative. Some of them, mostly lawyers, simply assume that AWS will behave predictably without much bothering to examine why they would do so. Others take pains to explain how unpredictability can be avoided by design. This is the case of Ronald Arkin, who excludes that future AWS could use artificial neural networks precisely because that would hinder the transparency and the predictability of the system.

Contrary to these views, it is here contended that, while design choices (like that propounded by Arkin) may have a role in reducing the overall unpredictability of the AWS’ behaviour, they will never eliminate it completely. As aptly noted by Guglielmo Tamburrini, the predictability of a machine is not only a function of its algorithms: it largely depends on the environment where the machine is meant to act. In this perspective, an effective way to prevent the machine from acting unpredictably is to structure its operational environment so as to keep out all those factors that are likely to put the machine’s perceptual, cognitive and behavioural capabilities under stress, with the first factor among these being the presence of humans. This is the reason why industrial robots are normally segregated from human workers or their interactions with humans are strictly regimented. Yet, this option is generally unavailable when it comes to AWS, which are intended to operate in deeply unstructured warfare scenarios, characterised by the intermingling of legitimate targets with protected persons and objects and where “[e]ach fighting side strives to generate unexpected events that defy the opponent’s predictions”. The risk of AWS’ acting unpredictably as a consequence of their operational scenario is compounded by the prospect, which cannot be discarded as wholly far-fetched, that AWS belonging to opposing belligerents interact with each other on the battlefield. Therefore, it could be plau-

19 Arkin, cit. supra note 2, p. 108.
21 Ibidem, p. 129. In a similar vein, see International Committee of the Red Cross (ICRC), Autonomy, artificial intelligence and robotics: Technical aspects of human control, Geneva, August 2019, pp. 11-13. The most notable exception in this respect is provided by autonomous anti-materiel defensive systems, like the Israeli Iron Dome, which operates in sufficiently structured areas pre-set by military officers on the basis of IHL-informed judgments. See above Chapter I.
sibly maintained that, no matter how they are designed, autonomous weapons systems are bound to act in ways that cannot be foreseen in advance by their users in a non-negligible number of cases.

As to the second question (viz. whether it is possible to disconnect legal responsibility from moral blameworthiness), the answer cannot be separated from the kind of legal responsibility one has in mind. At this preliminary stage of the analysis, one may observe that, as far as individual criminal responsibility is concerned, the introduction of a strict liability regime is simply out of the question. Indeed, the need to align moral blameworthiness and legal responsibility – and, as a consequence, to comply with the “control requirement” – is enshrined by a cornerstone of international criminal law, namely the principle of culpability, whereby “no one may be held accountable for an act he has not performed or in the commission of which he has not in some way participated, or for an omission that cannot be attributed to him”.

As will be seen below, however, such a theoretical hurdle fades away once we move from individual criminal responsibility to State responsibility or corporate civil liability, in relation to which an obligation to pay compensation could be validly grounded on ethical bases other than moral blameworthiness, such as the duty to assume the risk for the harmful events caused by dangerous activities.

### 2.2. The “Many Hands” Problem

An additional, structural source of responsibility gaps comes from what is commonly labelled as the problem of “many hands”. This expression was coined by Dennis Thompson, who used it in relation to the ascription of responsibility in complex organisations, with particular regard to public administration. According to Thompson, the dilution of the decision-making power among many different officials would make it “difficult even in principle to identify who is morally responsible for political outcomes”. This is because when many hands are at work “the different traditional preconditions for responsibility, like intent, knowledge, and freedom of action are distributed over many different individuals and none of them might meet all the conditions”. As a consequence, the “salient and immediate causal an-

---

24 Ibidem, p. 905.
tecedents” of a harmful event (i.e. who or what materially caused the event) “do not converge with its locus of decision-making” (i.e. who, if anyone, actually wanted it to occur).26

Helen Nissenbaum found this problem to be magnified by the increasing reliance of our society on computer technology. And this for a variety of reasons. Firstly, not only are software systems generally developed by complex organisations (software companies), but they are also normally used in articulated institutional settings (corporations, government agencies). Secondly, software programs are often composed of different modules generally developed by different teams of individuals, different producers or even asynchronously and for different purposes. When the software is conceived of in this way, “there may be no single individual who grasps the whole system or keeps track of all the individuals who have contributed to its various components”. Thirdly, the most sophisticated programs function as “systems of systems”, namely they work in pair with (or on the top of) other software systems, which may well be produced by different developers, with the risk that unforeseen incompatibilities between them may result in harmful events. Finally, in many cases software programs operate in a “symbiotic relationship” with robotic machines, which are in turn designed and manufactured by other teams of individuals. Should something go wrong, it could be tough to ascertain whether this is due to the machine, to the computer system or to a faulty interaction between the two.27

The foregoing analysis is relevant to the discussion on AWS in multiple respects. First of all, the deployment of an AWS presupposes an institutional setting that typically entails the involvement of “many hands”, including “the software programmers, those who build or sell

27 Ibidem, pp. 29-32. To illustrate how the problem of many hands “obscures” responsibility in computer-related accidents, it is customary to recall the case of the malfunctioning of the Therac-25, a computer-controlled machine designed for the radiation treatment of cancer. In the ’80s the machine overdosed six patients, causing the death of three of them and severe injuries to the others. As subsequent inquiries put into light, the overdosing was due to several factors, including “software errors, inadequate testing and quality assurance, exaggerated claims about the reliability, bad interface design, overconfidence in software design, and inadequate investigation or follow-up on accident reports” (M. Noorman, Computing and Moral Responsibility, in Stanford Encyclopedia of Philosophy, 2018). This had a remarkable backlash on the responsibility ascription process. Since “[t]he actions or negligence of all those involved might not have proven fatal were it not for the other contributing events”, it proved particularly challenging “to retrospectively identify the appropriate person that [could] be called upon to answer and make amends for the outcome.” (Ibidem).
hardware, [the procurement officials], military commanders, subordinates who deploy these systems and political leaders”. Moreover, given the extraordinarily demanding tasks they will be required to perform, one can easily foresee that AWS will develop as complex “systems of systems”, programmed and manufactured by joint ventures of private and public companies. Finally, AWS are precisely featured by the “symbiotic relationship” between software programme and robotic machines that might muddy the waters as to where the fault lies in case of undesired outcomes.

Indeed, one may easily imagine a “many hands” scenario for AWS-related accidents. Let us consider, for instance, an autonomous UAV model which is programmed to distinguish between military and civilian buildings. Evidence collected in the UAV testing phase suggests that the system fails to identify as civilian objects buildings presenting certain perceptual features in 0.05% of cases. These results, however, are inaccurate, as they do not take into account a number of environmental conditions that would raise the failure rate to 5%, as a consequence of the interaction between an undetected software bug and a design feature of the machine. Inadequate testing by the manufacturer(s) is partly due to the need to accelerate the launch of the UAV on the market and beat potential competitors to the punch. When delivering the purchased system to the battlefield unit, the procurement officer omits to mention the failure rate, solely recommending the greatest caution in the use of the aircraft against targets in the proximity of civilian objects. Basing her or his judgment on a liberal interpretation of this directive, a military commander deploys the aircraft over a sparsely populated area. As a consequence of this chain of events, the UAV targets and bombs a small school, causing several civilian victims.

Each character mentioned in this story contributed in distinctive ways to the accident occurring. Clearly, their acts are significant causal antecedents of the bombing event and, more significantly, the conduct of each involved individual is reprehensible for some reason: the programmer should have tested thoroughly to detect the software bug; similarly the manufacturer should also have carried out more rigorous testing; the procurement officer should have been clearer as to the risks associated with use of the AWS; the military commander should have been less confident in unknown technology. Nevertheless, none could be successfully charged with the war crime of “[i]nten-

tionally directing attacks against civilian objects”, since all of them would have a good chance at shrugging off responsibility by blaming someone else in the decision-making chain.

The example above shows that the many hands problem can be dissected into two aspects. On the one hand, there is the issue of discerning, in a tangle of causal antecedents and loci of decision-making, who to blame for an accident. This problem has been described as merely “practical”, in that it makes responsibility ascription difficult, but not impossible, or “epistemic”, to the extent that the acquisition of more pieces of information should help to find an acceptable solution for it. On the other hand, we have what has been called the “normative or moral problem”, which arises when a group of people can be held collectively responsible for a certain outcome, whereas none of the individual members of the group bear full-blown responsibility for it: we are confronted with a many hands problem in its normative or moral dimension, in other words, every time collective responsibility cannot be reduced to individual responsibility.

The latter aspect has proven to be more troublesome and has compelled ethicists to find a way to avoid responsibility gaps. Mark Bovens summarised these attempts by referring to three alternative models to individual accountability, namely i) corporate accountability; ii) hierarchical accountability; iii) collective accountability. The “corporate accountability” model places the blame directly on the organisation as a whole, so sparing the painstaking task of assessing the blameworthiness of each individual member of the group/organisation. The “hierarchical accountability” model, on the other hand, predicates that the blame for the misconducts of a group should be assigned to those who are on the top of its chain of command. Under the “collective accountability” model, finally, each member of the group is held personally responsible for the conduct of the organisation as a whole, simply because of his/her membership.

It is interesting to note that each of these models bears correspondence to international legal doctrines or regimes that will be analysed in the following pages. Notably, the legal counterparts of the

30 For an analogous conclusion, see N. BHUTA and S.-E. PANTAZOPOULOS, Autonomy and Uncertainty: Increasingly Autonomous Weapons Systems and the International Legal Regulation of Risk, in BHUTA et al. (eds), cit. supra note 20, p. 284 f., p. 289.
32 Ibidem, pp. 50-52.
“corporate accountability” model may easily be found in the regimes of State responsibility and corporate liability; the “hierarchical accountability” model has a match in international criminal law, notably in the doctrine of command responsibility; finally, the “collective accountability” model is echoed – with some qualifications – by the doctrine of the Joint Criminal Enterprise (JCE), whereby each member of a group is held individually responsible for crimes committed by the group within a common plan or purpose, regardless of whether (s)he directly participated to the commission of the crime.

3. Individual Criminal Responsibility

Most of the scholarly analysis coping with the responsibility gap problem is devoted to the problem of individual criminal responsibility. This might be explained with the widespread (but ultimately mistaken) perception that autonomy in weapons systems would solely affect the ascription of responsibility to the individual, while leaving untouched other forms of responsibility, with particular regard to that of the deploying State.33

An academic quest has thus been ushered in to find the most convenient individual(s) to blame in case of civilian casualties provoked by AWS. Lists of candidates for responsibility flourished in scholarly works on the subject, in a bizarre “dance” where potential defendants for AWS-related crimes alternate on the court’s dock. Some candidates are nothing new, as one would look at them in any situations involving the commission of an international crime, regardless of the weapon employed (the operator(s) activating the AWS, the fielding officer); some others are instead peculiar to this new technology and show the creative endeavour by scholarship to tread new paths in search of a responsible person (software developers, procurement officers).

In the following pages, their position will be examined. Our analysis, however, will refrain from using a listing method, but will consider two possible forms of responsibility for AWS-related crimes, namely direct responsibility and superior responsibility. This will assist us in discerning the various problems raised by AWS, without incurring repetition.

3.1. Direct Responsibility

Human wickedness in warfare has accustomed us to the worst

33 See below Section 4.
atrocities. Therefore, it is far from implausible that human agents willfully use AWS to commit international crimes. Let us start with considering the following scenarios:

1. An operator sets an autonomous Unmanned Ground Vehicles (UGV) so that it will shoot at every adult man in a village.
2. A commander orders to deploy a swarm of AWS, being aware that, due to their technical features, in the ordinary course of events the AWS will end up killing or maiming all the enemy combatants on the battlefield, regardless of whether they are already placed hors de combat.

In both cases AWS are just a tool in the hands of human agents, a (highly sophisticated) weapon through which these agents are perpetrating criminal conducts. This makes responsibility ascription relatively unproblematic, regarding both the actus reus and the mens rea.

As to the actus reus, we may refer to the traditional modes of participation to the crime, as codified by Article 25 of the ICC Statute. Notably, the operator in Scenario No. 1 “commits” (art. 25(3)(a) ICC Statute) the war crime of intentionally directing attacks against the civilian population or, depending on the context, murder as a crime against humanity, while the commander in Scenario No. 2 arguably “orders” (art. 25(3)(b) ICC Statute) the perpetration of the war crime of killing or wounding persons hors de combat.

The agents’ mental state, on the other hand, clearly meets the mens rea requirements as they are generally understood in international criminal law. Indeed, not only the harmful events caused by the machine are fully within the cognitive domain of the agents, but they also constitute the object of the latter’s volition. The operator (Sce-

---

34 As well summarized by Kai Ambos, modes of participation to international crimes under the ICC Statute are as follows: Direct perpetration, Co-perpetration, and Perpetration through another person (Art. 25(3)(a)); Ordering, soliciting and inducing (Art. 25(3)(b)); Assistance (Art. 25(3)(c)); Attempt (Art. 25(3)(f)). K. Ambos, Treatise on International Criminal Law. Volume I: Foundations and General Part, Oxford, 2013, p. 149.
35 Art. 8(2)(b)(i) of the ICC Statute.
36 Art. 7(2)(1)(a) of the ICC Statute.
nario No. 1), in particular, acts with *dolus directus* in the first degree, in that (s)he knows and wants the harmful consequences stemming from her or his wilful conduct; while the mental state of the commander (Scenario No. 2) may be qualified as *dolus directus* in the second degree (or *dolus indirectus*), as (s)he wants to undertake a certain action, being aware that harmful secondary consequences will ensue in the ordinary course of events. These are, in fact, the “easy cases”, which are often relied upon by supporters of AWS to argue that no responsibility gap would ensue from their deployment.

Not all cases of “direct” responsibility for AWS-related crimes are so straightforward, though. Autonomy in weapons systems is likely to raise unprecedented problems of responsibility ascription, as well as to shed new light on long-standing issues. Reference is made, in particular, to: i) the individual criminal responsibility of software developers, ii) the recurrence of the many hands problems, and iii) the increased role of *dolus eventualis*.

3.1.1. *The Participation to the Crime by Software Developers*

To the extent that military technology has been evolving to the point where a software is delegated the critical functions of selecting and engaging enemy targets, a brand-new category of candidates for international criminal responsibility enters the floor: software developers. To better grasp the role which software developers might play in the commission of international crimes, it could be useful to analyse two additional scenarios:

3. Upon request by the military, a team of software developers program a class of AWS so that they will target civilian objects along with military objectives with a ratio of 1:15. This is intended to sap the enemy troops’ morale. One of these AWS is subsequently employed to this end, resulting in the bombing of a hospital.

4. An unscrupulous arms dealer, eager to exacerbate tensions in conflict-torn regions, bribes the software developers of an AWS employed by peacekeeping forces, to program the latter so that it will attack religious buildings of a certain creed. This criminal plan is completely unbeknownst to the final users of the AWS.

---


39 Crootof, cit. supra note 2, pp. 1376-1377.


41 McFarland and McCormack, cit. supra note 6.
These two examples provide a glimpse at possible modes of participation to international crimes by software developers. Scenario No. 3 introduces the most likely ground for incriminating software developers, namely purposefully “aiding and abetting” the principal perpetrator(s) by providing the means for commission of a crime.42 This also is nothing new, indeed. After all, this situation is not much different from that of the German industrialists found to be guilty for supplying to the S.S. the infamous Zyklon B gas, with the knowledge that it would have been used to exterminate the internees in concentration camps.43

Scenario No. 4, instead, presents elements of novelty. It shows that autonomy in weapons systems might determine a radical shift in the locus of control for targeting decisions. The more weapons systems are endowed with the capability to perform the critical tasks of selecting and engaging targets without human intervention, the less military commanders and operators may be held primarily responsible for targeting decisions. Responsibility, indeed, is going to be shared with, or even wholly owned by, software developers, who might well have greater weight than final users on the way AWS take targeting decisions.44 This is precisely what happens in the scenario under analysis, where software developers – along with the instigator/inducer arms dealer – would bear responsibility for committing the war crime of intentionally directing attacks against buildings dedicated to religion.45 It is worthy of note that the crime is here committed “through another person”, namely the unaware and innocent operator of the peacekeeping force, “regardless of whether that other person is criminally responsible”.46 As this scenario illustrates, in other words, AWS technology makes it a concrete possibility to reduce members of military forces to puppets in the hands of war-criminal-geeks.

Ascription of responsibility to software developers is not bereft of difficulties. In the first place, there is the evidentiary problem of finding reliable proof of their involvement in the perpetration of international crimes. It is not unreasonable to assume that software experts will be able to conceal any trace of their participation to the commission of a crime, as well as to make the latter appear a tragic

42 Art. 25(3)(c) of the ICC Statute.
44 McFarland and McCormack, cit. supra note 6, p. 370.
45 Art. 8(2)(b)(ix) of the ICC Statute.
46 Art. 25(3)(a) of the ICC Statute.
accident. Clearly, this is a technical issue, whose solution cannot be conceived of in purely legal terms, but necessarily hinges on the bolstering of existing testing and evaluation procedures.

A second problem is genuinely legal and concerns the contextual element of war crimes, as codified by the Elements of Crimes annexed to the ICC Statute, whereby the criminal conduct must take “place in the context of and [be] associated with an […] armed conflict”.47 It has been rightly noted, in this regard, that the development of the software may be finalised well before the beginning of the armed conflict during which the crime is actually perpetrated. Should that happen, prosecution for war crimes could be barred by the impossibility to establish the existence of a contextual connection between the developers’ conduct and the relevant armed conflict.48 Absent an express amendment to the Elements of Crimes by the Assembly of States Parties, one way to overcome this hurdle is to construe this norm extensively, so as to include acts taking place ante bellum. One cannot but observe, however, that this interpretative solution would be at least problematic under the nullum crimen sine lege and the in dubio pro reo principles.

Finally, AWS are going to be developed by “many teams of developers in many organizations working on a multitude of subsystems with complex interdependencies”,49 so making it arduous to trace responsibility back to individual developers. This leads us again to the “many hands” problem, which we will now focus on.

3.1.2. The “Many Hands” Problem and Joint Criminal Enterprise

As observed above, the emergence of AWS technology may entail the proliferation of “many hands” scenarios, where the high number of people involved in the causal chain makes it difficult, if not impossible, to discern who is/are responsible for AWS-related harmful events.

This is especially likely to occur during the development phase, when various teams of individuals collectively work on the software that will shape an AWS’ targeting decisions. Let us assume, for instance, that the team of software developers involved in the programming of a war crimes-prone AWS is composed of tens of indi-

49 McFarland and McCormack, cit. supra note 6, p. 384.
individuals; that most of them do not code criminal instructions, but only those parts of the software that make the AWS operational; and that all team members knew that the AWS was intended for criminal uses, although at varying levels of awareness. Who will be held responsible for the crimes eventually perpetrated through the AWS? On what count?

In fact, international criminal law is well acquainted with these kinds of problems. International crimes stand out precisely in that they “connote organized, collective wrongdoing” whose implementation may require complex apparatuses. It is no surprise, therefore, that attempts have already been made in international practice to adequately grasp the responsibility of the individuals who, within those apparatuses, somehow contributed to the commission of criminal acts. In this respect, reference should be made – first and foremost – to the doctrine of the Joint Criminal Enterprise (JCE) developed by the International Criminal Tribunal for the former Yugoslavia (ICTY).

Broadly speaking, the JCE doctrine postulates that when one or more acts are perpetrated by a plurality of individuals in the furtherance of a criminal common plan, all of them should be treated as principal perpetrators, regardless of whether they were materially involved in the commission of the crime(s) and of whether their participation took place through a criminal or a legally neutral act. The ICTY famously distinguished three kinds of JCE. The basic form (JCE I) applies to the hypotheses where participants co-operate to pursue a common criminal plan (e.g. the persecution of a religious minority). The systemic form (JCE II) is quite specific and refers to the conscious and wilful participation in a system of ill-treatment (e.g. a concentration camp). The extended form (JCE III) comes into place when “one of the perpetrators commits an act which, while outside the common design, was nevertheless a natural and foreseeable consequence of the effecting of that common purpose”. Unlike the other forms, therefore, the mental element required is *dolus eventualis*: par-

---


participants are deemed responsible also for crimes that they did not in-
tent to perpetrate, nor were part of the original plan, only on the
ground that they foresaw and accepted the possibility that those crimes
would have been committed.

The JCE doctrine – especially as construed by the ICTY’s Ap-
peals Chamber – presents unique advantages for prosecutors. First, it
greatly alleviates the burden of proof lying with the prosecution, in-
sofar as it makes it sufficient to demonstrate that the defendant is
knowingly involved in a common criminal plan to hold her/him re-
 sponsible, as a principal perpetrator, for every (foreseen and foresee-
able) crime perpetrated in the furtherance of the plan. Second, the de-
 fendant’s contribution to the common purpose does not need to be
a criminal act ex se: to be criminalised is the participation to the JCE,
while it is immaterial whether this participation takes place through
a criminal or a legally neutral act. Third, in its extended form, JCE
significantly lowers the required mens rea by applying a dolus even-
tualis standard.

Is this of any relevance for responsibility ascription in AWS-re-
lated “many hands” scenarios? Clearly, it is. The JCE doctrine is in-
deed particularly suited to cope with complicated causation problems,
viz. the kind of problems that are likely to arise in cases where an
AWS is programmed to further a criminal purpose. On the one hand,
 once the existence of a common plan is established, the JCE would
allow to incriminate all the software developers who knowingly con-
tributed to the plan, regardless of the (apparently) innocuous charac-
ter of the part of the software they were tasked to code (e.g. the one
concerning AWS’ movements on a rugged terrain). On the other
 hand, as we will see below, the mental state of those involved in the
criminal use of AWS is more likely to be qualified as dolus eventualis,
rather than as dolus directus in the first or second degree.

So far, so good. Except that the JCE doctrine, especially in its ex-
tended form, happens to be particularly controversial, since it brings
imputation of criminal responsibility very close to a criterion of “guilt
by association”, in blatant contrast with the principle of culpability.

In this respect, it is significant that the drafters of the ICC Statute

53 At the same time, the common plan/common purpose requirement generally
rules out the applicability of the JCE doctrine (even in the extended form) to “many
hands” scenarios – like the one described in sub-Section 2.2 – where the harmful
event caused by the AWS ensues from an unfortunate chain of uncoordinated cul-
pable actions.

54 See e.g. M. BADAR, “Just Convict Everyone!” - Joint Perpetration: From Tadić
to Stakić and Back Again, in Int. Criminal L. Rev., 2006, p. 293 f.
chose to follow a different approach, aimed at striking a fairer balance between the need to prosecute those taking part, in any capacity, to a common criminal plan and the demands of international criminal law, especially in terms of culpability. The main deviations from the JCE model are two. On the one hand, not all participants to the common plan are treated as principal perpetrators, but only those who dominate the causal chain of events culminating in the realisation of the common criminal plan and could therefore be qualified as (direct or indirect) co-perpetrators (Article 25(3)(a)). Lower level participants – who may include, to bring the focus back to our subject, most software developers involved in the criminal programming of an AWS – will not be punished as principal perpetrators, but on the basis of an “accessorial” common purpose liability (Article 25(3)(d)). On the other hand, the ICC Statute does not envisage dolus eventualis as a culpable state of mind.\footnote{Article 30 of the ICC Statute.} This aspect, which has an important backlash on AWS-related responsibility issues, will now be addressed.

3.1.3. Limited Predictability of AWS and the Role of Dolus Eventualis

Until now, in the possible scenarios explored, we have assumed that human agents wilfully and knowingly caused a harmful event through the AWS, or were at least virtually certain that such an event would have occurred. As AWS will be endowed with broader and broader operational discretion, this is going to be the exception rather than the rule. In most cases the human agent will not be able to foresee with certainty what the machine is going to do, with inevitable repercussions on his/her mental state in relation to the harmful event. Consider the case where an AWS operator does not intend to move an attack against civilians, nor does (s)he have the virtual certainty that this will flow from the course of action undertaken; but activates the weapon by merely accepting the risk of civilian casualties as a possible (but uncertain) consequence of the machine’s autonomous functioning. In this hypothesis, the human agent does not wish the harmful event to occur, but merely “reconciles” him or herself with the fact that something wrong could happen as a result of the course of action chosen (\textit{dolus eventualis}).\footnote{\textsc{Van der Vyver}, cit. \textit{infra} note 38, p. 63.}

This, more tenuous, volitional relationship between action and event does not have a clear status under international criminal law, especially with regard to targeting decisions.\footnote{J.D. \textsc{Ohlin}, \textit{The Combatant’s Stance: Autonomous Weapons on the Battlefield}, in \textit{Int. Law. Studies}, 2016, p. 1 f., p. 21.} On the one side, the ICTY
proved ready to accept *dolus eventualis* as a culpable state of mind in relation to the war crime of intentionally directing an attack against the civilian population, basing itself on the ICRC’s reading of the notion of “wilfully [...] making the civilian population or individual civilians the object of attack” under Article 85 of the First Additional Protocol to the Geneva Conventions (AP I). On the other side, the drafters of the ICC Statute did not include *dolus eventualis* in the general description of the mental element of the crime under Article 30 – a decision which, after some jurisprudential swinging, seems to have been construed by the Court as ruling out the relevance of *dolus eventualis* as a culpable state of mind in the ICC system.

Such a fragmented legal framework led Paola Gaeta to suggest, with specific regard to AWS’ misdoings, that war crimes committed with *dolus eventualis*, while falling outside the jurisdiction of the ICC, could be tried before national courts, which remain obliged to punish the “grave breaches” of the Protocol by Article 85 of API. The problem with this view is that it takes for granted the extensive interpretation of Article 85 provided by the ICTY and the ICRC. Yet, this reading is far from solidly grounded in international practice and raises hardly surmountable logical problems. Indeed, if the mere acceptance of the risk of civilian casualties is equated to an intentional attack against the civilian population, collateral damage should always be deemed unlawful, regardless of whether it is “excessive” under the principle of proportionality. Perhaps, that would be a step forward in the protection of civilians. But this is simply not how things stand in IHL.

58 ICTY, *Prosecutor v. Galić*, IT-98-29-T, Trial Judgment, 5 December 2003, paras. 57-58. The Trial Chamber uses, quite interchangeably, the civil law notion of *dolus eventualis* and the common law one of “recklessness”, although only the former can be properly characterized as involving intent.


63 *Ibidem*, p. 113.
The foregoing flags a warning sign concerning individual responsibility for AWS-related crimes. Autonomy in weapons systems will increase the incidence of cases where the human agent can at best formulate probability assessments as to what the weapon will actually do in the theatre of war. This is likely to create serious hurdles for responsibility ascription, given that international criminal law fails to recognise mere risk acceptance, or dolus eventualis, as a culpable state of mind in relation to direct responsibility for targeting decisions.\(^{64}\)

Awareness of this problem led many authors to look at the doctrine of superior responsibility, and in particular at the lower mens rea required thereby, as an alternative route to establish individual responsibility. The viability of this alternative will be examined in the next sub-paragraph.

### 3.2. Superior Responsibility

In the literature dealing with the responsibility problems posed by AWS, a fair amount of attention has been paid to the possibility of applying the doctrine of superior responsibility to the officer who ordered their deployment on the battlefield\(^{65}\) and to the AWS operator.\(^{66}\) According to the authoritative definition provided by the ICRC, the doctrine of superior responsibility dictates that:

> [c]ommanders and other superiors are criminally responsible for war crimes committed by their subordinates if they knew, or had reason to know, that the subordinates were about to commit or were committing such crimes and did not take all necessary and reasonable measures in their power to prevent their commission, or if such crimes had been committed, to punish the persons responsible.\(^{67}\)

\(^{64}\) C. McDougall, Autonomous Weapon Systems and Accountability: Putting the Cart before the Horse, in Melbourne J. of Int. L., 2019, p. 58 f., p. 68; Ohlin, The Combatant’s Stance, cit. supra note 57, pp. 21-29.


\(^{66}\) Malik, cit. supra note 61, pp. 635-637.

\(^{67}\) J.-M. Henckaerts and L. Doswald-Beck, Customary International Humanitarian Law, Cambridge, 2005, p. 558. This picture is slightly complicated by the ICC Statute, whose Article 28 adopts different thresholds for holding military and civilian superiors accountable. The implications for our analysis, however, are negligible. See Saxon, cit. supra note 37, p. 25.
Reliance on this doctrine signals a shift in the way AWS are conceptualised. AWS are no longer considered as tools in the hands of human agents, being instead treated as soldiers, whose misdemeanours may trigger their superiors’ responsibility.  

Sci-fi fascinations aside, what makes the doctrine of command responsibility a good candidate for filling the AWS-related accountability gap is its lower mens rea requirement. As the formula “knew, or had reason to know” makes clear, indeed, the responsibility of the superior is triggered also in cases of culpable disregard of the risk that a crime is going to be perpetrated – a mental state that is even lower than dolus eventualis. This doctrine, therefore, would render it possible to incriminate the officer who deployed an AWS (and, according to some authors, even the soldier who actually operates it) in circumstances where one should reasonably expect that it could behave harmfully, even if there is no virtual certainty that it will happen.

In fact, the application of the doctrine of superior responsibility in the case under analysis is hindered by a number of problems. On a formal level, it should not be overlooked that the doctrine, at least as it currently stands, refers to human-to-human relationships. This means that, to be operational in relation to AWS, relevant norms should undergo a contra reum interpretation by analogy, which is generally prohibited under international criminal law. A related and similarly formal problem stems from the fact that the superior’s responsibility is derivative in character, which means that it arises only if a veritable crime is perpetrated or attempted by her/his subordinates. Yet, unless one considers the weapon itself to have committed a crime (which sounds simply nonsensical), it is well possible that the only culpable conduct is that of the commanding officer who deployed the AWS, with the consequence that the doctrine would be barred from coming into play.


Admittedly, these formal problems could be tackled by updating the doctrine of superior responsibility, by way of a reform of the ICC Statute or an ad hoc international treaty, or by modifying State practice (in particular, army manuals, military penal codes) in order to adapt the customary content of the doctrine to the specificities of AWS technology. To date, the most detailed proposal in this sense was made by Peter Margulies, who worked out a revised version of the doctrine of superior responsibility, based on what he called the standard of “dynamic diligence”. The standard envisaged by Margulies would be threefold, in that it would require: i) “continual adjustments in the machine-human interface”, to be performed by adequately trained persons within a military command structure; ii) “frequent, periodic assessments of an AWS’s learning in the field, to ensure that field calculations enabled by the machine’s software are IHL-compliant”; iii) “flexibility in the parameters governing the machine’s operation, with a presumption favoring interpretability of the AWS’s outputs”.

This undoubtedly constitutes an original and intriguing attempt to adapt the doctrine of superior responsibility to the emerging features of AWS. It also offers some interesting insights as regards the notion of Meaningful Human Control (MHC), to be analyzed in Chapter V. Yet, even assuming that it gathers sufficient international support, this proposal would still be unable, in a significant number of cases, to provide a suitable solution to the accountability gap problem.

As discussed above, there are good reasons to maintain that AWS – especially if employed in unstructured environments – are prone to take (harmful) courses of action that could have not been predicted at the time of deployment, so once more raising the issue of the existence of a proper legal basis for incrimination. For superior responsibility, even in Margulies’ revised version, does not amount to a form of strict responsibility, but envisages an element of “personal dereliction”, grounded on the knowledge or, at the very least, the


74 McDougall, cit. supra note 64, pp. 77-78; Geiss and Lahmann, cit. supra note 70, p. 393; Saxon, cit supra note 37, p. 27; J.M. Beard, Autonomous Weapons and Human Responsibilities, in Georgetown J. of Int. L., 2014, p. 617 f., p. 658.

75 US Military Court, Trial of Wilhelm von Leeb et al. (The German High Command Trial), 30 December 1947-28 October 1948, The UN War Crimes Commis-
foreseeability of the subordinates’ (be them human or artificial) unlawful behaviour and the failure to take appropriate actions to prevent.76

To overcome this hurdle, Neha Jain suggested further lowering the mens rea threshold, by adopting “opaque recklessness” as a culpable state of mind in relation to AWS-related crimes.77 Under this proposal, the defendant would act culpably whenever (s)he “knows his or her conduct is risky but either fails to realize or consciously disregards the specific reasons for the riskiness”.78 This would in fact allow holding the officer accountable for AWS, “crimes”, even if (s)he was “unaware of the exact risk of harm posed by the AWS’s conduct” and even if the latter’s actions were “uncertain and unpredictable,” provided that (s)he was aware of “a substantial and unjustified risk” of some unspecified “dangerous occurrence”.79 Even indulging the conceptual “opaqueness” of the notion of “opaque recklessness”80 this proposal comes across as problematic. In dealing with accountability issues, one should always take care not to confuse the fight against impunity with “scapegoat[ing] proximate human beings”.81 Insofar as the notion of opaque recklessness stretches individual blameworthiness to the outer limits of no-fault criminal responsibility for hazardous activities, however, it leans towards the substitution of scapegoating for accountability, in breach of the principle of culpability.82

An interesting objection to the claim whereby AWS’ unpredictability would inevitably create new and troublesome accountability gaps is that human soldiers are far more “autonomous” than AWS, and can
act in no less unforeseeable ways, e.g. by disobeying the superior’s
orders. While this is certainly true, this objection does not take into
due account that, in case of misconduct by human soldiers, the su-
perior may (and indeed must) exercise her or his punitive power over
them. Since this option is clearly precluded when the “wrongdoers”
are artificial agents, to which “punishment” is a meaningless concept,
in case of AWS’ unpredictable misdoings the officer ordering their
deployment would have no means to take appropriate measures in
response to the unwanted behaviour.

This brings to the limelight a further problem with the applica-
tion of the doctrine of superior responsibility, namely that the offi-
cer deciding the deployment of AWS is very unlikely to be able to
exercise an “effective control” over their activities. It has been noted,
in this regard, that there is a world of difference in the ways a mil-
itary officer might influence human and non-human behaviours. Hu-
man soldiers are subject to an unceasing training process, aimed at
instilling awareness of their obligations under IHL, which contin-
ues even when they are fielded into a combat scenario, inter alia,
by way of the establishment of a “command culture that emphasizes […]
humanitarian constraint”. By contrast, an autonomous weapon is
“instructed” to respect IHL during the development and testing phase,
so well before its deployment in combat scenarios and, more rele-
antly, with no opportunity for the fielding officer to influence its
behaviour by way of training or military discipline.

On these assumptions, Geoffrey S. Corn maintains that account-
ability should be burdened on the procurement officer, who may play
– especially in the development phase – a far more incisive role in
“training” AWS to comply with IHL, by equipping them with all the
required cognitive and evaluative capabilities, and assuming responsi-
bility thereof. This is undoubtedly a useful proposal, which could
help adapt the law of superior responsibility to AWS’ specific fea-

85 See, in particular, the obligation envisaged by Art. 87(2) AP I.
86 CORN, cit. supra note 83, pp. 222-223.
87 Ibidem, p. 224.
tures. Yet, it does not help overcome the predictability issue set out above.89 On the contrary, this difficulty would be only exacerbated by an anticipation of the locus of responsibility that ends up identifying, as responsible superiors, individuals who are temporally and physically detached from the battlefield.90

In sum, the doctrine of superior responsibility – as it currently stands – is formally inapplicable to the relationship between an AWS and the officer deploying it. Moreover, even if the doctrine were reformed so as to include also non-human subordinates, it would remain unable to offer a viable solution in many scenarios involving AWS-related accidents, because of the unpredictability inherent in autonomous decision-making.

One may wonder, at this juncture, whether the hypotheses analysed so far actually differ from the case where a harmful event is caused by the unpredictable mishap of a non-autonomous weapon (e.g. a laser-guided missile that, because of a mechanic breakdown, hits a hospital). Even in this case, none of the people involved in the accident (the operator who launched the missile, his/her superior, and so on) could be accused of intentionally attacking a civilian object. On reflection, it is not self-evident why this case should be distinguished from the one where an AWS goes awry and attacks a hospital: in both hypotheses, there is a deviation from the ordinary course of events that is unwanted, unknown and unforeseen, thereby excluding the responsibility of the proximate human agents.91 However, there is an important element of differentiation, which makes the unpredictability of AWS far more troublesome. In the case of non-autonomous technology, a harmful mishap just represents an unfortunate exception within a decision-making process that is still entirely in the hands of human beings. Such state of things is reversed in the case of AWS, which are not entrusted merely with the execution of human decisions, but rather enjoy “discretionary autonomy”92 in the performance of an entire chunk of the targeting process (in particu-

89 McDougall, cit. supra note 64, p. 80; Saxon, cit. supra note 37, p. 28.
90 For the same reason, the incrimination of procurement officers for AWS’ misdoings is thwarted by the temporal scope of the contextual element of war crimes, which – as seen above with regard to the responsibility of software developers (text accompanying notes 47-48) – requires that the criminal conduct takes “place in the context of and [be] associated with an […] armed conflict”.
92 Liu, Refining Responsibility, cit. supra note 69, p. 328.
lar, the critical functions of selecting and engaging targets). That being the case, accountability gaps will no longer be the exception, but the rule.

4. State Responsibility

A survey of the scholarly works on AWS shows that the issue of State responsibility has received scant attention. The reason for this should be probably sought for in the widespread perception that ascription of responsibility to States in these cases would be unproblematic. It is all the more telling, in this regard, that Human Rights Watch decided not to deal with this issue in its report on AWS-related accountability gaps on the grounds that “state responsibility for the unlawful acts of fully autonomous weapons could be assigned relatively easily to the user state”.

The argumentative basis for this conclusion may be summarized as follows. Under general international law, State responsibility arises whenever an action or omission that is attributable to the State (subjective element) is in breach of an international obligation of that State (objective element). As is well known, these conditions for State responsibility have been codified by Article 2 of the Draft Articles on the Responsibility of States for Internationally Wrongful Acts, adopted in 2001 by the International Law Commission (ILC). Significantly enough, the 2001 Draft Articles do not include State’s “fault” as a condition for international responsibility, which seems to suggest that, in order to hold a State to account for an internationally wrongful act, the ascertainment of the mental element would not be required.

The latter understanding is pivotal to maintain that, as far as State responsibility for AWS’ misdoings is concerned, all the above quandaries concerning the establishment of a culpable state of mind would not come into play. Indeed, once it is found that an AWS has been...

93 This led Hin-Yan Liu to suggest that the very “introduction of such irresponsible systems onto the battle space” should be criminalized, by drawing an analogy with the war crime of recruitment and use of child soldiers (ibidem, p. 344).
activated by a State organ, or by someone acting on the State’s behalf (subjective element), and that it carried out targeting activities that are in breach of international law (objective element), State responsibility would ensue for the harm caused. Nor, it is added, could an unexpected malfunction be invoked as a force majeure precluding wrongfulness: under the law of State responsibility, force majeure cannot excuse a breach of a peremptory norm of international law, a category that certainly includes fundamental norms of international humanitarian law (IHL), such as the principles of distinction and proportionality, and the prohibition of arbitrary deprivations of life under international human rights law (IHRL).

Such unconditional optimism is in fact unsubstantiated. Inaccurate, specifically, is the idea whereby State responsibility for IHL violations (or, where applicable, for arbitrary deprivations of life under IHRL) would not require the ascertainment of a mental element. While it is true that “intent” or “fault” do not count as constitutive elements of internationally wrongful acts under “secondary” customary norms on State responsibility, it is also generally acknowledged – including by the ILC itself – that the relevance of a culpable mental element may be envisioned by the “primary” norm whose alleged violation is at stake.

This is precisely the case of rules governing targeting decisions under IHL and IHRL. The prohibition to direct an attack against the civilian population, to begin with, arguably presupposes an element of intentionality. An attack that, because of an unexpected technical problem, causes unwanted civilian losses does not qualify in itself as a breach of the principle of distinction, because it cannot be said to have been directed against civilians. As we noted in Chapter II, the

96 It is indeed a well-established principle of IHL that a State “shall be responsible for all acts committed by persons forming part of its armed forces”. See Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (AP I), 8 June 1977, Art. 91.
primary norm on proportionality incorporates a mental element as well, insofar as it revolves around the expectations and anticipations of those in charge of the attacks and their conformity with a standard of reasonableness. Turning to IHRL, it is noteworthy that the Human Rights Committee defined “deprivation of life” as “an intentional or otherwise foreseeable and preventable life-terminating harm or injury, caused by an act or omission”, so implying that the prohibition of arbitrary deprivations addresses State conducts that involve intentionality (or at least awareness) on the part of its agents.

Therefore, every time the establishment of individual criminal responsibility is hindered by AWS’ unpredictability and thus by the lack of mens rea, State responsibility is not necessarily less problematic to ascertain. And this because in the absence of State’s fault the principle of distinction and proportionality under IHL, as well as the prohibition on arbitrary deprivations of life under IHRL, are simply not violated, with the consequence that the objective element for the purposes of State responsibility (i.e. a breach of an international obligation of that State) is not fulfilled. In this respect, there is no point in observing that force majeure cannot excuse a violation of peremptory norms of international law, since such violation – as we have just noted – cannot be deemed to have occurred in cases of harm caused by AWS’ unpredictable behaviour.

State responsibility, in other terms, does not offer a legal panacea to resolve all accountability issues raised by autonomy in weapons systems. Significantly enough, the authors who look more optimistic in this respect conceive of State responsibility for AWS’ misdoings as a strict liability regime. While this is clearly not the case de lege

101 For a similar conclusion, see Geiss and Lahmann, cit. supra note 70, pp. 386-387.
102 This finding has implications for the recent scholarly attempt to bridge the accountability gap by resorting to a notion that is half-way between individual criminal responsibility and State responsibility, namely that of individual administrative accountability for non-criminal violations of IHL (L.A. Dickinson, Lethal Autonomous Weapons Systems: The Overlooked Importance of Administrative Accountability, in R.T.P. Alcala and E. Talbot Jensen (eds), The Impact of Emerging Technologies on the Law of Armed Conflict, New York, 2019, p. 69 f.). Once State responsibility is ruled out, indeed, so will be individual administrative accountability.
103 D.M. Hammond, Autonomous Weapons and the Problem of State Accountability, in Chicago J. of Int. L., 2015, p. 652 f., pp. 668-671 (who seems to consider such a regime to be the actual state of the law); Crootof, cit. supra note 2, pp.
lata,\textsuperscript{104} the introduction of a strict liability regime for civilian damages – along the lines, for instance, of the 1972 Convention on the International Liability for Damage Caused by Space Objects\textsuperscript{105} – undoubtedly constitutes a promising option in a \textit{de lege ferenda} perspective, with the caveat that such regime should be shaped so as to take into due account that, unlike space objects, weapons systems “are by definition designed to lawfully cause certain damage”.\textsuperscript{106}

But what about the law as it currently stands? The above remarks, in fact, do not warrant the conclusion that State responsibility would be of no avail in the quest to reduce AWS-related accountability gaps. It should be observed, in this respect, that IHL and IHRL norms obliging States to take precautionary measures to avoid or minimize unlawful harm, while still envisaging an element of fault, set the bar at a lower level than international criminal provisions. To the extent that they envisage – as we saw in Chapter II – an obligation to “do everything feasible” to prevent infringements of the principles of distinction and proportionality, as well as of the right to life under IHRL, they arguably embody a due diligence standard,\textsuperscript{107} whose violation may be determined by a negligent conduct by those deciding or planning the use of an AWS, both in armed conflicts and in peacetime.\textsuperscript{108}


\textsuperscript{105} See Convention on International Liability for Damage Caused by Space Objects, 29 March 1972, Art. II (“A launching State shall be absolutely liable to pay compensation for damage caused by its space object on the surface of the earth or to aircraft flight”). See also Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies, 27 January 1967, Art. VII (“Each State Party to the Treaty that launches or procures the launching of an object into outer space […] and each State Party from whose territory or facility an object is launched, is internationally liable for damage to another State Party to the Treaty or to its natural or juridical persons by such object or its component parts on the Earth, in air space or in outer space, including the Moon and other celestial bodies”).

\textsuperscript{106} Geiss, cit. \textit{supra} note 99, p. 118.


\textsuperscript{108} This is not to say that “subjective" negligence by State organs (or individuals acting on State’s behalf) should be considered as a \textit{sine qua non} condition for breaching the due diligence principle, but rather that subjective attitudes of fault may be
An additional advantage of State responsibility lies in the fact that IHL and IHRL norms envisioning the positive obligation to take precautionary measures crucially differ from the doctrine of superior responsibility in that their application does not presuppose the commission of international crimes by human subordinates. This entails that – already de lege lata – the failure to take proper risk mitigation measures resulting in civilian losses (or otherwise unlawful deprivations of life) is likely to trigger State responsibility, even when it is not possible to establish the direct criminal responsibility of AWS’ operators. Furthermore, the application of a due diligence standard would make it possible to hold a State responsible in “hard” many hands scenarios, as the one described in sub-Section 2.2., where the harmful event caused by the AWS is the precipitate of a chain of uncoordinated faulty (but not intentionally criminal) actions.¹⁰⁹

5. Corporate Liability

It has been sometimes suggested that the corporations involved in the programming and manufacturing of AWS could be held liable for the latter’s misdoings. Since corporations are far from having an established status in international law,¹¹⁰ the issue of their responsibility in the AWS debate is mainly couched in domestic legal terms,¹¹¹ which puts it somehow outside the scope of the present analysis. Nevertheless, an inquiry upon the accountability gap problem would be incomplete without a mention of corporations, with specific regard to tort liability for international crimes, product liability and strict liability.

5.1. Tort Liability for International Crimes

The possibility to hold corporations responsible for international crimes...
crimes has been at the centre of a broad and well-known academic and jurisprudential debate, which is far from settled and certainly cannot be reproduced here.\textsuperscript{112} For the sake of the present discussion, therefore, we will assume that corporations can be regarded as recipients of the norms prohibiting international crimes. Given the lack of international forums having jurisdiction over corporate crimes,\textsuperscript{113} however, we must limit ourselves to gauging this question from the vantage point of national courts, where the issue is likely to be addressed through the lens of tort liability.\textsuperscript{114}

Reliance on corporate tort liability in cases concerning international crimes committed through AWS would help to reduce the responsibility gap stemming from the “many hands” problem. As we have seen, if the perpetration of an international crime ensues from the joint action of a large number of individuals, e.g. a team of software programmers, it could prove highly challenging to ascertain individual (criminal) responsibility for it, having particular regard to the \textit{mens rea}. If responsibility is traced to the corporation as a whole, instead, these conceptual hurdles could be addressed, at least in part, by resorting to judicial made doctrines aimed at establishing corporate criminal intent, such as the “collective knowledge” doctrine, followed by some US courts, which assumes that the corporation’s knowledge relating to a certain harmful conduct is constituted by the aggregate knowledge of all its employees.\textsuperscript{115}

Although this is an undoubted advantage, even this path to accountability is strewn with obstacles. To say the least, corporate tort liability for international crimes has been clearly asserted – and not without contestations – only by US courts, on the basis of a very pe-


\textsuperscript{113} See, above all, Article 21 of the ICC Statute, which rules out the Court’s jurisdiction over “legal persons”, including corporations. As carefully demonstrated by Andrew Clapham, however, this limitation was due to reasons other than the their (alleged) lack of legal personality under international criminal law (A. Clapham, \textit{The Question of Jurisdiction Under International Criminal Law Over Legal Persons: Lessons from the Rome Conference on the International Criminal Court}, in M.Y. Kamminga and S. Zia-Zarifi (eds), \textit{Liability of Multinational Corporations Under International Law}, The Hague, 2000, p. 139 f.)


\textsuperscript{115} \textit{United States v. Bank of New England}, 821 F.2d 844 (1st Cir. 1987)
cular piece of legislation, the 1789 Alien Tort Statute, and it is far from sure that it will ever develop in other legal systems. Furthermore, if the action for damages touches upon the legality (or even the appropriateness) of the military choices by the forum State, the risk is high that the court will decline jurisdiction on the basis of judge-made doctrines aimed at insulating issues of defence and foreign affairs from judicial scrutiny (political question, non-justiciability, atto politico, acte de gouvernement doctrines). Finally, and more importantly, corporate tort liability for international crimes (and the related “collective knowledge” doctrine), while providing a viable solution to the many hands problem, does not take away from the fact that a mens rea should be established. Accordingly, problems stemming from AWS’ inherent unpredictability are left unresolved.

5.2. Product Liability

It has been observed that “[p]roduct liability laws are largely untested in robotics”. AWS make no exception, indeed, although some have alluded to product liability as a feasible alternative to individual and State responsibility. Since the matter is generally gov-

116 B. Stephens, Translating Filártiga: A Comparative and International Law Analysis of Domestic Remedies for International Human Rights Violations, in Yale J. of Int. L., 2002, p. 1 ff. See, however, the recent judgment by the Supreme Court of Canada, somehow opening up the doors to actions in damages against corporations for the breach of customary international norms, including those prohibiting international crimes (Nevsun Resources Ltd. v. Araya, 2020 SCC 5). For a comment, see J. Hughes Jennett and M. Parcasio, Corporate Civil Liability for Breaches of Customary International Law: Supreme Court of Canada Opens Door to Common Law Claims in Nevsun v Araya, in EJIL:Talk!, 29 March 2020. The applicability to corporations of another piece of US legislation usually relied on in human rights litigations, the Torture Victim Protection Act (TVPA), seems to have been impliedly ruled out by the US Supreme Court in Mohamad v. Palestinian Authority, 566 U. S. (2012). For the view whereby, after Mohamad, there would still be room for invoking the TVPA against corporations, see the thought-provoking reflection by F. Seatzu, Speculating on the Future of the Torture Victim Protection Act (TVPA) after Mohamad and Kiobel, in Inter-American and European Human Rights J., 2016, p. 23 ff.


erned by domestic law (with the notable exception of European Directive 85/374/EEC), a comprehensive exposition of the problem is not practicable here. Hence, we will just put forth some observations of general character.

Despite the undoubted (and sometimes profound) differences in the various legal systems of the world, product liability regimes generally share a plaintiff-friendly nature, in that they set the fault element threshold to a standard of negligence and, in certain hypotheses, shift the burden of proof from the damaged party to the defendant company, which will have to demonstrate the applicability of one of the excuses provided by the law. If we consider the above-discussed difficulties concerning the proof of the fault element in AWS-related accidents, it would appear clear that this feature of product liability regimes is likely to provide a major advantage to the damaged party.

Even in this case, however, legal and conceptual hurdles stand in the way of redress for victims. First and foremost, although the evidentiary regime might be more favourable to the plaintiff, product liability remains associated with negligent behaviour by the defendant corporation. However low, this standard requires that the malfunction complained of is at least foreseeable by the manufacturer, the designer or the programmer. As repeatedly underlined, however, this is exactly what autonomy in weapons systems is bound to rule out in most cases of harmful events. And in fact some authors came to

---

120 Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products. Quite remarkably, an earlier attempt to regulate this matter at the international level, i.e. the Council of Europe European Convention on Products Liability in regard to Personal Injury and Death of 27 January 1977, turned out to be a failure, being signed by only 4 States and ratified by none.


122 *Ibidem*, para. 2(b) (“[A product] is defective in design when the foreseeable risks of harm posed by the product could have been reduced or avoided by the adoption of a reasonable alternative design”); Directive 85/374/EEC, Article 7 (“The producer shall not be liable […] if he proves: […] e) that the state of scientific and technical knowledge at the time when he put the product into circulation was not such as to enable the existence of the defect to be discovered”).

123 A. Finn and S. Scheding, *Developments and Challenges for Autonomous Unmanned Vehicles. A Compendium*, Berlin/Heidelberg, 2010, p. 183 (highlighting the difficulty “to functionally establish or fully test the response of the system to repeatable or verifiable system stimuli”); ICRC, *Autonomy*, cit. * supra* note 21, p. 12 (“it is not possible to test all the potential inputs and outputs of the system for all
question the very possibility to characterise robots, including but not limited to AWS, as “products” in a legal sense and, accordingly, to treat their harmful decisions as “defects”. For, to the extent that a robot is endowed with learning capabilities and autonomous decision-making power, “it is hardly plausible that [it] was defect[ive]; it did what it was supposed to do: It reacted to new inputs and adapted its behaviour – thus the machine is not defective as such.”

Besides, it is worth recalling that, before the courts of the country that is likely to be among the largest producers and users of AWS, i.e. the United States, product liability lawsuits would be inexorably barred due to the so-called “government contractor” defence. Under this doctrine, issued by the US Supreme Court in the Boyle case, military contractors are deemed immune from product liability whenever the product at stake (including a weapons system) has been manufactured according to specifications provided or approved by a federal or state government agency – quite a likely occurrence in the case of AWS, due to the delicate and complex activities they are meant to carry out. The scope of this defence, moreover, has been subsequently expanded by federal courts so as to make it applicable in virtually every case concerning AWS’ misdoings. Reference is made to the approach followed by the Court of Appeals for the Ninth Circuit in Koohi v. United States et al., which – remarkably enough – concerned a forerunner of modern AWS, the Aegis air defence system. On that occasion, the Court held this doctrine to shield contractors from claims for damages “arising out of the combatant activities of the military or naval forces [...] during time of war” – basically what AWS are expected to do in the near and mid-future.

5.3. No-Fault Liability

The problems emerged above could arguably be fixed by enacting a no-fault liability regime. As is known, the introduction of a regime of this kind might be appropriate when an inquiry as to the fault el-
ement is particularly complex. Moreover, both at the national and international level there is a tendency to establish regimes of this kind in relation to the performance of activities that may prove particularly dangerous because of their very nature or because of the means normally used to carry them out. No-fault liability is often resorted to, for example, in the field of environmental law. One could recall, in this respect, the International Convention on Civil Liability for Oil Pollution Damage (CLC), whose Article III states that:

The owner of a ship at the time of an incident [...] shall be liable for any pollution damage caused by oil which has escaped or been discharged from the ship as a result of the incident.

The rationale behind this kind of rules lies in the need to relieve the community from the cost burden of environmental damage, by transferring it to the person or entity taking advantage from the activity that caused it.

The adoption of such a liability regime would be certainly advisable in relation to the damages caused by AWS. After all, their use in armed conflicts could be properly characterised as “ultra-hazardous”, as “it involves a risk of serious harm that cannot be eliminated, even if utmost care is exercised”. In a de lege ferenda perspective, therefore, one could think of a treaty establishing, along the lines of the CLC, that corporations involved in the development and production

128 See, also for further references, Crootof, cit. supra note 2, p. 1395.
130 Ibidem, p. 1148. This constitutes an embodiment of the so-called “polluter pays” principle enshrined in Principle 16 of the Rio Declaration on Environment and Development, 12 August 1992, UN Doc. A/CONF.151/26 (Vol. I). On this principle, see A.E. Boyle, Making the Polluter Pay? Alternatives to State Responsibility in the Allocation of Transboundary Environmental Costs, in F. Francioni and T. Scovazzi (eds), International Responsibility for Environmental Harm, London, 1991, p. 363 f.; F.M. Palombino, Il significato del principio “chi inquina paga” nel diritto internazionale, in Rivista giuridica dell’ambiente, 2003, p. 871 f. Another example of corporate strict liability regime under international law is provided by a soft law instrument, the 2006 ILC Draft Principles on the Allocation of Loss in the Case of Transboundary Harm Arising out of Hazardous Activities, whose Art. 4 establishes that States “should take all necessary measures to ensure that prompt and adequate compensation is available for victims” including “the imposition of liability on the operator or, where appropriate, other person or entity. Such liability should not require proof of fault”.
131 Crootof, cit. supra note 2, p. 1395.
of AWS are strictly obliged to compensate the civilian damages caused by them.

6. Conclusions: The Inevitability of Accountability Gaps and Its Implications for the Legality of Autonomous Weapons Systems

In Furundžija, the ICTY’s Trial Chamber concluded a thorough review of the international legal regime outlawing torture by asserting that “[n]o legal loopholes have been left”.132 Should it have reviewed the legal consequences of AWS’ deployment, it would have surely passed a less optimistic judgement. As we have here attempted to demonstrate, autonomy in weapons systems is likely to magnify the proliferation of legal loopholes, making it difficult – when not impossible – to identify a responsible person or entity in case of harmful events.

On the one hand, indeed, the enforcement of individual criminal responsibility would be precluded in a fair number of hypotheses, partly because of technical-legal problems concerning the formulation of international criminal norms (notably, the temporal scope of the contextual element of war crimes and the derivate character of command responsibility), partly because of the difficulties inherently related to machines’ autonomous decision-making, viz. unpredictability and the “many hands” problem. On the other hand, State and corporate responsibility can fill individual accountability gaps only to a limited extent: they both allow overcoming the “many hands” problem and are less strict as to the mens rea requirement, but cannot provide a comprehensive solution to the issues associated with AWS’ unpredictability, not to speak of the procedural hurdles that could make it particularly difficult for individuals to actually obtain judicial redress.133

Against this backdrop, proposals aimed at adjusting existing law in order to address the accountability gap problem are generally of two kinds. In the first place, it has been suggested to reform the law of superior responsibility with a view to expanding its scope (and, relatedly, the lower mens rea associated thereto) and making it applicable to human-to-weapon relationships, through a fictio iuris equat-

133 See above the text accompanying notes 117 and 125-127. With regard to the procedural hurdles likely to thwart action for compensation against States, see Malik, cit. supra note 61, pp. 639-640.
ing AWS to human soldiers. As has been observed multiple times throughout this Chapter, however, lowering the required \textit{mens rea} does not suffice to address the major source of accountability gaps: namely AWS’ unpredictability. This explains why the best candidate to effectively “close” the accountability gap is represented by the second kind of proposals, i.e. the adoption of an international regime of no-fault liability for States and/or corporations involved in the development and use of AWS.

Looking at State and corporate (strict) liability as a substitute for individual criminal responsibility seemingly rests on the assumption that the various responsibility regimes provided by international law are ultimately fungible and replaceable with each other. Yet, this assumption would completely overlook the “complementarity” among these regimes, which is rooted in the distinction between the “predominantly reparational aspect of state responsibility and the punitive character of criminal law proceedings against individuals”\textsuperscript{134} Individual responsibility, therefore, should not too easily removed – \textit{de jure} or \textit{de facto} – from the international legal landscape for the reason, well-expressed in a celebrated \textit{dictum} by the Nuremberg Tribunal, that international crimes “are committed by men, not by abstract entities, and only by punishing individuals who commit such crimes can the provisions of international law be enforced”.\textsuperscript{135}

There is, of course, a third option to “plug” the accountability gap, which is the one propounded by those advocating for Meaningful Human Control (MHC) over weapons systems, namely to preserve a kind and degree of human control over the critical functions of se-


\textsuperscript{135} International Military Tribunal, \textit{The Trial of German Major War Criminals}, Judgment, 1 October 1946. Proceedings of the International Military Tribunal sitting at Nuremberg, Germany, Part 22 (22nd August, 1946 to 1st October, 1946), p. 447. At the 2016 CCW Informal Meeting of Experts the Norwegian delegate put the problem in very similar terms: “Another intrinsic challenge with fully autonomous weapons would be ensuring individual […] responsibility for unlawful acts in times of armed conflict. This is a cornerstone of modern international law. Without accountability, deterring and preventing international crimes becomes all that much harder” (Statement at the Panel on “Challenges to International Humanitarian Law”).
lecting and engaging targets, which is capable of ensuring a just ascription of individual responsibility, instead of thwarting it.

The soundness of this conclusion has been recently questioned in a thought-provoking article by Carrie McDougall. Her main claim is that, even if one accepts (as she seems to) that autonomy in weapons systems is bound to create gaps in the ascertainment of individual criminal responsibility under existing law, "accountability is not the right lens through which to examine the issue of whether AWS should be prohibited or regulated".137

To substantiate her criticism to the prominence of accountability in the AWS debate, McDougall firstly recalls the instrumental character of international criminal norms with respect to the other branches of international law (IHL, international human rights law, jus ad bellum), which entails that the criminalization of the use of certain weapons is dependent on the pre-existence of prohibitions "well entrenched in either conventional or customary international law and broadly supported by the international community". Hence, according to the Author, arguments relying on the risk of "accountability gaps" to justify a ban on AWS are legally flawed in that they unduly overturn this logical relationship ("a conduct is criminalized because it is prohibited"), by maintaining that the use of AWS should be prohibited because it will not be possible to criminalize someone in case of unwanted outcomes. In fact, she argues, it is for policymakers to deliberate whether and to what extent international criminal law should be amended to avoid accountability gaps by ensuring that someone in the decision-making chain is held responsible for the unpredictable harms caused by AWS: this, however, is immaterial as to whether autonomy in weapons systems should be prohibited or otherwise regulated.140

McDougall’s objections are not entirely convincing. To start with, it is difficult to agree with her view whereby international criminal law may be “malleable” to the point of allowing States settling the “accountability gap” problem by establishing that, in case an AWS acts in breach of IHL, “one or more proximate humans should always be held criminally responsible”, i.e. – it would seem – even in the absence of a culpable state of mind. As a matter of fact, States

136 McDougall, cit. supra note 64.
137 Ibidem, p. 82.
138 Ibidem.
139 Ibidem, pp. 84-85.
140 Ibidem, pp. 85-86.
141 Ibidem, p. 86.
would hardly deem this solution viable. Moreover, even if international criminal norms were actually enacted to this effect, with all likelihood they would be interpreted by competent courts so as to make them compatible with the principle of culpability, which is a backbone of international criminal law and is acknowledged in most national constitutional traditions.142 Contrary to what McDougall’s appears to imply, in other words, international criminal law cannot expand in whatever direction policy-makers wish. More importantly, it is precisely because of one of the bulwarks against such an unconstrained expansion, namely the principle of culpability, that autonomy in weapons systems is unavoidably conducive to individual accountability gaps.

McDougall, however, is correct in denying that AWS’ proneness to create accountability gaps is sufficient to infer ipso iure their unlawfulness.143 Yet, the point made in this Chapter is slightly different. The analysis carried out so far was indeed aimed at showing that the more weapons systems will become autonomous in the performance of targeting decisions, the lesser will be the role of individual criminal responsibility (and, thus, of international criminal law) in governing the use of armed violence. This would be all the more true should the international community opt for filling accountability gaps by enacting compensation mechanisms based on a strict liability regime for States (and/or corporations) involved in the development and use

142 See K.J. Heller and M.D. Dubber (eds), The Handbook of Comparative Criminal Law, Stanford, 2011, pp. 24 (Argentina), 58 (Australia), 106-107 (Canada), 148-149 (China), 187 (Egypt), 216 (France), 261 (Germany), 294-295 (India), 326-327 (Iran), 357-358 (Israel), 400-401 (Japan), 420 (Russia), 463 (South Africa), 491 (Spain), 536-537 (United Kingdom: with some qualifications), 573-576 (United States), to which one must add Art. 27(1) of the Italian Constitution. This entails that domestic courts could refuse to apply international criminal norms establishing a strict responsibility regime by invoking the constitutional supremacy of national fundamental principles. On this issue, see F.M. Palombino (eds), Duelling for Supremacy. International Law vs. National Fundamental Principles, Cambridge, 2019.

of AWS. The latter, notably, would bring about a veritable paradigm shift\textsuperscript{144} towards an “insurance model” in the regulation of armed conflicts (and armed violence in general), which would prevent international criminal law from playing its crucial function of “pronounce[ing] the wrongfulness of actions that harm the interests of the international community as a whole”\textsuperscript{145}.

Such a development would be in itself neither prohibited nor conceptually wrong. But that would be a radical departure from the international legal order built up starting from the end of World War II. Whether there is, at the international level, an actual willingness to take this step is not just an issue of policy, as McDougall suggests. It has rather to do with the underlying ethical convictions of the international community on the issue of human responsibility for armed violence, in the face of the technical possibility to increasingly replace human judgment and decision-making with artificial ones. This question, which relates more in general to the stance of the international community on human-machine relationship in ethically and legally sensitive areas, will be addressed in the next Chapter.

\textsuperscript{144} Geiss and Lahmann, cit. \textit{supra} note 70.