

The Human-Weapon Relationship in the Age of Autonomous Weapons and the Attribution of Criminal Responsibility for War Crimes

Dr. Marta Bo*

I. Introduction

This paper is concerned with the implementation of automation and artificial intelligence in weapon technology and with its impact on the ascription of criminal responsibility. Many have focused on the problem of ascribing legal responsibility from the crimes committed via the use of fully autonomous weapons and on the difficulties of ascribing responsibility when machines attain a certain level of autonomous decision making capacity (the so-called responsibility gap).¹ However, before considering the responsibility gap that could arguably arise if and when fully autonomous weapons are developed and deployed, it is important to focus on a current (and likely future) feature of warfare and targeting: human-machine teaming. Full autonomy in weapon systems is not the only scenario that may affect the ascription of criminal responsibility. How automation and autonomy in *some functions* of weapons (leading to semi-autonomous weapons) bear on the ascription of criminal responsibility and on the culpability of the perpetrator have not yet been acknowledged. This paper discusses the effects on human decision making of automation of some functions of a weapon system. For the purposes of this paper, these effects are equated to the impact that the implementation of artificial intelligence in weapons generates (and with AI these effects could be even exacerbated). Therefore, reference to an automated system or semi-autonomous weapons, for the purpose of this paper, also covers artificial intelligent systems.²

* Researcher at the T.M.C. Asser Institute in The Hague (The Netherlands), post-doctoral researcher at the Graduate Institute for International and Development Studies in Geneva (Switzerland). The author would like to sincerely thank Taylor Woodcock for her instrumental research assistance on the topic and editorial work on the article. The usual disclaimers apply.

¹ Thompson Chengeta, 'Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law', (2016) 45 Denv. J. Int'l L. & Pol'y 1 1; Robert Sparrow, 'Killer Robots' (2007) 24 Journal of Applied Philosophy 62; ICRC, 'Autonomous Weapon Systems Implications of Increasing Autonomy' (2016) 44.

² In relation to the difference between automation and autonomy, Marra and McNeil affirm that: 'although the terms "automation" and "autonomy" are alike in that both "refer to processes that may be executed independently from start to finish without any human intervention", the processes'

The growing implementation of automation and artificial intelligence into some functions of weapon systems increasingly requires humans to take decisions in tandem with machines. In particular, the use of artificial intelligence, such as machine learning, as a decision making aid in the targeting process is a prime of example of this currently under development. The time has come to focus on and pay more attention to the impact on human decision making flowing from the implementation of automation in weapons technology.

This paper will take into consideration another crucial, but overlooked, starting point for assessing the problems surrounding the ascription of responsibility for crimes committed via the deployment of autonomous weapons: the law of the war crimes committed during the conduct of hostilities, so-called conduct of hostilities crimes. Analyzing the features of the relationship between humans and semi-autonomous weapons against the legal framework of conduct of hostilities crimes shows that the determination of the *mens rea* element will be problematic due to the stringent subjective requirements for these crimes. However, this analysis will further show that ‘risk taking behavior’ is an important area to capture in relation to human-machine interaction during targeting, and for establishing *mens rea* for war crimes. It follows that an inquiry into whether and, if so, how automation influences the awareness of risks and risk propensity is in order.

Reflecting on human decision making in the age of semi-autonomous weapon systems is crucial for criminal law. Human decision making is the process underlying how human will, knowledge and assessment of and subjective attitudes towards risk are shaped. These are all the various components of different *mens rea* requirements (i.e. direct intent, indirect intent, *dolus eventualis*, recklessness or negligence). The fulfillment – or lack thereof – of these components leads to the determination of criminal responsibility for war crimes. How is human decision making affected by automation? How does the implementation of automation affect risk assessments and human predisposition towards those risks? Is the use of automated systems more

differences are more revealing than their similarities. Automated systems are not self-directed. They also lack decision-making capability. Instead, these systems simply have "the capacity to operate without [human intervention]". By contrast, autonomous entities are capable of being "independent in the establishment and pursuit of their own goals", see William C Marra and Sonia K McNeil, 'Understanding the Loop: Regulating the Next Generation of War Machines' (2013) 36 Harv . J . L . & Pub . Pol ' y 1139, 1150.

conducive to risk taking behavior? And more specifically: how is the knowledge and the assessment of risks of committing a war crime affected by automation?

In order to inquire into the problem of determining responsibility for war crimes in committed with semi-autonomous weapons, Section II highlights the importance of human-machine teaming as a present and future feature of warfare. It also maps out some developments in weapon systems that could serve as an example to understand new features of the human-weapon relationship. Section III underlines the specific and fundamental function under international law of determining individual criminal responsibility for serious violations of international humanitarian law (IHL). The importance of ascribing individual criminal responsibility in multi-layered decision making environments, such as the military, will be underscored.

Section IV and V.I will describe the normative and descriptive functions of the principle of culpability. Section V.II will offer a comparative law analysis of the different *mens rea* requirements and will provide the necessary notions to approach the debates surrounding the specific *mens rea* standards for conduct of hostilities crimes (Section V.II), and in particular of the war crime of attacking civilians. Finally, Section VI will attempt to make an examination of the cognitive environment where human agents using semi-autonomous weapon systems operate, focusing on the cognitive phenomena that could bear on human decision making and on the shaping of the *mens rea* of the operator of semi-autonomous weapon systems.

II. Human-Machine Teaming as Present and Future Feature of Warfare

For the purposes of the present discussion, autonomous weapons are those that once activated, select and engage targets whose engagement was not previously decided by a human.³ Most weapon systems today can be defined as semi-autonomous. In these systems, autonomy is implemented in one or more targeting functions (i.e. the search, identification, tracking or prioritization of targets) and the human operator retains the decision to engage specific targets.⁴ According to Scharre, the human operators of contemporary semi-autonomous weapon systems carry out three functions during the

³ Paul Scharre and Michael C Horowitz, 'Autonomy in Weapon Systems' (2015) 16.

⁴ Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (WW Norton and Company 2018) 44.

targeting process, at times simultaneously. These roles include the ‘essential operator’ responsible for the final engagement step; the ‘moral agent’ who decides whether engagement is appropriate; and finally the ‘fail-safe’ responsible for intervening in the operation of the weapon if deemed necessary.⁵ [SEP]

Moreover, thirty states employ so-called supervised autonomous weapons. Supervised autonomous weapons present similar characteristics to autonomous weapons, i.e. the whole engagement cycle is automated, but humans supervise the operations and can intervene.⁶

However, the above are far from being commonly accepted concepts and distinctions, with much of the debate today centering on the lack of and potential need for an agreed-upon definition of autonomous weapons. Definitional ambiguity is also reflected in the debates on the development of these weapons, with many arguing that we are inevitably going towards full autonomy in weapons,⁷ while others maintain that autonomous weapons are already here.⁸ While uncertainty as to the definition of autonomous weapons and as to a possible arms race loom, the study of the challenges brought about by the increased use of automation and autonomy in weapons should not be hampered by the absence of an agreed-upon definition.

It is indeed necessary to keep in mind that the concept of autonomy itself is contentious.⁹ Moreover, agreement on a definition might be difficult to reach since ‘there may be no single general level of autonomy across the system and different functions of a weapons system may have different levels of autonomy’.¹⁰ In the comprehensive report of the Stockholm International Peace Research Institute (SIPRI)

⁵ Paul Scharre, ‘Centaur Warfighting : The False Choice of Humans vs. Automation’ (2016) 151 *Temp. Int’l & Comp. L.J.* 154.

⁶ Scharre (n 4) 45.

⁷ Marra and McNeil (n 2) 1153–1157; Timothy Coughlin, ‘The Future of Robotic Weaponry and the Law of Armed Conflict: Irreconcilable Differences?’ (2011) 17 *UCL Jurisprudence Review* 67, 70–74; Peter W Singer, *In the Loop? Armed Robots and the Future of War*, Brookings (Jan. 28, 2009).

⁸ Rebecca Crootoft, ‘The Killer Robots Are Here: Legal and Policy Implications’ [2015] *Cardozo L. Rev* 1837, 1868–1872; Scharre (n 4). See also ‘Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems’ 12.

⁹ Marcus Schulzke, ‘Autonomous Weapons and Distributed Responsibility’ (2013) 26 *Philosophy and Technology* 203, 207, 208 and 211. See also Ekelhof affirming: ‘The term autonomy generates the most intricate, emotionally-laden and connotation-loaded debates. “There is a myth that autonomy is some single thing and that everyone understands what it is”. However, the word autonomy is employed with different meanings and intentions and can be viewed from many different angles’, Merel AC Ekelhof, ‘Complications of a Common Language: Why It Is so Hard to Talk about Autonomous Weapons’ (2017) 22 *Journal of Conflict and Security Law* 311, 323.

¹⁰ ‘Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems’ (n 8) 12.

on autonomy in weapon systems, which is a point reference in any debate on lethal autonomous weapon systems (LAWS), it is contended that the relevant discussion should relate to the autonomy of functions.¹¹ Shifting the focus to the autonomy of *some functions* of a weapon system makes it fundamental to analyze how humans deal with the increased autonomy of some steps of the targeting cycle.

The focus of this paper is therefore on the main feature characterizing the present era of autonomy in weapons: one where humans work in tandem with machines in the targeting process. Scharre calls this ‘centaur warfighting’, which refers to ‘human machine teaming in targeting decisions’.¹² According to him, systems that incorporate both automation and human decision making are the ideal systems in the battlefield as they are able to ‘leverage the precision and reliability of automation without sacrificing the robustness and flexibility of human intelligence’.¹³ Since prominent experts in the field argue that this should also be the trend in next-generation weapons,¹⁴ this discussion is also of relevance for the future.

Discussions at the Group of Governmental Experts (GGE) on emerging technologies in the area of LAWS¹⁵ show that future trajectories do indeed appear to be heading in this direction. The delegations to the GGE on emerging technologies in the area of LAWS agree on a few basic principles, one of which is ‘the essential importance of human control, supervision, oversight or judgement in the use of lethal force’.¹⁶ The ‘human element in the use of lethal force’ will certainly be a decisive factor in establishing restrictions on the development of autonomy in LAWS, making human-machine teaming a highly probable feature of future warfare.

¹¹ Vincent Boulanin and Maaike Verbruggen, ‘Mapping the Development of Autonomy in Weapon Systems’ (2017) 7. See also Scharre (n 4) 28.

¹² Scharre (n 5) 151.

¹³ Scharre (n 5).

¹⁴ See presentation of Professor Mary Cummings at GGE meetings, on 12 April 2018, ‘Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems’ (n 8) 17. On human-machine teaming, she said that the ideal system in the battlefield would be the one that could strongly leverage the human-machine team. See *contra*, Sparrow (n 1) 69.

¹⁵ The GGE on emerging technologies in the area of LAWS was established by the 2016 Fifth Review Conference of the High Contracting Parties to the Convention on Certain Conventional Weapons (CCW) and meets twice a year in Geneva.

¹⁶ See Chair’s summary of the discussions of the GGE on emerging technologies in the area of LAWS established by the 2016 Fifth Review Conference of the High Contracting Parties to the Convention on Certain Conventional Weapons, April 2018, available at [https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/DF486EE2B556C8A6C125827A00488B9E/\\$file/Summary+of+the+discussions+during+GGE+on+LAWS+April+2018.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/DF486EE2B556C8A6C125827A00488B9E/$file/Summary+of+the+discussions+during+GGE+on+LAWS+April+2018.pdf).

Military culture is also a factor that underpins states' aversion toward full autonomy in weapon systems. One of the highest priorities of the militaries in developing LAWS is ensuring that the weapons will be more predictable than humans and that they will not 'achieve the same degree of autonomy as humans'.¹⁷ This suggests that autonomous weapons will be designed to have 'scripted or supervised autonomy' and that they will depend on human control or that a human will remain in a command or supervisory role.¹⁸

A good example of human-machine teaming is Automated Target Recognition (ATR) software implemented in semi-autonomous weapons. ATR software relies upon pattern recognition to identify military objects predefined by humans.¹⁹ ATR software has always been quite inefficient and rudimentary, with the ability to recognize only predetermined targets and with their performance depending on the environment.

There have been attempts to overcome the limitations of ATR software. Defense Advanced Research Projects Agency (DARPA) has begun a project aiming at improving ATR algorithms. The Target Recognition and Adaption in Contested Environments (TRACE) project by DARPA exploits advances in computer vision and machine learning. TRACE's aim is to develop better ATR algorithms and techniques that could 'rapidly and accurately identify military targets'.²⁰

Besides target recognition, another example concerns intelligence for targeting. There has been a significant increase in the amount of full-motion video produced by unmanned aerial vehicles (UAVs) in the past 10-15 years, the analysis of which is undertaken by hundreds of analysts.²¹ The US Department of Defense's (DOD) Algorithmic Warfare Cross-Functional Team (AWCFT), also known as Project Maven, was established in 2017 and aims to 'accelerate DoD's integration of big data and machine learning' and to 'turn the enormous volume of data available to DoD into actionable intelligence and insights at speed'.²² In essence, Project Maven involves the use of machine learning in software and automates the work of analysts

¹⁷ Schulzke (n 9) 208.

¹⁸ *ibid.*

¹⁹ Boulanin and Verbruggen (n 11) 24 and 25.

²⁰ Scharre (n 4) 84-88.

²¹ Merel AC Ekelhof, 'Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting' (2018) 71 *Naval War College Review* 61, 77.

²² *ibid.*

looking at drone video footage. It autonomously extracts objects of interest from moving or still imagery and it then relays its discoveries to a human analyst.²³

The use of automation and autonomy as a decision aid in the targeting process makes it crucial to understand the cognitive phenomena underlying human-machine teaming. In human-machine teams, human agents are increasingly required to take decisions in concert with automated systems. Human operators are required to process information provided by automated systems in order to take targeting decisions; or they oversee automated systems and are required to intervene when something in the targeting process goes wrong. This also entails the creation of new cognitive environments where decisions of targeting are taken. We witness the creation of new, hybrid cognitive architectures combining human intelligence with machine intelligence.

As aptly affirmed by Ekelhof, ‘introducing autonomous technologies into the targeting process presents fundamental challenges, not only to military structures and the military mind-set, but most importantly to decision-making processes and the relationships between human actors and technologies in the targeting process’.²⁴

Before analyzing how human agents’ targeting decisions take shape when automation or artificial intelligence is implemented in weapons and how this should influence the assessment of the *mens rea* of a perpetrator of war crimes, it is important to make a digression on the centrality of criminal responsibility for war crimes for IHL.

III. Criminal Responsibility for War Crimes

IHL regulates the rights and duties of the belligerents (be they states or armed groups) during armed conflict and the protection of those not or no longer taking part in hostilities. On the one hand, international humanitarian law lays down obligations for states: the fundamental rule of common Article 1 of the Geneva Conventions (GCs) being that ‘*the High Contracting Parties undertake to respect and to ensure respect for the present Convention in all circumstances*’. On the other hand, international humanitarian law also enacts provisions (i.e. prohibitions to commit serious violations) addressed to individuals. The four GCs of 12 August 1949 for the protection of war victims and the 1977 Additional Protocol I to the Geneva Conventions (AP I) stipulate that individual criminal responsibility arises for those

²³ Cheryl Pellerin, ‘Project Maven to Deploy Computer Algorithms to War Zone by Year’s End’, US Department of Defense News <<https://dod.defense.gov/News/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/>> accessed 25 January 2019.

²⁴ Ekelhof (n 21) 3.

individuals who commit or order the commission of serious violations of international humanitarian law. The GCs and the Additional Protocols distinguish between the most serious violations of the treaties, so-called ‘grave breaches’ (which, in the wording of these treaties, must be regarded as war crimes), and on the other hand, other violations. The GCs and AP I establish a specific framework for the prevention and punishment of ‘grave breaches’ which entails the obligation to provide for effective penal sanctions and mandatory prosecution under the scheme *aut dedere aut judicare*.²⁵

Violations of obligations incumbent upon states to the GCs and its Protocols give rise to the international responsibility of the state that has committed it. The consequences stemming from the violations of IHL committed by states fall within the law of state responsibility. For a violation of IHL to entail state responsibility it is necessary that the conduct underlying the violation is attributable to a state, according to the norms on attribution of the ILC Articles on state responsibility.²⁶

State responsibility is not dependent on and does not entail individual responsibility. However, international state responsibility can exist alongside individual responsibility.²⁷ IHL is one of the few regimes that, next to state responsibility, envisages individual criminal responsibility for the violations of some of its norms. The GCs, in particular, are the first IHL instruments to expressly criminalize breaches of the rules contained therein. As correctly pointed out, the Geneva Conventions (together with AP I) are based on the ‘humanization and individualization of inter-belligerent enforcement’.²⁸

²⁵ Article 85(5) AP I. See on grave breaches, Paola Gaeta, ‘Grave Breaches of the Geneva Conventions’ in Andrew Clapham, Paola Gaeta and Marco Sassòli (eds), *The 1949 Geneva Conventions: A Commentary Edited* (OUP 2015) 615 ff.

²⁶ See Article 91 of AP I: ‘A Party to the conflict [...] shall be responsible for all acts committed by persons forming part of its armed forces’. See on attribution of violations of IHL, Marco Sassòli, ‘State Responsibility for Violations of International Humanitarian Law’ (2002) 84 *Int’l Rev. Red Cross* 401, 404–412.

²⁷ On the non-exclusive nature of individual and state responsibility see Andre Nollkaemper, ‘Concurrence between Individual Responsibility and State Responsibility in International Law’ (2003) 1 *Int’l & Comp. L.Q.* 615–615, 618–621; Elies Van Sliedregt, *Individual Criminal Responsibility in International Law* (OUP 2012) 5–7.

²⁸ Paola Gaeta, quoting D Jinks, ‘Humanization and Individualization in the Enforcement of International Humanitarian Law’, paper submitted at a Colloquium on *Virtues, Vices and Human Behavior and Democracy in International Law*, organized by the Institute for International Law and Justice of the NYU Law School in Spring 2009, 9, ‘The Interplay Between the Geneva Conventions and International Criminal Law’ in Andrew Clapham, Paola Gaeta and Marco Sassòli (eds), *The 1949 Geneva Conventions: A Commentary Edited* (OUP 2015) 737.

In international humanitarian law, individual criminal responsibility is an important addition to forms of collective responsibility, such as state responsibility. The fact that international state responsibility is ‘conceived in terms of an essentially reparational view as well as a sort of civil liability in the international order’ and ‘it does not have the object of punishing the State’,²⁹ does not promote strong deterrent effects.

Moreover, in the law of state responsibility, human beings, their obligations, and their responsibility become ‘invisible’.³⁰ Nollkaemper sees the ‘invisibility’ of human beings in several parts of the law of state responsibility:

‘The principles governing breach and attribution are indifferent to the subjective conduct of the author of the act. The conduct of the state as a legal person is assessed against an objective standard. Fault may be determined by national law, but in principle does not enter the international legal sphere. The individual also is invisible in the principles governing remedies. Remedies fall on the state, not on individuals whose acts triggered state responsibility. Sanctions on individuals are left to national law’.³¹

In contrast, individual criminal responsibility presupposes the attribution of violations of its norms to individuals and the existence of norms directed at the *punishment* of individuals. The enforcement of international humanitarian law via the prosecution and punishment of individuals concentrates on the obligations that IHL imposes on individuals.³² By attaching responsibility to individuals, the moral and legal obligations of human beings under IHL are reinforced. As affirmed by Sassoli, ‘[p]rosecution clarifies that IHL is law, especially for those who consider “law” to be only that which is regularly enforced through sanctions’.³³

Criminalization ensures compliance with IHL³⁴ and has strong deterrent effects.³⁵ The individualization of responsibility and punishment has therefore become a crucial means to enforce IHL.³⁶

²⁹ Pierre-Marie Dupuy, ‘International Criminal Responsibility of the Individual and International Responsibility of the State’ in Antonio Cassese, Paola Gaeta and John RWD Jones (eds), *The Rome Statute of the International Criminal Court: A Commentary, Vol. 2*, vol II (2002) 1097.

³⁰ Nollkaemper (n 27) 617.

³¹ *ibid.*

³² Antonio Cassese, ‘On the Current Trends towards Criminal Prosecution and Punishment of Breaches of International Humanitarian Law’ (1998) 9 *European Journal of International Law* 2, 4.

³³ Marco Sassòli, ‘Humanitarian Law and International Criminal Law’ in Antonio Cassese (ed), *The Oxford Companion to International Criminal Justice* (OUP 2009) 113.

³⁴ Gaeta (n 28) 737.

³⁵ Sassòli (n 33) 113.

Individualization of responsibility becomes all the more important in modern warfare. Modern warfare is a complex military and political structure. It is characterized by many layers of decision making within the military and within the political sphere. Politicians take decisions to initiate wars. Politicians and military commanders define objectives and strategies. Then, decisions on how to achieve military and political objectives are taken at different levels of the military hierarchy.³⁷

Multi-layered decision making is to an even greater extent a feature of targeting, the specific focus of this paper. Targeting is a complex process involving different levels of decision making.³⁸ All the four different targeting steps (i.e. developing objectives and guidance; planning; execution and assessment) involve different actors, such as commanders, operational planners, tactical planners, component commanders and intelligence experts.

The main consequence of multi-layered military decision making is distribution of responsibility. Distribution of responsibility may turn into dispersal or diffusion of responsibility. The decision making process being fragmented across multiple actors may render it difficult to ascribe responsibility for a certain outcome to a single actor participating in the decision making process.

This feature is exacerbated by the increased use of military technology. The dramatic advancements in military technologies, and their integration into military structures and tactics may indeed add new actors to the already complex military theater: machines. Human-machine teaming has become and will probably remain a common feature of warfare. As such, we witness an increasing trend towards the sharing of responsibility between humans and machines.

In the relationship between the human and the machine, individualization of the responsibility of the human agent is crucial. This will lead to responsabilize the human behind the machines, strengthen precautionary standards, reinforce compliance with the law and prevent crimes. With regard to the individualization of responsibility for violations of IHL, an analysis of the principle of culpability and its centrality in criminal law is in order.

³⁶ See also Sparrow: '[t]he assumption and/or allocation of responsibility is also vital in order for the principles of *jus in bello* to take hold at all. [...] [A]pplication of the principles of *jus in bello* requires that we can identify the persons responsible for the actions that these principles are intended to govern', Sparrow (n 1) 67.

³⁷ Marcus Schulzke, 'Robots as Weapons in Just Wars' (2011) 24 *Philosophy and Technology* 293, 209 and 210.

³⁸ Ekelhof (n 21) 66–74.

IV. The Central Importance of Culpability in Criminal Law Theory

The process of individualization of responsibility for grave violations of international law through war crimes is contingent on one fundamental concept and principle: the concept of ‘guilt’ or the principle of culpability.

The concept of ‘guilt’ or principle of culpability in its descriptive meaning refers to the *mens rea* requirements contained in each offence, which will be specifically analyzed in Section V and even more specifically with respect to conduct of hostilities crimes in Section V.I. In brief, criminal liability is best thought of as the culmination of an act with a culpable state of mind.³⁹ *Mens rea* is, next to the *actus reus*, one of the fundamental requisites of a criminal offence and for criminal responsibility. The *mens rea* element of a crime expresses the legal and descriptive dimension of culpability: the requirement that in order to ascribe criminal responsibility an accused must carry out a voluntary physical act with a prescribed mental state.

Culpability also has a fundamental moral and normative dimension. In this sense, it refers to blameworthiness and the conditions under which one can be fairly and morally blamed for a criminal behavior.⁴⁰

Moral culpability can be considered the basis of the modern criminal justice system.⁴¹ This argument rests on the idea, found in both criminal law and ethics, that individuals should be held “‘responsible’ for their volitional conduct’.⁴² In this sense, culpability signifies ‘moral fault, blameworthiness [and] guilt’.⁴³ The principle of culpability is the primary means through which the law imposes blame when individuals commit crimes.⁴⁴ Punishment is legal coercion inflicted on an individual because they have culpably committed a violation of a norm.⁴⁵ *Mens rea* serves to

³⁹ Douglas Husak, *The Philosophy of Criminal Law: Selected Essays* (OUP 2010) 27.

⁴⁰ George P Fletcher, *Basic Concepts of Criminal Law* (OUP 1998) 99.

⁴¹ Jerome Hall, *General Principles of Criminal Law* (The Bobbs-Merrill Company Publishers 1947) 168; see *contra* HLA Hart, *Punishment and Responsibility: Essays in the Philosophy of Law* (OUP 2008) 36.

⁴² Jerome Hall, *General Principles of Criminal Law* (The Bobbs-Merrill Company Publishers 1947) 166.

⁴³ Edward M. Wise, ‘The Concept of Desert’, (1987) 33 *Wayne Law Review* 1343, 1352; Michael Moore, *Placing Blame: A General Theory of Criminal Law* (OUP 2010) 403.

⁴⁴ Douglas Husak, ‘Broad Culpability and the Retributivist Dream’ (2012) 9 *Ohio State Journal of Criminal Law* 449, 449.

⁴⁵ Guyora Binder, ‘Punishment Theory: Moral or Political?’(2002) 5 *Buffalo Criminal Law Review* 321, 353.

ensure that only those who are ‘genuinely blameworthy’⁴⁶ are punished in the criminal process.

The criminal justice system seeks to be distinguished from civil law on the grounds that criminal offenses are contingent on fault and blameworthiness such that the basis of guilt is moral culpability.⁴⁷ Such distinction emerges through a focus on ‘intentional and knowing behaviour’, expressed in terms of *mens rea* requirements, the analysis of which will follow.⁴⁸

V. *Mens Rea* Requirements – Legal and Descriptive Dimension

The descriptive function of *mens rea* constitutes one of the elements of a criminal offence. While the *actus reus* is the external manifestation (i.e. objective element) of a criminal offence, the *mens rea* is the necessary state of mind (i.e. subjective element) that must accompany the prohibited behavior or consequence. Assessment of a perpetrator’s *mens rea* is carried out only once the *actus reus* has been fulfilled. The *actus reus* element may consist of proscribed conduct or a proscribed consequence. This distinction, which will be relevant for the analysis of conduct of hostilities crimes,⁴⁹ leads to the classification of offences as either conduct or result crimes. The former are consummated when the conduct is carried out; the latter are consummated only when a certain result – causally linked to the act or omission of the perpetrator – takes place. Therefore, only in relation to result crimes, is proof of causation necessary.⁵⁰

The starting point of this paper is the fulfillment of the *actus reus* element of the relevant war crime. Given the fact that this paper focuses on human-machine teaming and on war crimes committed by human agents in concert with machines, it is unnecessary to delve into the thorny and debated question of whether artificial intelligent systems may be able to carry out an *actus reus* themselves.⁵¹ This paper

⁴⁶ Samuel W Buell, ‘Culpability and Modern Crime’ (2015) 103 *The Georgetown Law Journal* 547, 586.

⁴⁷ Darryl K Brown, ‘Criminal Law Theory and Criminal Justice Practice’ (2012) 49 *American Criminal Law Review* 73, 78.

⁴⁸ William Shabas, *An Introduction to the International Criminal Court* (5th edn, CUP 2017) 221.

⁴⁹ See different formulations in the Rome Statute and AP I.

⁵⁰ Johannes Keiler, ‘Actus Reus and Mens Rea’ in Johannes Keiler and David Roef (eds), *Comparative Concepts of Criminal Law* (Intersentia 2016) 60.

⁵¹ The voluntary act requirement – entailing that the conduct must be the result of a voluntary and conscious action on the part of the perpetrator – is provided for in the Rome Statute: see Gerhard Werle and Florian Jeßberger, *Principles of International Criminal Law* (3rd Edition, OUP 2014) para 474 and

will focus on the *mens rea* of the human operator and how this is affected by the introduction of automation and artificial intelligence in weapon systems. Before turning to the *mens rea* requirements for conduct of hostilities crimes it is necessary to first provide some definitions and frame the debates surrounding the mental element of the war crimes that could be committed with semi-autonomous systems.

V.I A Comparative Criminal Law Analysis of *Mens Rea* Standards

Comparative criminal law studies show that, broadly speaking, criminal law addresses and punishes *intentional* harm and *risky behaviors*. As aptly put by Fletcher,

‘[t]he criminal law should begin, [the subjective theory] holds, by identifying interests that are worthy of protection. The next step should be preventing conduct that threatens those interests. The reasons that humans sometimes threaten those interests is either that they *intend to do so or that they take risks* that subject the protected interest to danger’.⁵²

Across different common law and civil law systems, a person can only be blamed and convicted for conduct or consequences that he brought about intentionally or at least negligently.⁵³ The ground rule is that criminal offences must be committed intentionally, while negligence – which implies the failure to comply with a standard of diligence, and the failure to perceive a risk –⁵⁴ is the exception and must be explicitly provided for by law.⁵⁵

also in several national systems (such as Italy and the US). The requirements of voluntariness and consciousness of a criminal act cast doubt as to whether fully autonomous are capable of carrying out an *actus reus*: ‘the voluntary aspect of *actus reus* may never be met since the idea that an [artificial agent] can act voluntarily is contentious’, Thomas C King and others, ‘Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions’ [2019] *Science and Engineering Ethics* 1, 7. See also Pagallo: ‘[t]he common legal standpoint excludes robots from any kind of criminal responsibility because they lack psycho-logical components such as intentions or consciousness, i.e., the set of preconditions for attributing liability to someone in the case of violation of criminal laws like murdering or stealing’, Ugo Pagallo, ‘Killers, Fridges, and Slaves: A Legal Journey in Robotics’ (2011) 26 *AI and Society* 347, 349. See on the impossibility to hold a machine morally responsible for its actions, Sparrow (n 1) 71–73. On autonomous weapons’ lack of self-consciousness, see also Schulzke (n 9) 207.

⁵² George P Fletcher, *Rethinking Criminal Law* (OUP 2000) 118 and 119 (emphasis added).

⁵³ Jeroen Blomsma and David Roef, ‘Forms and Aspects of Mens Rea’ in Johannes Keiler and David Roef (eds), *Comparative Concepts of Criminal Law* (Intersentia 2016) 127.

⁵⁴ Civil law countries recognize two forms of negligence (advertent and inadvertent). Common law countries accept that certain offences ‘can be committed negligently, or even unintentionally (eg strict or absolute liability offences), see Sarah Finnin, ‘Mental Elements under Article 30 of the Rome Statute of the International Criminal Court: A Comparative Analysis’ (2012) 61 *International and Comparative Law Quarterly* 325, 328.

⁵⁵ Michael Bohlander, *Principles of German Criminal Law* (Studies in International and Comparative Criminal Law 2009) 59.

Within intention, or *dolus*, there are different stages which correspond to different degrees of the volitional and cognitive components. Civil law countries recognize direct intent or *dolus directus* in the first degree, indirect intent or *dolus indirectus* in the second degree, and *dolus eventualis*.⁵⁶ Common law countries know direct and indirect intent, and recklessness, which, however, is not a form of intent.

Direct intent or *dolus directus* in first degree reflects the perpetrator's aim and is characterized by the accused's 'purposeful will' to engage in the prohibited conduct or bring about the forbidden results. Common law systems define it as acting 'purposely', i.e. having a 'conscious object' to engage in the forbidden conduct or to cause the forbidden results (US Model Penal Code); or acting 'intentionally' (UK and Australia). It is apparent that for direct intent, the volitional component is very strong. The volitional component of indirect intent or *dolus indirectus* in second degree is instead weaker and is substituted by the cognitive component. Indirect intent or *dolus indirectus* in second degree refers to the consequences which are not part of the accused's aim (so they are unwanted) but which are inevitably the result of his aims. In civil law systems, the perpetrator must foresee certain consequences 'as a certainty or as highly probable'.⁵⁷ The perpetrator who 'acts despite knowing' is treated as if he 'acted wanting'. In common law systems, this mental state is known as acting 'knowingly' (US Model Penal Code), or still broadly defined as 'intentionally' (UK and Australia). 'Knowingly' according to the US Model Penal Code means being 'practically certain' that one's conduct will cause the forbidden result; whereas in the UK the accused is required to have been 'virtually certain' that the forbidden result will occur.⁵⁸

The third and lowest level of intent in civil law systems is *dolus eventualis*, which refers to the accused's knowledge of *possible* or *probable* harmful consequences (cognitive element) 'combined with a positive mental or emotional disposition towards it' (volitional element).⁵⁹ In *dolus eventualis*, the cognitive element consists of the accused's awareness 'of the fact that his actions *may* lead to an offence being committed'.⁶⁰ Different theories exist as to the cognitive component of *dolus eventualis*. This has been formulated in terms of awareness of the *possibility* or of the

⁵⁶ There are of course differences between and within countries.

⁵⁷ Finin (n 54) 332.

⁵⁸ *ibid.*

⁵⁹ G Taylor, 'Concepts of Intention in German Criminal Law' (2004) 24 Oxford Journal of Legal Studies 99, 102.

⁶⁰ Bohlander (n 55) 64.

probability that a certain forbidden result will occur,⁶¹ or in Dutch law for example, as awareness of a considerable risk.⁶²

The accused's knowledge must be accompanied by a subjective attitude towards the possible or probable consequences.⁶³ The subjective posture towards the possible or probable consequences has been theorized and interpreted differently in civil law systems: depending on the system, the perpetrator must *accept* or *approve* the forbidden consequence, or *be reconciled* or *make peace with* their occurrence, or *be indifferent to* their occurrence.⁶⁴ These different theories show differing degrees of the volitional element that range from approval to indifference, the latter resembling the standard for recklessness.⁶⁵

Common law systems do not utilize the category of *dolus eventualis* and have instead a different parallel category of *mens rea*: recklessness. Despite some similarities, recklessness is not a form of intent but constitutes a separate form of *mens rea*. Acting 'recklessly' refers to the mental state of the perpetrator who foresees that his conduct may bring about the forbidden harmful result, but nevertheless 'takes a deliberate and unjustifiable risk of bringing it about'.⁶⁶ Recklessness is the conscious taking of an unreasonable risk. Therefore, despite being aware of the risk of causing the result proscribed by the criminal offence, the 'reckless' perpetrator disregards it and persists. The US Model Penal Code defines recklessness as 'conscious risk creation'.⁶⁷

The function of *dolus eventualis* and recklessness is to include and make punishable those 'side effects' of the actor's intentional conduct that would otherwise fall outside the scope of intention.⁶⁸ Foreseen and unwanted side-effects are equated with intended results.⁶⁹ There has been enduring tension over the demarcation between *dolus eventualis* and recklessness. In brief, the difference boils down to the perpetrator's posture toward risks *vis-à-vis* the objective, actual risk of harm. While

⁶¹ *ibid* 65.

⁶² Blomsma and Roef (n 53) 133.

⁶³ Taylor (n 59) 102.

⁶⁴ Mohamed Elewa Badar, *The Mental Element in the Rome Statute of the International Criminal Court: A Commentary from a Comparative Criminal Law Perspective*, vol 19 (2008) 489–491; Fletcher (n 52) 445 and 446; Blomsma and Roef (n 53) 137.

⁶⁵ Bohlander (n 55) 65.

⁶⁶ See Badar referring to *R. v. G and Another* [2004] 1 AC 1034 (HL) Badar (n 64) 488.

⁶⁷ Model Penal Code, para 2.02(2)(c). Knowledge of the risk is the crux of the distinction between recklessness and negligence.

⁶⁸ Fletcher (n 52) 443.

⁶⁹ Blomsma and Roef (n 53) 132.

dolus eventualis stresses the actor's attitude towards those risks, recklessness focuses on the actual risk and the likelihood of harmful consequences (the 'reckless' perpetrator has, as opposed to the perpetrator acting with *dolus eventualis*, an 'affirmative aversion to the harmful side effects').⁷⁰ Finally, what grounds recklessness is the perpetrator's awareness of the risks he has taken. This distinguishes it from negligence, which instead requires lack of knowledge of such risks.⁷¹

To conclude, determining criminal liability in a way that is consistent with the principle of culpability should be based on what the accused believed they were doing or on what the accused believed they were consciously risking. Criminal law responses are required for intentional harm or risky side effects of intentional conduct. With regard to risky behavior, the cognitive element is predominant over volition and requires the accused to recognize the danger of harmful consequences that they are risking. Moreover, for a determination of *dolus eventualis* or recklessness the perpetrator's attitude towards the risk of crimes being committed is determinative.

This digression on *mens rea* gives us the necessary framework to now approach the analysis of *the constituent elements* of conduct of hostilities crimes. The analysis will be conducted with the specific aim of inquiring into the relevance of risk taking behavior in the law of conduct of hostilities crimes. The intentional commission of conduct of hostilities crimes by the human operator of semi-autonomous systems falls squarely within the legal framework of the Rome Statute and AP I, which respectively require 'intentionality' and willfulness as mental elements. However, the scenarios that could become more relevant in the context of deployment of semi-autonomous weapons are those where the military lack direct intent but instead incur some risks of killing civilians. Some relevant developments have occurred in the law and prosecution of conduct of hostilities crimes concerning the interpretation of the mental element for the war crime of attacking civilians as including *dolus eventualis* and recklessness. An analysis of these developments is provided in the next Section.

V.II Zoom in on Conduct of Hostilities Crimes

In the debates on criminal responsibility for the use of semi-autonomous weapons, little attention is currently being paid to the specific crimes that could be committed

⁷⁰ Fletcher (n 52) 448.

⁷¹ Blomsma and Roef (n 53) 142.

with autonomous and semi-autonomous weapons, i.e. conduct of hostilities crimes, and to the proof of the elements of these crimes. For a meaningful discussion of responsibility for crimes committed with semi-autonomous weapons, but also, autonomous weapon systems, such an analysis is in order.

As seen above, criminal responsibility is ascribed on the basis of the fulfillment of an *actus reus* accompanied by a specific mental state of the perpetrator. Given its relevance and possible occurrence when semi-autonomous weapons are used, I shall focus here on the war crime of attacking civilians as set out in Article 8(2)(b)(i) of the Rome Statute of the International Criminal Court (ICC)⁷² and Article 85(3)(a) of AP I.

A comparison of the two definitions is in order. Article 8(2)(b)(i) of the Rome Statute requires that the perpetrator *intentionally* direct ‘attacks against the civilian population as such or against individual civilians not taking direct part in hostilities’. Insofar as the *actus reus* is concerned, the crime of attacking civilians and civilian objects – and similarly other war crimes relating to methods of warfare – is formulated in terms of conduct crimes. The *actus reus* is fulfilled by the launch of the attack regardless of civilian deaths or injuries resulting from it. While the required *mens rea* in Article 8(2)(b)(i) of the Rome Statute is ‘intentionality’, the *mens rea* of the war crime of attacking civilians as defined in Article 85(3)(a) of AP I is ‘willfulness’. Finally, insofar as the *actus reus* is concerned, AP I contains the additional requirement that an attack against civilians cause ‘*death or serious injury to body or health*’. Under AP I, the war crime of attacking civilians (as well as the other grave breaches) therefore amounts to a result crime, i.e. the offence is consummated only if certain proscribed consequences (i.e. deaths or injury) take place. This aspect is fundamental for the analysis of *mens rea* under the Rome Statute, since Article 30 on *mens rea* defines different gradings of intent, one of which (resembling the *dolus eventualis* standard) is only applicable to result crimes.⁷³ Arguably, this different formulation of the crime of attacking civilians also bears on the *mens rea* requirements under AP I. It has indeed been argued that AP I lowers the *mens rea* requirement for grave breaches (requiring ‘willfulness’ instead of

⁷² Hereinafter Rome Statute.

⁷³ See discussion thereafter on the application of Article 30(2)(b) to result crimes.

intentionality) because of the additional requirement of the proof of result, i.e. deaths or injuries as a result of the attack.⁷⁴

Fundamentally, this offence criminalizes the violation of the IHL principle of distinction between civilian and military targets. It prohibits military forces from *intentionally* targeting civilians. The intentionality requirement would seem to exclude the risk taking conduct of the human operator who is aware of possible civilian deaths but still decides to take the risk and launches the attack. However, two caveats are important for our purposes since they open the door to the prosecution of risk taking behavior for this crime.

First, despite the differing ‘intentionality’ and ‘willfulness’ requirement and the underlying rationale that this offence prohibits the attacks intentionally directed at civilian targets, there are interpretations of the mental element that allow for the consideration of risk taking behavior in the prosecution of the war crime of attacking civilians. These interpretations concern the mental elements established by both the Rome Statute and AP I.

Second, this offence not only encompasses intentional attacks against civilians, but also attacks that ‘lack targeting’. In particular, this includes the following: 1) attacks that are not directed against a specific military target; 2) attacks employing indiscriminate weapons (i.e. weapons incapable of distinguishing between civilian and military targets); and 3) attacks carried out without taking the necessary precautions to spare civilians, ‘especially failing to seek precise information on the objects or persons to be attacked.’⁷⁵ Attacks effectuated without taking the necessary precautions is a likely scenario in the deployment of semi-autonomous weapons. The inclusion of this type of attack represents a gateway for considering mental elements lower than direct intent (i.e. the violation of the obligation to take all necessary precautions) for the purposes of ascribing criminal responsibility for the war crime of attacking civilians.

Turning to the interpretations of the mental element under the Rome Statute, the general rule on *mens rea* is the most stringent mental element in both civil law and common law systems.⁷⁶ Article 30 of the Rome Statute requires that, in the absence of

⁷⁴ Hector Olasolo, ‘Subjective Elements’, *Unlawful Attacks in Combat Situations: From the ICTY’s Case Law to the Rome Statute* (Martinus Nijhoff Publishers 2008).

⁷⁵ Knut Dörmann, *Elements of War Crimes under the Rome Statute of the International Criminal Court: Sources and Commentary* (CUP 2003) 131 and 132; Olasolo (n 74) 218.

⁷⁶ Finnin (n 54) 355.

specific mental state requirements contained in the definition of the crimes (the ‘unless otherwise provided’ clause) the *actus reus* of an offence be committed with ‘intent’ and ‘knowledge’. Under the Rome Statute, the perpetrator must intend to engage in the forbidden conduct (‘intent’ as defined by Article 30(2)(a)) or must intend to cause the forbidden result (‘intent’ as defined by Article 30(2)(b)). It follows that Article 30(2)(a) defines intent with respect to conduct crimes while Article 30(2)(b) applies to result crimes. These two definitions of intent amount to direct intent or *dolus directus* in the first degree.⁷⁷

There is agreement in scholarship and in the case law of the ICC that the mental state of the perpetrator who *does not intend* to cause the forbidden result, but *foresees* its occurrence as a *necessary* or *certain* or *highly probable* consequence of the achievement of his main purpose and nevertheless engages in the conduct (indirect intent or *dolus directus* in the second degree) also falls within Article 30.⁷⁸ In contrast, disagreement prevails over the question of whether the conduct of the perpetrator acting with *dolus eventualis* could be punished under the Rome Statute. As mentioned above, *dolus eventualis* refers to the mental state of the perpetrator who is aware that his conduct *may* produce the forbidden result and he is indifferent to or is reconciled with or accepts the occurrence of the forbidden result,⁷⁹ depending on the theory to which one subscribes.

Before dealing with the applicability of *dolus eventualis* for the purposes of the present investigation, it is important to bear in mind that Article 30 of the Rome Statute defers to specific *mens rea* requirements contained in the definition of the crimes. According to Article 8(2)(b)(i) of the Rome Statute, the subjective element for the war crime of launching an attack against civilians is ‘intentionality’. The question then arises as to whether this is a different standard from Article 30 of the Rome Statute. ‘Intentionality’ under Article 8(2)(b)(i) of the Rome Statute has been interpreted as covering *dolus directus* in the first degree and *dolus directus* in the

⁷⁷ Gerhard Werle and Florian Jeßberger, *Principles of International Criminal Law* (3rd edn, OUP 2014) para 474 and 475, Olasolo (n 74) 209.

⁷⁸ Finnin (n 54) 355; Ex plurimis, Olasolo (n 74) 218. See Decision Pursuant to Art 61(7)(a) and (b) of the Rome Statute on the Charges of the Prosecutor against Jean-Pierre Bemba Gombo, *Bemba* (ICC-01/05-01/08-424), Pre Trial Chamber II, 15 June 2009, para 359; Decision on the Confirmation of Charges, *Lubanga*, (ICC-01/04-01/06-803, Pre Trial Chamber I, 29 January 2007, paras 351-352; Decision on the Confirmation of Charges, *Katanga and Ngudjolo Chui* (ICC-01/04-01/07-717), Pre Trial Chamber I, 30 September 2008 para 530.

⁷⁹ Finnin (n 54) 332.

second degree.⁸⁰ It follows that Article 8(2)(b)(i) covers the mental state of the perpetrator who means to engage in the attack and intends the civilian population to be the object of an attack.⁸¹ But, it also covers situations where the perpetrator does not want to hit civilians but ‘*is aware that civilians or civilian objects will be necessarily hit (dolus directus in the second degree)*’.⁸² In *dolus directus* in the second degree, the cognitive component, i.e. the perpetrator’s knowledge with regard to the civilians being inevitably hit, becomes decisive to prove his *mens rea*.

Is also knowledge of the *possibility* or *probability* of hitting civilians a culpable mental state under Article 30 and/or Article 8 of the Rome Statute? Does *dolus eventualis* – which refers to awareness of the *possibility* or *probability* of hitting civilians combined with a positive or indifferent attitude towards this – have application in conduct of hostilities crimes? This question is of crucial importance in the debate concerning criminal responsibility for crimes committed with semi-, but also, fully-autonomous weapon systems.

As mentioned, Article 30(2)(b) of the Rome Statute clarifies that intent with regard to consequences means ‘mean[ing] to cause that consequence’. This Article also contains a second definition of intent as ‘awareness’ that a consequence ‘will occur in the ordinary course of events’. This is a specific definition of ‘intent’ applying only to offences that entail certain consequences (result crimes). According to some legal scholars and to some decisions of the ICC, the ‘awareness that a consequence will occur in the ordinary course of events’ amounts to *dolus eventualis*.⁸³ This author’s view is that the formulation of Article 30(2)(b) does not make it clear whether the required level awareness is that a consequence will *certainly* occur (indirect intent or *dolus directus* in the second degree) or that the consequence will *probably* or *possibly* occur in the ordinary course of events’.⁸⁴ The acceptance of the latter standard would

⁸⁰ Olasolo (n 74) 218.

⁸¹ Dörmann (n 75) 181.

⁸² Olasolo (n 74) 218.

⁸³ Hans-Heinrich Jescheck, ‘The General Principles of International Criminal Law Set Out in Nuremberg, as Mirrored in the ICC Statute’ (2004) 2 J. Int’l Crim. Just. 38, 45; Otto Triffterer, The New International Criminal Law – Its General Principles Establishing International Criminal Responsibility, in The New International Criminal Law, 639-727 (Kalliopi Koufa edn, 2003), 706. See Decision on the Confirmation of Charges, *Lubanga*, (ICC-01/04-01/06-803, Pre Trial Chamber I, 29 January 2007, paras 352 and 354; see also Decision on the Confirmation of Charges, *Katanga and Ngudjolo Chui* (ICC-01/04-01/07-717), Pre Trial Chamber I, 30 September 2008 para 251, footnote 329.

⁸⁴ See on the debates concerning the interpretation of the wording ‘will occur’ Finnin (n 54) 346 and 347. For an interpretation excluding *dolus eventualis* from Article 30(2)(b) see Johan D Van der Vyver,

lead to the inclusion of *dolus eventualis*.⁸⁵ However, due to the lack of clarity and poor formulation of Article 30 of the Rome Statute, diverging positions among legal scholars and the lack of conclusive determinations by the ICC, the debate remains open.⁸⁶

At first glance, these discussions do not appear to be relevant for the war crime of attacking civilians. In fact, Article 30(2)(b) of the Rome Statute only applies to result crimes, i.e. those for which a certain consequence is an element of the crime. It does not apply to conduct crimes, such as launching an attack against civilians, to which only the article 30(2)(a) definition of ‘intent’ as meaning to engage in the conduct applies.⁸⁷ However, there are two observations that must be taken into account. First, the war crime of launching an attack against civilians has been construed as a result crime, the result not being constituted by ‘deaths or injuries’ but by ‘hitting civilians’. Accordingly, the conduct element would be fulfilled by ‘dropping a bomb’ and its consequence produced with the ‘hitting of civilians’.⁸⁸ Moreover, in the context of the

‘The International Criminal Court and the Concept of Mens Rea in International Criminal Law’ (2005) 12 U. Miami Int’l & Comp. L. Rev. 57, 70 and 71.

⁸⁵ See, ‘[t]he concept ‘dolus eventualis’ does not have a monolithic or uniform meaning within all civil law systems. It generally includes awareness of a substantial or high degree of probability that the consequence will occur, and in many systems, it also includes awareness of a serious risk that a consequence will occur, coupled with indifference to that outcome or reconciliation with that outcome. Some systems may also include some forms of inadvertence in this concept. Due to different national conceptions, attempts to define the concepts were abandoned during the negotiations. Whatever may be the merits of the distinction under national legal systems, it was clear that the first-noted meaning of ‘dolus eventualis’ is included (i.e., ‘will occur in the ordinary course of events’). The latter meanings of ‘dolus eventualis’ or ‘recklessness’ were not incorporated explicitly into article 30, although it may be open to the Court to interpret the provision to include some of these aspects’, Donald K Piragoff and Darryl Robinson, ‘Article 30, Mental Element’ in Otto Triffterer and Kai Ambos (eds), *The Rome Statute of the International Criminal Court A Commentary* (Beck/Hart 2016) 1122 footnote 77.

⁸⁶ See for positions excluding *dolus eventualis*, Kai Ambos, ‘General Principles of Criminal Law in the Rome Statute’ (1999) 10 Criminal Law Forum 1, 21 and 22; Gerhard Werle and Florian Jessberger, ‘Unless Otherwise Provided: Article 30 of the ICC Statute and the Mental Element of Crimes under International Criminal Law’ (2005) 3 J. Int’l Crim. Just. 35 35, 53; Finin (n 14) 349. See *contra* Decision Pursuant to Art 61(7)(a) and (b) of the Rome Statute on the Charges of the Prosecutor against Jean-Pierre Bemba Gombo, *Bemba* (ICC-01/05-01/08-424), Pre Trial Chamber II, 15 June 2009, paras 360-369; Concurring Opinion of Judge Van den Wyngaert, *Ngudjolo Chui* (01/04-02/12), Trial Chamber II, 18 December 2012, para 15.

⁸⁷ On the ‘element analysis’ approach adopted by the Rome Statute see Badar (n 64) 474–484.

⁸⁸ Michael Bothe, ‘War Crimes’ in Antonio Cassese, Paola Gaeta and Jones, John R.W.D. (eds), *The Rome Statute of the International Criminal Court, A Commentary* (OUP 2002) 389. See *contra* Schabas affirming: ‘[t]hat no result need be demonstrated, and that the offence is committed by the mere launching of an attack, is well established in the case law of the International Criminal Court’, William Schabas, *The International Criminal Court: A Commentary on the Rome Statute* (2nd edn, 2016 OUP) 257. Schabas cites *Katanga* (ICC-01/04-01/07), Judgment pursuant to article 74 of the Statute, 7 March 2014, para 799: ‘[i]t further considers that the crime of attack against civilians proscribes a certain conduct and that the material element is established where the attack is launched and its object is a civilian population as such or individual civilians not taking direct part in hostilities; no result need ensue from the attack. Indeed, the Chamber considers that the absence of a result requirement in

prosecution of unlawful attacks, according to Fenrick the requirement of result is implied in Article 8(2)(b)(i). Fenrick indeed affirms that ‘[i]n any event, proof of death, injury or damage would usually be necessary to establish the attack was directed at civilian persons or objects’.⁸⁹

Second, even if one does not subscribe to the interpretations of unlawful attacks as result crimes, we must take into account that it has been argued that Article 30(2)(a) (defining intent as desire to engage in the prohibited conduct) also covers *dolus eventualis*.⁹⁰ This would lead to the application of *dolus eventualis* to conduct crimes, such as the war crime of attacking civilians. The vastly diverging positions as to the inclusion of *dolus eventualis* in Article 30 leads to a lack of clarity that cannot but bear on the problem of ascription of criminal responsibility for crimes committed with (semi) autonomous weapons.

Turning to Article 8(2)(b)(i) and its ‘intentionality’ requirement, the question arises as to whether this standard encompasses *dolus eventualis*? Importantly for our discussion, the ‘intentionality’ requirement of Article 8(2)(b)(i) has been interpreted by some scholars as not deviating from Article 30 of the Rome Statute.⁹¹ Instead, it has been construed as simply restating the general subjective element provided for by Article 30 of the Rome Statute.⁹² Therefore, if we adhere to the interpretations of Article 30 of the Rome Statute advocating for an inclusion of *dolus eventualis*, Article 8 would also cover actions of those who are *aware of the likelihood* or the possibility that, instead of hitting a military target, they will end up hitting civilians or civilian objects and is reconciled with or indifferent to such a result.

the Elements of Crimes is not accidental, insofar as, where such a requirement exists, the Elements of Crimes refer to it and specify the consequence thereof.

⁸⁹ WJ Fenrick, ‘A First Attempt to Adjudicate Conduct of Hostilities Offences: Comments on Aspects of the ICTY Trial Decision in the Prosecutor v. Tihomir Blaskic’ (2000) 13 *Leiden Journal of International Law* 931, 940. See also Dörmann highlighting that during the negotiations of the Rome Statute, the minority affirmed that ‘it had always been tacitly understood that the grave breach threshold would be applicable. If there is a weapon failure the conduct should only be charged as an attempt’ Dörmann (n 75) 130.

⁹⁰ See Van der Vyver, ‘[t]he phrase designating the mental element in the former context (the accused “means to engage in the conduct”) would, according to general principles of criminal law, include *dolus directus*, *dolus indirectus*, and *dolus eventualis*’, Vyver (n 84) 70.

⁹¹ According to Piragoff and Robinson: ‘[a] number of the definitions of crimes specifically require that the material elements be “intentional” or be committed “intentionally”. This is likely superfluous, given the general rule in article 30. The specific presence of these terms is likely a product of the negotiation process whereby the terms were imported from other international instruments, such as the Geneva Conventions, or where certain delegations wished to make clear the intentional nature of the crimes before they agreed to their inclusion’, Piragoff and Robinson (n 85) para 15. See *contra* Bothe (n 88) 89.

⁹² Dörmann (n 75) 131 and 148. See *contra* Bothe (n 88) 389.

What about the reckless human operator of semi-autonomous weapons? The perpetrator acting ‘recklessly’ takes an unreasonable and unjustifiable risk but has an affirmative aversion to the harmful effects of this risky behavior (as opposed to the approval or indifference of the subject acting with *dolus eventualis*).⁹³ Is reckless conduct relevant in the law of conduct of hostilities crimes? Antonio Cassese has – perhaps too overzealously – advocated for recklessness to be included in the Rome Statute and to apply at least to war crimes.⁹⁴ Piragoff and Robinson argue that, despite the fact that ‘recklessness’ was not included in Article 30 of the Rome Statute, ‘it may be open to the Court to interpret the provision to include some of these aspects’.⁹⁵

A further entry point for the relevance of risk taking behavior and to standards lower than direct intent in the law of conduct of hostilities crimes has been provided by a string of decisions of the International Tribunal for the former Yugoslavia. As has been mentioned, an additional distinction between the legal framework of conduct of hostilities crimes in the Rome Statute and AP I is the requirement of ‘willfulness’. Article 85 AP I requires that ‘grave breaches’, which include the war crime of launching an attack against the civilian population, must be committed ‘willfully’. In a series of decisions, the International Tribunal for the Former Yugoslavia⁹⁶ and the International Committee of the Red Cross (ICRC)⁹⁷ interpreted ‘willfulness’ as

⁹³ Olasolo (n 74) 208.

⁹⁴ Cassese affirmed that ‘Article 30 of the Rome Statute defines the mental element of crimes as consisting of intent and knowledge. While it is no doubt meritorious to have defined these two notions, it appears questionable to have excluded recklessness as a culpable *mens rea* under the Statute. One fails to see why, at least in the case of war crimes, this last mental element may not suffice for criminal responsibility to arise. Admittedly, in the case of genocide, crimes against humanity and aggression, the extreme gravity of the offence presupposes that it may only be perpetrated when intent and knowledge are present. However, for less serious crimes, such as war crimes, current international law must be taken to allow for recklessness: for example, it is admissible to convict a person who, when shelling a town, takes a high and unjustifiable risk that civilians will be killed — without, however, intending, that they be killed — with the result that the civilians are, in fact, thereby killed’, Antonio Cassese, ‘The Statute of the International Criminal Court: Some Preliminary Reflections’ (1999) 10 EJIL 144, 153-154.

⁹⁵ Piragoff and Robinson (n 85) 1122 footnote 77.

⁹⁶ Judgment, *Galić* (IT-98-29-T, TC I), Trial Chamber, 5 December 2003, para 54, affirmed in Judgment, *Galić* (IT-98-29-A), Appeals Chamber, 30 November 2006, para 140. Subsequently, Judgment, *Strugar* (IT-01-42-A), Appeals Chamber, 17 July 2008, para 270; Judgment, *Martić* (IT-95-11-T), Trial Chamber I, 12 June 2007, para 72; Judgment, *Perišić* (IT-04-81-T), Trial Chamber I, 6 September 2011, para 100.

⁹⁷ ICRC Commentary to Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, of 1987, Article 85, para 3474. The ICRC interprets ‘willfully’ as follows: ‘*the accused must have acted consciously and with intent, i.e., with his mind on the act and its consequences, and willing the ("criminal intent" or "malice aforethought"); this encompasses the concepts of "wrongful intent" or "recklessness", viz., the attitude of an agent who, without being certain of a particular result, accepts the possibility of it happening; on the other hand, ordinary negligence or lack of foresight is not*

encompassing direct intent, indirect intent, as well as ‘recklessness’.⁹⁸ Interpreting ‘willfulness’ as also encompassing ‘recklessness’ makes criminally responsible for this offence the military commander who launches an attack and foresees the possibility of hitting civilians.

This development has been criticized for diluting the rationale underlying the notion of the war crime of attacking civilians, which is also the main feature distinguishing this offence from disproportionate attacks: the concept of intent.⁹⁹ While the former criminalizes violation of the obligation to distinguish civilian targets from military targets (principle of distinction),¹⁰⁰ the prohibition of disproportionate attacks criminalizes the violation of the principle of proportionality.¹⁰¹ Therefore, it prohibits *attacks directed at military targets causing civilian harm*, but only when civilian deaths are *unintended* and when they *outweigh the expected military advantage* of the attack. The lowering of the mental element to prosecute for the war crime of attacking civilians the military commander who launches an attack and foresees the possibility of hitting civilians would: 1) ‘squeeze almost all of the targeting cases into’ the prosecution of violations of the principle distinction (i.e. into the crime of attacking civilians); and therefore 2) sidestep the more complicated proportionality assessment, making it ‘legally irrelevant’.¹⁰²

While these criticisms are founded and, more generally, although this discussion has demonstrated how taking into account propensity for risk taking behavior in the *mens rea* analysis could allow for perpetrators of violations of IHL via semi-autonomous weapons to be prosecuted for war crimes through the standards of *dolus eventualis* or recklessness, these issues are not without controversy.

covered, i.e., when a man acts without having his mind on the act or its consequences (14) (although failing to take the necessary precautions, particularly failing to seek precise information.)

⁹⁸ See *contra* Eser who considers ‘willfulness’ a stricter requirement than ‘intention’, Albin Eser, ‘Mental Elements—Mistake of Fact and Mistake of Law’ in Antonio Cassese, Paola Gaeta and John RWD Jones (eds), *The Rome Statute of the International Criminal Court A Commentary* (OUP 2002) 899.

⁹⁹ Jens David Ohlin, ‘Targeting and the Concept of Intent’ (2013) 35 *Michigan Journal of International Law* 79, 86.

¹⁰⁰ Article 48 AP I with accompanying rules in Articles 51 and 52 AP I.

¹⁰¹ The principle of proportionality ‘expressly prohibits attacks against a military objective if such attacks are expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects or a combination thereof which would be excessive in relation to the concrete and direct military advantage anticipated’, see Emanuela-Chiara Gillard, ‘Proportionality in the Conduct of Hostilities The Incidental Harm Side of the Assessment Proportionality in the Conduct of Hostilities’ (2018) 6. In AP I the rule appears three times: first, in Article 51(5)(b), as an example of an attack that is indiscriminate and therefore prohibited; and in Article 57(2)(a) (ii) and (iii) in relation to precautions in attack.

¹⁰² Ohlin (n 99) 86 and 87.

Nonetheless, it is important to acknowledge the existence of authoritative interpretations of the mental element of the crime of attacking civilians as including recklessness. Moreover, it is submitted that mental elements lower than direct intent for the war crime of attacking civilians could have important effects in terms of increasing standards of precautions and deterrence when semi-autonomous and autonomous weapons are deployed.

To conclude, the above shows there might be room to take into account risk taking behaviors in the law of conduct of hostilities crimes. The decisive factors that make risk taking behaviors in the use and deployment of semi-autonomous weapons punishable might be (depending on the theory adopted): risk awareness, distinguishing *dolus eventualis* and recklessness from negligence; the subjective posture (approval or indifference vs aversion) towards these risks, distinguishing *dolus eventualis* from recklessness; and the actual risks (certain or highly probable in *dolus directus* in the second degree, probable or possible in *dolus eventualis*). Given the place of forms of responsibility based on risk awareness in the prosecution of conduct of hostilities crimes, in the debate on autonomous weapons and criminal responsibility we should therefore critically analyze the factors and processes that might affect the relevant components of these *mens reas*.

VI How New Features of the Human-Weapon Relationship May Affect the Proof of the Mental Element

The human operators of semi-autonomous systems who risk harm to civilians may, depending upon the interpretation one adopts and the degree of likelihood of these risks, be prosecuted for the war crime of attacking civilians.

The following discussion will therefore scrutinize the introduction of automated aids or artificial intelligence in weapons against studies on decision making and risks in order to analyze its main effects on the human operators of semi-autonomous weapons systems and on their risk taking behavior. It is submitted that these effects must be taken into consideration in the determination of *mens rea* for the war crime of attacking civilians.

Affirming that risk taking (and not only direct intent) has a place in the law and in the prosecution of conduct of hostilities crimes renders it important to look at the cognitive factors that affect militaries' risk perception and preferences. This makes

research on the cognitive science of the psychology of decision making, heuristics and biases applicable and relevant to this context.

Our starting point is that studies show that (military) decision makers do not necessarily adhere to the principle of rational decision making. The rational analytical approach to decision making involves the weighting of different options, and the choosing of the optimum solution, (i.e. the solution with the highest expected value) by application of the law of probability.¹⁰³ However, it has been largely proven that humans deviate from the rational model. There are instead a broad range of naturalistic decision making theories,¹⁰⁴ hinting at the departure from the law of probabilities and highlighting the role of other factors affecting decision making.

Klein found that concurrent option comparison is hardly ever employed by experienced decision makers, such as military commanders. Instead of analytical decision making they employ recognition-primed decision making, which is described as the ability

‘to use their experience to recognize a situation as familiar, which gives them a sense of what goals are feasible, what cues are important, what to expect next and what actions are typical in that situation. The ability to recognize the typical action means that experienced decision makers do not have to do any concurrent deliberation about options’¹⁰⁵

Experienced decision makers employ an intuitive ‘pattern-matching process’ that allow them to draw from experience in order to generate the first solution to a problem.¹⁰⁶ As aptly affirmed by Knighton: ‘an individual’s experience will be key to his understanding of the situation, his perception of the outcome and the likelihood that it will occur’.¹⁰⁷ Recognition-primed decision making, which is a form of intuitive decision making, appears to be best suited in stressful and fast-paced scenarios and appropriate for well-experienced decision makers. In contrast, according to Klein, the rational analytical model is apt to be employed in scenarios

¹⁰³ RJ Knighton, ‘The Psychology of Risk and Its Role in Military Decision-Making’ (2005) 4 Defence Studies 309, 311–313.

¹⁰⁴ A good summary of different studies is provided by Knighton (n 103).

¹⁰⁵ Gary A Klein, ‘Strategies Decision Making’ [1989] Military Review 56, 58.

¹⁰⁶ *ibid* 59.

¹⁰⁷ Knighton (n 103) 323.

where there is ample time for decisions and by novice decision makers or by experienced ones facing novel situations.¹⁰⁸

Before turning to the heuristics and biases that affect intuitive decision making, it is important to first clarify which type of decisions become relevant for the fulfillment of the *mens rea* of the crime of attacking civilians. Relevant for our discussion is the decision making process of the commander/soldier who does not want to hit civilians but who nevertheless accepts some risk that this will occur. Therefore, in relation to the crime of attacking civilians the decision making process concerns the likelihood and probability of a certain outcome, i.e. of killing civilians.

The probability assessment occurring in the *mens rea* of the perpetrator of the war crime of attacking civilians is an important difference with the war crime of disproportionate attacks (Article 8(2)(b)(iv) of the Rome Statute and Articles 85(3)(b) and 51(5)(b) AP I), the analysis of which is beyond the scope of the present paper. In the definition of the Rome Statute, which departs partially from the norms contained in AP I, disproportionate attacks are those attacks against military objectives which are expected to incidentally cause civilian damage that is clearly excessive in the relation to the concrete and direct overall military advantage anticipated. The assessment of incidental civilian damage is a complex qualitative judgment weighing civilian losses and damages against the military advantage anticipated,¹⁰⁹ yet it also contains some predictions about the likelihood that civilian harm will occur.¹¹⁰ Rather than being only a judgment on probability of civilian losses, it is submitted that the incidental loss assessment can first and foremost be equated to a decision problem. Decision problems are ‘defined by the acts or options which one must choose, the possible outcomes or consequences of these acts, and the contingencies or conditional probabilities that relate outcomes to acts’.¹¹¹ Relevant to decision problems are studies on the impact on decision making of the framing of a problem. It has been proved that when a problem is framed in terms of loss (such as excessive collateral damage for a particular target) rather than gains, decision makers ‘exhibit a propensity for risk-taking’.¹¹² Military decision makers show an even greater readiness to take

¹⁰⁸ Klein (n 105) 61.

¹⁰⁹ See Marco Sassòli, Can Autonomous Systems Respect the Principles of Distinction, Proportionality and Precaution? ICRC (n 1) 83.

¹¹⁰ Gillard (n 101) 17 and 18.

¹¹¹ Amos Tversky and Daniel Kahneman, ‘The Framing of Decisions and the Psychology of Choice’ (1981) 211 *Science* 453, 453.

¹¹² Knighton (n 103) 314.

risks when problems are expressed in terms of losses,¹¹³ such as in incidental harm assessments.

Having defined the problem at stake as one of probability, the main submission of this paper is that military personnel are not exempt from heuristics and biases, and these are exacerbated by the implementation of automation and artificial intelligence in weapon systems. The increasing importance of cognitive ecology for cognitive science requires us to approach cognitive phenomena in context.¹¹⁴ It presses us to take into consideration the web of elements that form part of a cognitive ecosystem, the environment in which cognitive processes develop and operate. Human cognitive activity does not operate in isolation and is integrated in its environment. Human cognitive activity and the human decisions are the result of the interaction between internal cognitive processes and the environment, which in the case of targeting also comprises machines. When automation or artificial intelligence is introduced the decision making environment changes. The introduction of automated aids has an impact on the cognitive analysis of a situation on the part of the human operator. Whether the human agent retains the final decision to engage or has the power to intervene in an otherwise autonomous targeting cycle, it is compelling to analyze how the decision making process may be affected by the introduction of automated aids or artificial intelligence and the heuristics and biases relevant to this.

When discussing heuristics and biases in decision making, one cannot but apply the seminal studies on ‘Judgment under Uncertainty’ conducted by Kahneman and Tversky. In their work, they studied the cognitive shortcuts employed in decision making. They describe three heuristics.

The first is the ‘representativeness’ heuristic, according to which individuals evaluate the probability of a particular outcome by the degree to which such an outcome is representative of stereotypes given in the problem scenario. For our purposes this reflects the evaluation of the probability that civilians will be hit through the integration of automated systems or AI in the targeting process on the basis of the narrative surrounding the use of these systems. In the ‘representativeness’ heuristic the stereotypes concerning and the narrative surrounding automated systems or AI systems play a major role. If machines are described as highly reliable, the military will tend to predict scenarios that appear most representative of this description,

¹¹³ *ibid* 315.

¹¹⁴ Edwin Hutchins, ‘Cognitive Ecology’ (2010) 2 *Topics in Cognitive Science* 705.

leading them to, for example, not intervene in an autonomous targeting cycle or to not confirm the data provided by machines. This will also lead them to ignore other factors that limit the accuracy of the prediction, such as prior probabilities and the reliability of the evidence.¹¹⁵

To the extent that automated information and automation feedback is believed to be highly reliable, the human operator reduces his cognitive efforts to seek out other information. This process can be referred to as the ‘tendency toward “satisficing” in judgment and decision making’.¹¹⁶ Moreover, the human cognitive process may be affected by automation bias which leads the human operator to over-attend and over-trust the automated data or information. By the same token, contradictory information tends to be interpreted by the human operator as consistent with automation feedback (confirmation bias).¹¹⁷ Automation biases may therefore interfere with a soldier or commander’s judgement and decision making. For example, commanders or soldiers deploying LAWS may overly rely on the information and recommendations provided by automated processes without seeking contradictory information regarding the lawfulness of the target. Or, they might have no reason to believe that the device will malfunction and violate international humanitarian law.

The second heuristic concerns the ‘availability of instances and scenarios, which is often employed when people are asked to assess the frequency of a class or the plausibility of a particular development’.¹¹⁸ The probability of an event is assessed by the ‘ease by which instances or occurrences can be brought to mind’.¹¹⁹ The availability bias stresses the importance of the retrievability of previous instances and accidents. If an outcome, such as LAWS mistakenly hitting civilians or an automated system providing wrong data targets, is vividly imaginable, the probability of its occurrence is likely to be overestimated. This might in turn lead to, for example, lack of reliance on automated systems or reluctance to incorporate or use data provided by AI systems, impairing the whole targeting process. Military culture could also be relevant here, in terms of there being a preference towards human ‘expertise’.

¹¹⁵ Amos Tversky and Daniel Kahneman, ‘Judgment under Uncertainty: Heuristics and Biases’ (1974) 185 *Science* 1124, 1124–1126.

¹¹⁶ Kathleen L Mosier and Linda J Skitka, ‘Human Decision Makers and Automated Decision Aids: Made for Each Other?’ in Raja Parasuramana and Mustapha Mouloua (eds), *Automation and Human Performance: Theory and Applications* (CRC Press 1996) 205.

¹¹⁷ *ibid* 204.

¹¹⁸ Tversky and Kahneman (n 115) 1131.

¹¹⁹ *ibid* 1127.

Thirdly, anchoring and adjusting is the final heuristic identified by Kahneman and Tversky. ‘Adjustment from an anchor [...] is usually employed in numerical predictions when a relevant value is available’.¹²⁰ This shows that people make estimations and probability judgments adjusting the ‘initial value’ or ‘anchoring point’ that is suggested by the formulation of the problem or that may be the result of a partial calculation.¹²¹ It is reasonable to assume that this heuristic could lead, for example, military commanders to not depart from or only minimally question the data provided by automated systems.

Moreover, there are problems linked to the complexity of the weapon system and of the targeting cycle as a whole and the fragmentation of tasks and roles. By delegating some targeting functions to the system, there is also an overall decrease in vigilance and situational awareness by the human operator. This may lead to his failure to detect errors or to assess the risks related to a certain targeting decision.¹²²

Connected to anchoring is a general tendency to underestimate the probability of disjunctive events and overestimate the probability of conjunctive events. Applying this to complex systems, such as semi-autonomous weapons, entails that people underestimate the probability of failures of complex systems. However, even if the likelihood of a failure of each single part of the weapon system is low, ‘the probability of an overall failure can be high if many components are involved’.¹²³ As also affirmed by Dalcher, ‘the more complex the system, the more failure points it harbors. Automated protection eliminates some of them, but at the cost, and risk, of introducing others. The possibility of massive failures that may affect the whole system is increased as more components (including humans and their decisions) interact’.¹²⁴

To conclude, biases in the decision making processes of military operators of semi-autonomous weapons could be broadly due to the: (1) narrative of high-reliability surrounding automated and autonomous systems; (2) complexity of the targeting cycle and of the weapon system; and (3) availability of instances of blatant failures or mistakes by automated or autonomous systems.

¹²⁰ *ibid* 1131.

¹²¹ *ibid* 1128.

¹²² Mosier and Skitka (n 116) 207 and 208.

¹²³ Tversky and Kahneman (n 115) 1129.

¹²⁴ Darren Dalcher, ‘Why the Pilot Cannot Be Blamed: A Cautionary Note about Excessive Reliance on Technology’ (2007) 7 *International Journal of Risk Assessment and Management* 350, 363.

Biases may differently affect the different roles that human operators perform with regard to target selection and engagement. Their effects on the awareness of risks and on risk propensity have obvious repercussions on the shaping of *mens rea* of the human perpetrator of the war crime of attacking civilians.

Conclusion

This paper has dealt with the advent of automation and artificial intelligence in warfare and with the creation of new forms of human-weapon command-and-control relationships. Human-machine teaming, in different forms, appears to be the present and likely future of targeting processes. This entails the creation of new cognitive environments in which human operators are required to take decisions.

Given the centrality of culpability to criminal responsibility, this paper has inquired into the standards of *mens rea* for conduct of hostilities crimes, focusing on the war crime of attacking civilians. Having established the possible relevance of forms of criminal responsibility based on risk-taking behaviors, the paper has addressed the processes of military decision making. This paper highlighted several heuristics that could bear on decision making processes and focused on how human-machine interaction may alter military decision making processes. It is submitted that biases in military decision making are exacerbated by the introduction of automation and autonomy in weapons. A further relevant finding is that most heuristics and biases could depend on the narrative of semi-autonomous weapons adopted by the military. Biases bear on humans' risk awareness and risk propensity. These are elements that are crucial for the determination of *mens rea*. However, given the different roles that human could perform with regard to semi-autonomous weapons, further research that concentrates on the effects of heuristics and on the exact extent of their impact on the assessment of *mens rea* for conduct of hostilities crimes is required.