

OF SOULS, SPIRITS AND GHOSTS: TRANSPOSING THE APPLICATION OF THE RULES OF TARGETING TO LETHAL AUTONOMOUS ROBOTS

TETYANA (TANYA) KRUPIY*

The article addresses how the rules of targeting regulate lethal autonomous robots. Since the rules of targeting are addressed to human decision-makers, there is a need for clarification of what qualities lethal autonomous robots would need to possess in order to approximate human decision-making and to apply these rules to battlefield scenarios. The article additionally analyses state practice in order to propose how the degree of certainty required by the principle of distinction may be translated into a numerical value. The reliability rate with which lethal autonomous robots need to function is identified. The article then analyses whether the employment of three categories of robots complies with the rules of targeting. The first category covers robots which work on a fixed algorithm. The second category pertains to robots that have artificial intelligence and that learn from the experience of being exposed to battlefield scenarios. The third category relates to robots that emulate the working of a human brain.

CONTENTS

I	Introduction.....	2
II	Definition of Lethal Autonomous Robots.....	3
III	The State of Scientific Knowledge and Aspirations	8
IV	State Practice.....	11
V	The Rules of Targeting	12
	A The Principle of Distinction	12
	B The Rule of Target Verification	14
	C The Principle of Proportionality	16
	D The Principle of the Least Feasible Damage	18
VI	Criteria for Regulating Lethal Autonomous Robots	20
	A The Role of ‘Situational Awareness’ and Judgement.....	20
	B The Role of Ethics	28
	C An Ability to Experience Emotions and Compassion	31
	D A Numerical Expression of the Principle of Distinction	33
	1 Landmines	34
	2 Air Bombing.....	35
	3 Cluster Munitions.....	38
	4 Synthesis of the Analysis	40
	E Reliability Rate.....	41
VII	Ability of Autonomous Robots to Fulfil Legal Obligations	44
	A Lethal Autonomous Robots that Work on ‘Simple’ Algorithms	44
	B Legality of Autonomous Robots that Emulate the Human Brain.....	51
VIII	Conclusion	57

* LLB (London School of Economics), LLM (London School of Economics), PhD (University of Essex), freelance legal consultant. I would like to thank Karen Hulme, Marco Sassòli, Geoff Gilbert and the anonymous referees for their valuable feedback on this article. Additionally, I would like to thank Joanna Bryson for the opportunity to discuss issues relating to artificial intelligence with computer scientists who specialise in this area.

I INTRODUCTION

There has been a rapid pace in the development of technologies that are available to states for use on the battlefield. In the past 15 years the academic debate has shifted from the question of whether it is lawful for pilots to operate at high altitude¹ to the question of whether it is lawful for machines with artificial intelligence to make decisions about whom to target without human oversight.² States are currently using the United Nations as an arena in which to address the pivotal question of whether the employment of lethal autonomous robots ('LARs') complies with international humanitarian law ('IHL').³ At this stage, different groups of states have adopted varying positions on this issue. On one end of the spectrum, states such as Costa Rica⁴ and Pakistan⁵ have declared that these systems should not be developed. On the other end of the spectrum, the United States has said that it will persist in developing this technology.⁶ The observation of Steve Omohundro, a physicist and artificial intelligence specialist at the research centre Self Aware Systems,⁷ provides context for the discussions which are taking place between states in the United Nations. According to him '[a]n autonomous weapons arms race is already taking place'.⁸

The military utility of employing LARs makes this technology appealing to some states.⁹ Since robots are a force multiplier, militaries require fewer soldiers when they have robots.¹⁰ Robots allow parties to the conflict to conduct military operations over a wider area, in addition to allowing them to strike the enemy at longer range.¹¹ Moreover, the employment of robots reduces soldier casualties because robots may be tasked with very dangerous missions.¹² Another appeal of

¹ Marina Mancini, 'Air Operations against the Federal Republic of Yugoslavia' (1999) in Natalino Ronzitti and Gabriella Venturini (eds), *The Law of Air Warfare: Contemporary Issues* (Eleven International, 2006) 273, 275–8; A P V Rogers, 'Zero-Casualty Warfare' (2000) 837 *International Review of the Red Cross* 165; Alexandra Boivin, 'The Legal Regime Applicable to Targeting Military Objectives in the Context of Contemporary Warfare' (Research Paper Series No 2, Geneva Academy of International Humanitarian Law and Human Rights, 2006) 46–7.

² William Boothby, 'Some Legal Challenges Posed by Remote Attack' (2012) 94 *International Review of the Red Cross* 579, 584; Jonathan David Herbach, 'Into the Caves of Steel: Precaution, Cognition and Robotic Weapon Systems under the International Law of Armed Conflict' (2012) 4(3) *Amsterdam Law Forum* 3, 5.

³ Andras Kos, 'European Union Statement' (Speech delivered at the Meeting of the High Contracting Parties to the *Convention on Certain Conventional Weapons*, Geneva, 14 November 2013) 4; Thomas Gürber, 'Thematic Debate on Conventional Weapons' (Speech delivered at the UN General Assembly, 1st Comm, 68th sess, 28 October 2013) 3.

⁴ Maritza Chan, 'General Debate on Conventional Weapons' (Speech delivered at the UN General Assembly, 1st Comm, 68th sess, 18 October 2013) 3.

⁵ UN GAOR, 1st Comm, 68th sess, Agenda Items 89 to 107, UN Doc A/C.1/68/PV.9 (16 October 2013) 7.

⁶ Euthimis Tsiliopoulos, 'Killer Robots Are Coming!', *The Times of Change* (online), 15 May 2014 <<http://www.thetoc.gr/eng/technology/article/killer-robots-are-coming>>.

⁷ John Markoff, 'Fearing Bombs That Can Pick Whom to Kill', *The New York Times* (online), 11 November 2014 <http://www.nytimes.com/2014/11/12/science/weapons-directed-by-robots-not-humans-raise-ethical-questions.html?_r=1>.

⁸ *Ibid.*

⁹ Gary E Marchant et al, 'International Governance of Autonomous Military Robots' (2011) 12 *Columbia Science and Technology Law Review* 272, 275.

¹⁰ *Ibid.*

¹¹ *Ibid.*

¹² *Ibid.*

robots is that they hold a promise of what Martin Shaw calls ‘clean’ wars.¹³ Robots are programmed in such a way that their judgement is not clouded by emotions and by a desire for self-preservation.¹⁴ Proponents of the employment of robotic systems argue that they will, therefore, be more circumspect about when to open fire than humans.¹⁵ Finally, it is cheaper to maintain robots than the armed forces which are comprised of soldiers.¹⁶ Given the fact that states have adopted varying viewpoints on the question of whether the rules of targeting adequately address the employment of LARs on the battlefield, a fundamental question is whether such systems are capable of compliance with the relevant rules.

In order to provide a groundwork for this assessment, it will first be explained how states define LARs, what the current state of scientific knowledge is and what technologies scientists envision developing in the future. The obligations imposed by the rules of targeting and the qualities required for applying these rules will be examined. How these elements translate to the context of machine decision-making and what additional criteria the law might require of machines will then be analysed. After formulating the standards by which the rules of targeting regulate machine decision-making, three types of robot categories will be evaluated for their compliance with the relevant rules. The first category covers robots that function based on a fixed algorithm. The second category involves robots that have artificial intelligence and learn from being continuously exposed to battlefield scenarios, but do not emulate the working of a human brain. The third category relates to robots that work on an algorithm that mimics the working of a human brain.

II DEFINITION OF LETHAL AUTONOMOUS ROBOTS

The starting point of the discussion is that it is possible to design different kinds of robots. The British Army, for instance, employs robots for bomb disposal.¹⁷ The armed forces may start to use robots to carry equipment in the near future.¹⁸ Overall, although robots may perform different functions, they perform on the basis of common mechanisms. The two main categories of robots are those which are ‘automated’ systems and those which are ‘fully autonomous’ systems.¹⁹ There are also robots with varying degrees of autonomy which fall

¹³ Martin Shaw, *The New Western Way of War* (Polity, 2005) 87–8.

¹⁴ Marchant et al, above n 9, 280.

¹⁵ Ibid.

¹⁶ International Committee of the Red Cross, ‘Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects’ (Report, 26–28 March 2014) 13 (*Autonomous Weapons Systems Report*).

¹⁷ The British Army, *Dragon Runner Bomb Disposal Robot* (2015) <<https://www.army.mod.uk/equipment/23256.aspx>>.

¹⁸ Stew Magnuson, ‘Robotic Mule Vendors Seek Opportunities Outside Military’, *National Defense Magazine* (online), July 2013 <<http://www.nationaldefensemagazine.org/archive/2013/July/Pages/RoboticMuleVendorsSeekOpportunitiesOutsideMilitary.aspx>>.

¹⁹ Development, Concepts and Doctrine Centre, ‘Unmanned Aircraft Systems: Terminology, Definitions and Classification’ (Joint Doctrine Note 3/10, Ministry of Defence, May 2010) [108], [109] (*Unmanned Aircraft Systems*).

in-between these two categories.²⁰ To explain, engineers use the term ‘automated systems’ to refer to unsupervised systems or processes that involve repetitive, structured and routine operations without much feedback information.²¹ An example of such an ‘automated’ system is a dishwasher.²² Turning to the battlefield context, an example of an automated weapon is an anti-vehicle landmine.²³ Depending on the model, these may be activated by wheel pressure from vehicles, acoustics, magnetic influence, radio frequencies, infrared signature or disturbance.²⁴ Another example is the Sensor Fuzed Weapon,²⁵ which detects objects that match a pre-programmed profile and that emit a heat signature.²⁶ Furthermore, sentry robots relay a signal to a commander that a human has been detected and may be instructed to engage the target.²⁷ In terms of the North Atlantic Treaty Organization (‘NATO’) four-tier test for autonomy, the sentry robots fall under level one. Level one systems are remotely controlled by an operator and depend on operator input.²⁸ On the other hand, anti-vehicle landmines and Sensor Fuzed Weapons fall under level two. Level two covers automated systems that rely on pre-programmed settings for their behaviour and that are not remotely controlled.²⁹

The engineers employ the term ‘autonomous’ to designate systems that: (1) are self-governing; (2) operate without direct human control or supervision; and (3) function in changing and unstructured environments.³⁰ These systems use

²⁰ NATO Industrial Advisory Group Sub-Group/75, ‘NIAG SG/75: UAV Autonomy’ (Paper, NATO Industrial Advisory Group, 2004) 43 <http://uvs-info.com/phocadownload/05_3g_2005/28_NATO.pdf> (‘NIAG SG/75’).

²¹ The dictionary meaning of to ‘automate’ is ‘to use machines or computers instead of people to do a particular task, especially in a factory or office’. This meaning appears across different dictionaries. The problem with the definition is that it does not distinguish between automated and autonomous systems. The better definition is that a key component of automation is that a machine performs a repetitive process. This definition also reflects the fact that engineers define autonomous systems as machines which operate without human oversight in unstructured environments. See Cambridge University Press, *Cambridge Dictionaries Online* <<http://dictionary.cambridge.org/dictionary/business-english/automate?q=automated>>; Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*, 23rd sess, Agenda Item 3, UN Doc A/HRC/23/47 (9 April 2013) 8; Peter Asaro, ‘On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making’ (2012) 94 *International Review of the Red Cross* 687, 690 n 5; Panos J Antsaklis, ‘Setting the Stage: Some Autonomous Thoughts on Autonomy’ (Paper presented at the IEEE International Conference on Robotics and Automation, Gaithersburg, 14–17 September 1998) 520.

²² Asaro, above n 21, 690 n 5.

²³ Alan Backstrom and Ian Henderson, ‘New Capabilities in Warfare: An Overview of Contemporary Technological Developments and the Associated Legal and Engineering Issues in Article 36 Weapons Reviews’ (2012) 94 *International Review of the Red Cross* 483, 488.

²⁴ United States General Accounting Office, ‘Military Operations: Information on US Use of Land Mines in the Persian Gulf War’ (Report No GAO-02-1003, September 2002) 5 (‘*US Use of Land Mines in the Persian Gulf War*’).

²⁵ Backstrom and Henderson, above n 23, 488.

²⁶ Textron Defense Systems, ‘CBU–105 Sensor Fuzed Weapon/BLU–108 Submunition’ (Information Pamphlet, 2014) <http://www.textrondefense.com/sites/default/files/pdfs/product-info/sfw_brochure-small.pdf> (‘*CBU–105/BLU–108*’).

²⁷ Backstrom and Henderson, above n 23, 488.

²⁸ NIAG SG/75, above n 20, 43.

²⁹ Ibid.

³⁰ Antsaklis, above n 21, 520.

feedback information from a variety of sensors to orientate themselves.³¹ At present, robots can be equipped with sensors such as cameras, infrared, sonars, lasers, temperature sensors and ladars.³² The sonars use sound wavelengths to determine the range and orientation of objects.³³ Passive sonars detect sounds.³⁴ The ladars employ light wavelengths (lasers) to measure distance at which the object is located and to re-create the object in 3D.³⁵ Infrared sensors detect the emission of infrared waves.³⁶ All sources of heat, including human beings, emit infrared waves.³⁷ Currently, NATO identifies two types of ‘autonomous’ systems. Level three systems are autonomous non-learning systems which function based on a pre-programmed set of rules.³⁸ At the end of the NATO scale (level four) are autonomous self-learning systems.³⁹ These systems function based on two sets of rules.⁴⁰ The system is unable to modify core rules, such as the rules of targeting.⁴¹ However, it is able to continuously modify non-core rules as it learns from experience by, for instance, being exposed to battlefield scenarios.⁴² Currently, researchers are studying how one could program LARs to learn from experience.⁴³ Their goal is to have a robot that has artificial intelligence and that, by integrating information about its previous experience, is able to respond to novel situations.⁴⁴

The ‘automated’ and ‘autonomous’ robots work on the same principles.⁴⁵ They follow ‘fixed and deterministic’ algorithmic instructions.⁴⁶ Algorithmic instructions are a set of rules which a computer follows in order to compute a number or to perform a task. They are written in the form: if condition X is true, then perform operation Z.⁴⁷ Thus, the robotic system determines the character of the object in front of it ‘based on pre-programmed characteristics, such as shape

³¹ United States Air Force, ‘Unmanned Aircraft Systems Flight Plan 2009–2047’ (Flight Plan, 18 May 2009) 33 (*‘Unmanned Aircraft Systems Flight Plan’*).

³² Noel E Sharkey, ‘The Evitability of Autonomous Robot Warfare’ (2012) 94 *International Review of the Red Cross* 787, 788.

³³ National Ocean Service, *What is Sonar?* (23 January 2014) National Oceanic and Atmospheric Administration <<http://oceanservice.noaa.gov/facts/sonar.html>>.

³⁴ *Ibid.*

³⁵ Sensors Unlimited, *Laser Radar/LIDAR/LADAR Including Eye-Safe Lasers* (2015) UTC Aerospace Systems <<http://www.sensorsinc.com/LADAR.html>>.

³⁶ National Aeronautics and Space Administration, Science Mission Directorate, *Infrared Waves* (14 August 2014) <http://missionscience.nasa.gov/ems/07_infraredwaves.html>.

³⁷ *Ibid.*

³⁸ *NIAG SG/75*, above n 20, 43.

³⁹ *Ibid.*

⁴⁰ *Ibid.*

⁴¹ *Ibid.*

⁴² *Ibid.*

⁴³ US Department of Defense, ‘Unmanned Systems Integrated Roadmap FY2013–2038’ (Roadmap No 14-S-0553, 2014) 67 <<http://www.defense.gov/pubs/DOD-USRM-2013.pdf>> (*‘Unmanned Systems Integrated Roadmap’*); Marchant et al, above n 9, 284.

⁴⁴ Marchant et al, above n 9, 284; Backstrom and Henderson, above n 23, 493.

⁴⁵ Roy Featherstone and David Orin, ‘Robot Dynamics: Equations and Algorithms’ (Paper presented at the IEEE International Conference on Robotics and Automation, San Francisco, 2000) 826–7.

⁴⁶ Asaro, above n 21, 690 n 5.

⁴⁷ Ronald C Arkin, ‘Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture’ (Technical Report No GIT-GVU-07-11, Georgia Institute of Technology, 2008) 67.

and dimensions'.⁴⁸ This means that once a robot identifies a sufficient number of characteristics which it can reconcile with the pre-programmed list of objects, the robot will classify the object in front of it as, for instance, a military objective.⁴⁹ 'This type of matching is mechanical, based on quantitative data'.⁵⁰ What distinguishes 'autonomous' systems from 'automated' systems is that they employ stochastic, or probability-based, reasoning.⁵¹ This means that there is a degree of uncertainty as to what decision an autonomous system will take.⁵²

Having looked at how engineers define the difference between autonomous and automatic systems, it will now be explained how states define LARs. So far only the US and the United Kingdom have made publicly available national policies on autonomous weapon systems.⁵³ The US Department of Defense defines an autonomous robotic system as a 'system that, once activated, can select and engage targets without further intervention by a human operator'.⁵⁴ In its more recent publication *Unmanned Systems Integrated Roadmap FY 2013–2038* the US provided a more detailed description of the capabilities of these systems.⁵⁵ In determining which target to engage, an LAR should be capable of responding to the unfolding situation on the battlefield and to deviate from the pre-programmed mission.⁵⁶ After choosing which target to engage, it will develop a plan of action to fulfil the selected mission independently of human control.⁵⁷ In order for LARs to be able to select and engage targets without human control once on the battlefield, they will need to have the ability to 'integrate sensing, perceiving, analyzing, communicating, planning, decision-making, and executing' capabilities.⁵⁸ In terms of the NATO autonomy scale, these systems fall under level four, namely under self-learning autonomous systems.

In a parliamentary debate the UK defined the term LARs as referring to: 'robotic weapons systems that, once activated, can select and engage targets without further intervention by a human operator'.⁵⁹ This definition is identical to that given by the US to 'autonomous weapon systems' in its Department of Defense Directive 3000.09.⁶⁰ Expanding on this definition, the UK military doctrine states that an autonomous system: (1) operates in an unstructured environment; (2) is capable of choosing between alternative courses of actions without human oversight or control following receipt of information from

⁴⁸ Markus Wagner, 'Taking Humans Out of the Loop: Implications for International Humanitarian Law' (2011) 21 *Journal of Law, Information and Science* 155, 161.

⁴⁹ *Ibid.*

⁵⁰ *Ibid.*

⁵¹ *Autonomous Weapons Systems Report*, above n 16, 13.

⁵² *Ibid.*

⁵³ *Ibid.* 8.

⁵⁴ US Department of Defense, 'Department of Defense Directive 3000.09: Autonomy in Weapon Systems' (Government Directive, 21 November 2012) 13 <<http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf>> ('*Autonomy in Weapon Systems*').

⁵⁵ *Unmanned Systems Integrated Roadmap*, above n 43, 66.

⁵⁶ *Ibid.* 66–7.

⁵⁷ *Ibid.*

⁵⁸ *Ibid.* 67.

⁵⁹ United Kingdom, *Parliamentary Debates*, House of Commons, 17 June 2013, vol 564, col 729 (Nia Griffith).

⁶⁰ *Autonomy in Weapon Systems*, above n 54, 13.

sensors; (3) makes decisions which contribute to the achievement of the strategic objectives of the military campaign; and (4) is capable of the same understanding of the situation as a human.⁶¹ Such a system, the doctrine goes on to suggest, will not follow a 'pre-written set of rules or instructions'.⁶² The doctrine concludes by saying that the current state of technology does not allow for autonomous robots.⁶³ Since these systems emulate the working of a human brain, they are capable of greater autonomy than that envisaged by the NATO scale.

There is a gap between the definition of LARs in the US and the UK military doctrines. The US doctrine merely envisages that an autonomous system is able to perform a mission to a high standard using algorithmic programming.⁶⁴ Meanwhile, the UK doctrine excludes all machines which work on preset instructions from being autonomous.⁶⁵ Furthermore, the US,⁶⁶ unlike the UK,⁶⁷ did not explicitly state that LARs should understand the situation and respond to it in the same way as a human would. Consequently, the UK requires that higher levels of autonomy should be achieved before the decision-making authority may be delegated to such systems.

Although the US and the UK have defined what LARs are, there is currently no common definition.⁶⁸ Japan articulated in the United Nations General Assembly the need for a definition of LARs.⁶⁹ It believes that meetings of high contracting parties to the *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to have Indiscriminate Effects* ('CCW 1980')⁷⁰ are the correct fora to achieve this.⁷¹ Kathleen Lawand, head of the arms unit at the International Committee of the Red Cross, similarly highlighted that there is no consensus on the definition of LARs.⁷² The UN Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns, regards the definition of LARs formulated by the US as a valuable starting point.⁷³ It is at this stage necessary to explain what the current state of robotic technology is and what scientists would like to achieve in the future.

⁶¹ *Unmanned Aircraft Systems*, above n 19, 1–5.

⁶² *Ibid* 1–6.

⁶³ *Ibid* 1–5.

⁶⁴ *Autonomy in Weapon Systems*, above n 54, 7.

⁶⁵ *Unmanned Aircraft Systems*, above n 19, 1–6.

⁶⁶ *Autonomy in Weapon Systems*, above n 54, 13; *Unmanned Systems Integrated Roadmap*, above n 43, 66–7.

⁶⁷ *Unmanned Aircraft Systems*, above n 19, 1–5.

⁶⁸ Kathleen Lawand, *Fully Autonomous Weapon Systems* (25 November 2013) International Committee of the Red Cross <<https://www.icrc.org/eng/resources/documents/statement/2013/09-03-autonomous-weapons.htm>>.

⁶⁹ UN GAOR, 1st Comm, 68th sess, 19th mtg, Agenda Items 89 to 107, UN Doc A/C.1/68/PV.19 (29 October 2013) 4.

⁷⁰ *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, opened for signature 10 October 1980, 1342 UNTS 137 (entered into force 2 December 1983) ('CCW 1980').

⁷¹ *Ibid*.

⁷² Lawand, above n 68.

⁷³ Heyns, above n 21, 7.

III THE STATE OF SCIENTIFIC KNOWLEDGE AND ASPIRATIONS

Militaries have been for some time using autonomous weapons in environments where civilian objects and military objects tend not to be intermixed, such as the high seas.⁷⁴ For instance, the Navy Phalanx system is located on a ship.⁷⁵ It detects the incoming fire and automatically fires back at the source of the threat.⁷⁶ Another autonomous weapon is the Wide Area Search Autonomous Attack Miniature Munition.⁷⁷ This is a miniature smart cruise missile which can loiter over an area and search for a particular target such as an armoured vehicle.⁷⁸ This munition can either autonomously engage the target or ask for permission to do so.⁷⁹ The MK 60 Encapsulated Torpedo ('CAPTOR') mine locates submarines by searching for a distinct acoustic signature and ignores friendly submarines, which have a different acoustic signature.⁸⁰ More recently, Lockheed Martin developed a long-range anti-ship missile for the US Air Force and Navy which flies itself for hundreds of miles and autonomously changes its flight-path in order to avoid being detected by radar.⁸¹

The limitation of these systems is that they are only capable of recognising military objectives which have a particular signature such as missiles and aircraft. To date, there are no technologies which enable a weapon system to distinguish between civilians, combatants and individuals who take a direct part in hostilities.⁸² To illustrate, the autonomous capability of the Lockheed Martin anti-ship missile is confined to detecting the radio waves emitted by radar⁸³ and planning its flight path in such a way as to avoid the space where there are radio waves. This is a very different capability from autonomous selection of targets in an environment where civilians and civilian objects on the one hand, and combatants, individuals who take a direct part in hostilities and military objectives on the other hand, are intermixed. Some militant groups choose not to wear a distinctive sign in order to protect themselves by blending in with the

⁷⁴ Kris Osborn, *Navy Overhauls Phalanx Ship Defense Weapon* (21 August 2013) Defense Tech <<http://defensetech.org/2013/08/21/navy-overhauls-phalanx-ship-defense-weapon>>; GlobalSecurity.org, *MK 60 Encapsulated Torpedo (CAPTOR)* (7 July 2011) <<http://www.globalsecurity.org/military/systems/munitions/mk60.htm>>.

⁷⁵ Osborn, above n 74.

⁷⁶ Ibid.

⁷⁷ Department of the Air Force, A: *Flight Test Demonstration of an Autonomous Wide Area Search Miniature Munition with Two-Way Data Link Capability* (1 April 2003) Federal Business Opportunities <https://www.fbo.gov/index?s=opportunity&mode=form&id=56f2b47c6b5d544c39ed579e4f301b94&tab=core&_cview=1>.

⁷⁸ Ibid.

⁷⁹ Ibid.

⁸⁰ Federation of American Scientists, *MK 60 Encapsulated Torpedo (CAPTOR)* (13 December 1998) <<http://fas.org/man/dod-101/sys/dumb/mk60.htm>>.

⁸¹ Markoff, above n 7.

⁸² Boothby, above n 2, 586.

⁸³ Radars send out electromagnetic waves. The objects that are in the path of these waves reflect the waves. The receiver of the radar detects the deflected waves. It displays the information about the speed and direction at which the object is moving. Additionally, the receiver displays on the screen the shape of the object and its density. Australian Government Bureau of Meteorology, *How Radar Works* (2015) <http://www.bom.gov.au/australia/radar/about/what_is_radar.shtml>.

population.⁸⁴ The militants may drive in civilian vehicles without distinct markings.⁸⁵ When this happens, troops find it difficult to distinguish between civilians and civilians who take a direct part in hostilities.⁸⁶ It is increasingly common for non-state groups not to wear a distinctive sign.⁸⁷ Michael Schmitt anticipates that this trend will accelerate as states obtain more advanced technologies.⁸⁸ Accordingly, it is necessary to examine what prospects exist for systems to be developed that will be able to distinguish civilians from civilians who take a direct part in hostilities.

Computers have been shown to be capable of logical analysis of the implications of each move such as in the case of a chess game.⁸⁹ IBM's Deep Blue computer beat the world chess champion Gary Kasparov in May 1997.⁹⁰ Although the designer of the Big Blue said that a glitch in the program helped the computer to win,⁹¹ it is evident that computer programs can play chess games to the same standard as grand masters. For example, Kasparov's match with the X3D Fritz computer in 2003 resulted in a draw.⁹² This information should, however, be understood in its context:⁹³ specifically that '[c]hess is a fairly well-defined rule-based game that is susceptible to computational analysis'.⁹⁴ Additionally, some computer programs have been successful at approximating the answers that a human would give to a set of questions.⁹⁵ On the other hand, at present computers cannot process visual data very well because computers read information pixel by pixel.⁹⁶ Since robots are equipped with sensors such as

⁸⁴ For instance, neither the members of the Taliban nor of Al-Qaeda wore a distinctive sign during Operation Enduring Freedom 2001 in Afghanistan. Jay S Bybee, 'Status of Taliban Forces under Article 4 of the *Third Geneva Convention* of 1949' (Memorandum Opinion, United States Department of Justice, 7 February 2002) 3; Jay S Bybee, 'Application of Treaties and Laws to Al Qaeda and Taliban Detainees' (Memorandum, United States Department of Justice, 22 January 2002) 10.

⁸⁵ C J Chivers and Eric Schmitt, 'In Strikes on Libya by NATO, an Unspoken Civilian Toll', *The New York Times* (online), 17 December 2011 <http://www.nytimes.com/2011/12/18/world/africa/scores-of-unintended-casualties-in-nato-war-in-libya.html?_r=0>.

⁸⁶ R Jeffrey Smith and Ann Scott Tyson, 'Shootings by US at Iraq Checkpoints Questioned', *The Washington Post* (online), 7 March 2005 <<http://www.washingtonpost.com/wp-dyn/articles/A12507-2005Mar6.html>>; Jonathan Steele, 'Iraq War Logs: Civilians Gunned Down at Checkpoints', *The Guardian* (online), 23 October 2010 <<http://www.theguardian.com/world/2010/oct/22/iraq-checkpoint-killings-american-troops>>.

⁸⁷ Richard Norton-Taylor, 'Asymmetric Warfare', *The Guardian* (online), 3 October 2001 <<http://www.theguardian.com/world/2001/oct/03/afghanistan.socialsciences>>; Craig Hatkoff and Rabbi Irwin Kula, 'A Fearful Scimitar: ISIS and Asymmetric Warfare', *Forbes* (online), 3 September 2014 <<http://www.forbes.com/sites/offwhitepapers/2014/09/02/the-asymmetric-scimitar-obamas-paradigm-pivot/>>.

⁸⁸ Michael N Schmitt, 'The Principle of Discrimination in 21st Century Warfare' (1999) 2 *Yale Human Rights and Development Journal* 143, 158–61.

⁸⁹ Marcus du Sautoy, 'Can Computers Have True Artificial Intelligence?', *BBC News* (online), 3 April 2012 <<http://www.bbc.co.uk/news/technology-17547694>>.

⁹⁰ Klint Finley, 'Did a Computer Bug Help Deep Blue Beat Kasparov?', *Wired* (online), 28 September 2012 <<http://www.wired.com/2012/09/deep-blue-computer-bug>>; Asaro, above n 21, 705.

⁹¹ Finley, above n 90, 28.

⁹² Ibid.

⁹³ Asaro, above n 21, 705.

⁹⁴ Ibid.

⁹⁵ du Sautoy, above n 89.

⁹⁶ Ibid.

cameras, sonars and infrared systems to enable them to orient in their environment, they lack adequate sensory or vision processing systems.⁹⁷ Robots are also not capable of interpreting the visual data in an abstract fashion.⁹⁸ For instance, a human can recognise words which have been warped to make them look slightly different.⁹⁹ A computer is incapable of doing so.¹⁰⁰ Furthermore, although researchers are trying to develop robots that can learn from experience and respond to novel situations, many believe that it is unclear whether at present 'it can be predicted with reasonable certainty what the robot will learn'.¹⁰¹

In a similar vein, scientists worldwide are working on an ambitious project of recreating how a human brain performs cognitive tasks.¹⁰² They would like to develop a software program that would work on this template.¹⁰³ Presently, some scientists argue that it could be possible to recreate the workings of a human brain, but that it would take from 50 to 100 years to achieve this.¹⁰⁴ They view human intelligence as a set of algorithms which are executed in the brain.¹⁰⁵ The algorithms interact with each other in order to switch the mental state from one moment to another.¹⁰⁶ Since computer programs use algorithms, these scientists believe that the function of the brain could be emulated by a computer.¹⁰⁷ Others argue, however, that highly abstract algorithms which operate on discrete symbols with fixed meanings are inadequate for capturing 'the adaptive flexibility of intelligent behaviour'.¹⁰⁸ Although Canadian scientists have built the world's largest simulation of a functioning human brain,¹⁰⁹ this model is still insufficiently complex to capture the entirety of the workings of a human brain.¹¹⁰ Additionally, unlike a human brain, their software program Spaun takes a lot of computing power to perform even the smallest of tasks.¹¹¹ To illustrate, the computer takes two hours to run to perform one second of a Spaun simulation.¹¹²

Some scientists, such as Noel Sharkey, are sceptical about efforts to create a software program that will enable a robot to comply with IHL.¹¹³ For instance, Sharkey argues that robots need to be capable of undertaking 'common sense

⁹⁷ At present robots can merely be equipped with 'sensors such as cameras, infrared sensors, sonars, lasers, temperature sensors and ladars'. Sharkey, above n 32, 788.

⁹⁸ du Sautoy, above n 89.

⁹⁹ Ibid.

¹⁰⁰ Ibid.

¹⁰¹ Marchant et al, above n 9, 284 (emphasis omitted).

¹⁰² George Dvorsky, *How Will We Build an Artificial Human Brain?* (2 May 2012) IO9 <<http://io9.com/5906945/how-will-we-build-an-artificial-human-brain>>.

¹⁰³ Ibid.

¹⁰⁴ Ibid.

¹⁰⁵ Ibid.

¹⁰⁶ Ibid.

¹⁰⁷ Ibid.

¹⁰⁸ Colin Allen, 'The Future of Moral Machines', *The New York Times* (online), 25 December 2011 <<http://opinionator.blogs.nytimes.com/2011/12/25/the-future-of-moral-machines/>>.

¹⁰⁹ Francie Diep, 'Artificial Brain: "Spaun" Software Model Mimics Abilities, Flaws of Human Brain', *The Huffington Post* (online), 30 November 2012 <http://www.huffingtonpost.com/2012/11/30/artificial-brain-spaun-softwaremodel_n_2217750.html>.

¹¹⁰ Ibid.

¹¹¹ Ibid.

¹¹² Ibid.

¹¹³ Sharkey, above n 32, 788–9.

reasoning' and 'battlefield awareness' in order to distinguish between combatants and civilians.¹¹⁴ He also thinks that developing machines which have these capabilities 'may be computationally intractable',¹¹⁵ unless there is an unforeseen breakthrough in science. Nevertheless, because these scientific breakthroughs could take place, it is crucial to analyse what conditions LARs would need to fulfil in order to comply with the rules of targeting. Before undertaking this analysis, the *opinio juris* of states on the legality of employing LARs will be surveyed.

IV STATE PRACTICE

The position of states on the employment of LARs may be subdivided into three main categories. Costa Rica¹¹⁶ and the Holy See¹¹⁷ argue that the employment of LARs should be banned. The Holy See views the removal of human control over the decision of whether or not to take someone's life as deeply problematic both from a legal and ethical point of view.¹¹⁸ Switzerland also explains that it wishes for human control to be retained over robotic systems.¹¹⁹ On the other hand, the European Union delegation¹²⁰ and countries such as Ecuador,¹²¹ Egypt,¹²² Greece,¹²³ Ireland,¹²⁴ Italy,¹²⁵ Japan,¹²⁶ Madagascar,¹²⁷ Lithuania,¹²⁸ Mexico,¹²⁹ Pakistan¹³⁰ and Ukraine¹³¹ state that a

¹¹⁴ Ibid 789.

¹¹⁵ Ibid.

¹¹⁶ Chan, above n 4, 3.

¹¹⁷ Archbishop Silvano M Tomasi, 'Statement' (Speech delivered at the Meeting of the High Contracting Parties to the *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects* (CCW), 14 November 2013) 2.

¹¹⁸ Ibid 2.

¹¹⁹ S E Urs Schmid, 'Exchange of Views' [author's trans] (Speech to the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, 13 April 2015) 2.

¹²⁰ Kos, above n 3, 4.

¹²¹ Permanent Mission of Ecuador to the United Nations, 'Statement' [author's trans] (Speech to the UN General Assembly, 1st Comm, 68th sess, 25 October 2013) 2.

¹²² UN GAOR, 1st Comm, 68th sess, 4th mtg, Agenda Items 89 to 107, UN Doc A/C.1/68/PV.4 (8 October 2013) 10.

¹²³ UN GAOR, 1st Comm, 68th sess, 19th mtg, Agenda Items 89 to 107, UN Doc A/C.1/68/PV.19 (29 October 2013) 4.

¹²⁴ Ibid 10.

¹²⁵ Vinicio Mati (Speech to the Meeting of the High Contracting Parties to the *Convention on Certain Conventional Weapons*, Geneva, 14–15 November 2013) 2.

¹²⁶ Toshio Sano, 'Statement' (Speech to the Meeting of the High Contracting Parties to the *Convention on Certain Conventional Weapons*, Geneva, 14–15 November 2013) 2.

¹²⁷ Annick H Andriamampianina, 'General Exchange of Views' [author's trans] (Speech to the Meeting of the High Contracting Parties to the *Convention on Certain Conventional Weapons*, Geneva, 14–15 November 2013) 2.

¹²⁸ The Republic of Lithuania, 'Statement' (Speech to the Meeting of the High Contracting Parties to the *Convention on Certain Conventional Weapons*, Geneva, 14–15 November 2013) 2.

¹²⁹ Head of the Delegation of Mexico, 'Statement' [author's trans] (Speech to the Meeting of the High Contracting Parties to the *Convention on Certain Conventional Weapons*, Geneva, 14–15 November 201) 2.

¹³⁰ UN GAOR, 1st Comm, 68th sess, Agenda Items 89 to 107, UN Doc A/C.1/68/PV.9 (16 October 2013) 6, 7.

treaty should be concluded which either regulates or restricts the way in which these systems may be employed. They view the negotiation of an additional protocol to *CCW 1980* as a suitable solution.¹³² Finally, countries such as the UK¹³³ and the US¹³⁴ explain that, at present, the military will have personnel controlling the robots. However, as new technologies are developed, there may well come a point when robots will make autonomous targeting decisions.¹³⁵ Given the wide ranging position of states in regard to the question of whether current norms adequately address the employment of LARs, it is necessary to analyse how the rules of targeting regulate such technologies.

V THE RULES OF TARGETING

A *The Principle of Distinction*

The principle of distinction requires the parties to a conflict to distinguish ‘at all times’ between the civilian population, individuals who take a direct part in hostilities and combatants on the one hand, and between civilian objects and military objectives on the other hand.¹³⁶ This rule may be found in art 48 of the *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)* (‘API 1977’).¹³⁷ The International Court of Justice (‘ICJ’) commented in its advisory opinion on the *Legality of the Threat or Use of Nuclear Weapons* (‘*Nuclear Weapons Case*’) that this rule is the bedrock of IHL.¹³⁸ A review of state practice shows that it is uncontested that this rule has customary international law status in international and non-international armed conflict.¹³⁹

When states originally formulated the principle of distinction, they contemplated that human beings would be making the decision of whether it was

¹³¹ Oleksandr Aleksandrovich, ‘Statement’ (Speech to the Meeting of the High Contracting Parties to the *Convention on Certain Conventional Weapons*, Geneva, 14–15 November 2013) 2.

¹³² See the state practice of the European Union delegation and states including Ecuador, Egypt, Greece, Ireland, Italy, Japan, Lithuania, Mexico, Madagascar, Pakistan and Ukraine. See above nn 120–9.

¹³³ United Kingdom, *Parliamentary Debates*, House of Lords, 7 March 2013, vol 743, col WA411 (Lord Astor of Hever); United Kingdom, *Parliamentary Debates*, House of Lords, 26 March 2013, vol 744, col 960 (Lord Astor of Hever); United Kingdom, *Parliamentary Debates*, House of Commons, 17 June 2013, vol 564, col 734 (Alistair Burt).

¹³⁴ *Autonomy in Weapon Systems*, above n 54, 2–3; *Unmanned Aircraft Systems Flight Plan*, above n 31, 41.

¹³⁵ *Unmanned Aircraft Systems*, above n 19, 1.5, 1.6; *Unmanned Systems Integrated Roadmap*, above n 43, 67.

¹³⁶ *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, opened for signature 12 December 1977, 1125 UNTS 3 (entered into force 7 December 1978) art 48 (‘API 1977’).

¹³⁷ *Ibid.*

¹³⁸ *Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion)* [1996] ICJ Rep 226, 35 [78]–[79].

¹³⁹ Jean-Marie Henckaerts and Louise Doswald-Beck (eds), *Customary International Humanitarian Law* (Cambridge University Press, 2005) vol 2, ch 1.

lawful to employ lethal force.¹⁴⁰ Although states have been using weapons that locate targets based on detecting a distinct signature, it is human beings who make the decision in what types of circumstances the employment of such weapons is lawful. Moreover, these weapons attack only objects that emit a distinct signal or signature, such as radar or acoustic waves. The following example demonstrates that such weapons may fail to discriminate between lawful and unlawful targets if human decision-makers do not make a careful assessment of whether it is lawful to use them for a particular attack. For instance, the Israeli Harpy munition detects radar signals and cannot tell whether the radar is located on a civilian object or on a military objective, such as an anti-aircraft station.¹⁴¹ The French Ambassador pointed out that LARs raise the question of whether removing the human from the decision to use lethal force has implications for compliance with the law.¹⁴² This is because LARs will autonomously select targets in environments such as cities and will, therefore, need to autonomously assess whether an attack complies with the principle of distinction.

To date, the academic discussion has mainly focused on what degree of certainty the principle of distinction requires human decision-makers to achieve.¹⁴³ Since the situation on the battlefield is unpredictable and constantly evolves,¹⁴⁴ individuals are unable to achieve complete certainty.¹⁴⁵ It is recognised that individuals may make genuine mistakes and target a civilian or a civilian object as a result. The principle of distinction prohibits individuals from launching an attack ‘when it is not reasonable to believe’, in the circumstances in which they find themselves, and on the basis of the information available to them, that the proposed target is a combatant or a military objective.¹⁴⁶ The prospect of LARs being developed has resulted in commentators analysing what qualities enable individuals to evaluate whether an attack will comply with the

¹⁴⁰ UN GAOR, 1st Comm, 68th sess, 4th mtg, Agenda Items 89 to 107, UN Doc A/C.1/68/PV.4 (8 October 2013) 3–4; Marco Sassòli, ‘Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to Be Clarified’ (2014) 90 *International Law Studies* 308, 323.

¹⁴¹ Sharkey, above n 32, 788.

¹⁴² UN GAOR, 1st Comm, 68th sess, 4th mtg, Agenda Items 89 to 107, UN Doc A/C.1/68/PV.4 (8 October 2013) 3.

¹⁴³ See generally Nils Melzer, ‘Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law — Adopted by the Assembly of the International Committee of the Red Cross on 26 February 2009’ (2008) 90 *International Review of the Red Cross* 991; Afsheen John Radsan and Richard Murphy, ‘Measure Twice, Shoot Once: Higher Care for CIA-Targeted Killing’ [2011] *University of Illinois Law Review* 1201, 1224; Geoffrey S Corn, ‘Targeting, Command Judgment, and a Proposed Quantum of Information Component: A Fourth Amendment Lesson in Contextual Reasonableness’ (2011) 77(2) *Brooklyn Law Review* 437, 485; Carla Crandall, ‘Ready ... Fire ... Aim! A Case for Applying American Due Process Principles before Engaging in Drone Strikes’ (2012) 24 *Florida Journal of International Law* 55, 87–8.

¹⁴⁴ Carl von Clausewitz, *Principles of War* (Stephen Austin and Sons, 1943) 51.

¹⁴⁵ Melzer, above n 143, 1039.

¹⁴⁶ *Prosecutor v Galić (Judgement and Opinion)* (International Criminal Tribunal for the Former Yugoslavia, Trial Chamber I, Case No IT-98-29-T, 5 December 2003) [51].

rules of targeting such as the principle of distinction.¹⁴⁷ Since human beings have traditionally been making this assessment, states have not commented in detail on what these qualities are. There simply was no need to do this. Consequently, it is necessary to untangle aspects which are implicit in the principle of distinction. These relate to: (1) qualities which enable individuals to distinguish lawful from unlawful targets; and (2) qualities which make it possible for individuals to evaluate whether an attack complies with the principle of distinction.

B The Rule of Target Verification

In order to put parties to the conflict in a position where they can distinguish between lawful and unlawful targets, customary international law requires them to 'do everything feasible to verify that the objectives to be attacked' are combatants, individuals who take a direct part in hostilities or military objectives.¹⁴⁸ This rule is found in *API 1977* art 57(2)(a)(i) and will be referred to as the rule of target verification.¹⁴⁹ The rule of target verification requires that those who plan, decide upon or execute an attack gather information in order to assist them in determining whether there are lawful targets in an area of attack.¹⁵⁰ Subsequently, they are to take steps to verify that the proposed target is in fact a lawful one.¹⁵¹

The obligation imposed by the rule of target verification is qualified by the term 'feasible'. States interpret the term 'feasible' as requiring them to take those precautions 'which [are] practicable or practically possible, taking into account all [the] circumstances ruling at the time, including humanitarian and military considerations'.¹⁵²

¹⁴⁷ Human Rights Watch and Harvard International Human Rights Clinic, *Losing Humanity: The Case against Killer Robots* (Human Rights Watch, 2012) 4; Robert Sparrow, 'Building a Better WarBot: Ethical Issues in the Design of Unmanned Systems for Military Applications' (2009) 15 *Science and Engineering Ethics* 169, 181; Jörg Wellbrink, 'Roboter Am Abzug' (Speech delivered at the Zebis Discussion Seminar, Berlin, 4 September 2013) <<http://www.zebis.eu/veranstaltungen/archiv/podiumsdiskussion-roboter-am-abzug-sind-soldaten-ersetzbar/>>; Asaro, above n 21, 699.

¹⁴⁸ Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 367.

¹⁴⁹ *API 1977* art 57(2)(a)(i).

¹⁵⁰ UK Ministry of Defence, 'Joint Service Manual of the Law of Armed Conflict' (Service Manual JSP 383, Joint Doctrine and Concepts Centre, 2004) [13.32] ('*Service Manual of the Law of Armed Conflict*').

¹⁵¹ *Ibid.*

¹⁵² Germany, Declarations Made upon Ratification of *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 1607 UNTS 526 (14 February 1991) 529 [2]; Canada, Reservations and Statements of Understanding Made upon Ratification of *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 1591 UNTS 464 (20 November 1990) 464; United Kingdom of Great Britain and Northern Ireland, Declarations Made upon Ratification of *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 2020 UNTS 75 (28 January 1998) 76 [b]. All sources are quoted in: Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 357–8.

This means that the measures parties to the conflict are able to take to gather intelligence, conduct reconnaissance and verify the character of the target depend on available resources and their quality.¹⁵³

The International Criminal Tribunal for the Former Yugoslavia ('ICTY') Committee explained that commanders 'must have some range of discretion to determine which available resources shall be used and how they shall be used'.¹⁵⁴ The law requires commanders to determine in good faith what resources and tactics it is 'feasible' to employ in the circumstances.¹⁵⁵ Their decision is judged against what a reasonable person would have done in the circumstances.¹⁵⁶

Unfortunately, states have not disclosed criteria which commanders apply in evaluating whether it is 'feasible' to adopt a particular precautionary measure. Although there clearly comes a point when it is not 'practicable or practically' possible to adopt a particular measure, such as to recruit more informants, the determination of when this point is reached is far from straightforward. Michael Walzer argues that the notion of necessity is fluid in nature, so that it is a product of subjective judgement whether necessity exists.¹⁵⁷ The commanders can always invoke necessity in order to maintain that it is, for instance, not 'feasible' to assume risk to the force.¹⁵⁸ Walzer further explains that military necessity is a term that is invoked to discuss 'probability and risk'.¹⁵⁹ When commanders talk about military necessity, they are really talking about reducing the risk of losing the battle or the risk of soldiers being killed.¹⁶⁰ Walzer concludes that in practice, 'a range of tactical and strategic options' usually exist which can improve the chances of victory.¹⁶¹

The rule of target verification should be understood in the context of rules which complement it. *API 1977* art 57(1) requires parties to the conflict to take 'constant care' in order to spare civilians and civilian objects.¹⁶² This rule has customary international law status in international and non-international armed conflicts.¹⁶³ It supplements and fleshes out the principle of distinction.¹⁶⁴ The

¹⁵³ Jean-François Quéguiner, 'Precautions under the Law Governing the Conduct of Hostilities' (2006) 88 *International Review of the Red Cross* 793, 797.

¹⁵⁴ Committee Established to Review the NATO Bombing Campaign against the Federal Republic of Yugoslavia, 'Final Report to the Prosecutor' (Report, International Criminal Tribunal for the Former Yugoslavia, 2000) [29] ('*NATO Bombing against Yugoslavia Report*').

¹⁵⁵ Yves Sandoz, Christophe Swinarski and Bruno Zimmermann (eds), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (International Committee of the Red Cross, 1987) 682 [2198].

¹⁵⁶ Canada Office of the Judge Advocate General, *Law of Armed Conflict at the Operational and Tactical Level* (National Defence, 1999) 4.3–4.4 §§25–27, quoted in Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 359.

¹⁵⁷ Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations* (Basic Books, 4th ed, 2006) 144.

¹⁵⁸ *Ibid.*

¹⁵⁹ *Ibid.*

¹⁶⁰ *Ibid.*

¹⁶¹ *Ibid.*

¹⁶² Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 336.

¹⁶³ *Ibid.* 336–7.

¹⁶⁴ Sandoz, Swinarski and Zimmermann, above n 155, 680 [2191].

upshot of this rule is that parties to the conflict should consider up to the point of carrying out the attack what verification measures it is 'feasible' to take. The following example illustrates why the rule of target verification is relevant to LARs. A robot belonging to a signatory to the *API 1977*, which encounters a hydroelectric dam, may need to determine what measures would be 'feasible' to take in order to check whether the enemy uses the dam in 'regular', 'significant' and 'direct' support of military operations.¹⁶⁵ The crucial question is whether LARs, depending on their software architecture, are capable of applying the rule of target verification. In order to address this question, it is necessary to analyse what qualities enable individuals to apply the rule to battlefield scenarios.

C The Principle of Proportionality

The principle of proportionality is designed to govern situations where civilians and civilian objects are incidentally injured as a result of an attack on a lawful target. The rule is formulated in *API 1977* art 51(5)(b) and prohibits attacks 'which may be 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'.¹⁶⁶ The principle of proportionality has customary international law status in international and non-international armed conflicts.¹⁶⁷ The rule requires commanders to balance the military value of the attack and the harm to civilians.¹⁶⁸ In applying the rule, commanders consider unlike values.¹⁶⁹ They rely on their moral judgement¹⁷⁰ when balancing military gains and harm to civilians. Since commanders rely on their judgement, the rule has an element of subjectivity.¹⁷¹ Commenting on the meaning of the term 'excessive', Michael Bothe, Waldemar Solf and Karl Partsch explain that an attack is disproportionate whenever there is an 'obvious imbalance' between the two sides of the proportionality test.¹⁷²

¹⁶⁵ *API 1977* art 56(2)(a).

¹⁶⁶ *API 1977* art 51(5)(b).

¹⁶⁷ Inter-American Commission on Human Rights, *Third Report on the Human Rights Situation in Colombia*, Doc OEA/Ser.L/V/II.102 Doc.9.rev.1 (26 February 1999) [77]; Jean-Marie Henckaerts and Louise Doswald-Beck (eds), *Customary International Humanitarian Law* (Cambridge University Press, 2005) vol 1, 46; Fausto Pocar, 'Protocol I Additional to the 1949 Geneva Conventions and Customary International Law' in Yoram Dinstein and Fania Domb (eds), *The Progression of International Law: Four Decades of the Israel Yearbook on Human Rights — An Anniversary Volume* (Brill, 2011) 197, 206.

¹⁶⁸ Christopher Greenwood, 'Customary International Law and the First Geneva Protocol of 1977 in the Gulf Conflict' in Peter Rowe (ed), *The Gulf War 1990–1991 in International and English Law* (Sweet & Maxwell, 1993) 63, 88.

¹⁶⁹ Kenneth Watkin, 'Assessing Proportionality: Moral Complexity and Legal Rules' (2005) 8 *Yearbook of International Humanitarian Law* 3, 5.

¹⁷⁰ *Ibid* 7.

¹⁷¹ Marco Sassòli and Lindsay Cameron, 'The Protection of Civilian Objects — Current State of the Law and Issues *de lege ferenda*' in Natalino Ronzitti and Gabriella Venturini (eds), *The Law of Air Warfare: Contemporary Issues — Essential Air and Space Law* (Eleven International, 2006) vol 1, 35, 63.

¹⁷² Michael Bothe, Karl Josef Partsch and Waldemar A Solf, *New Rules for Victims of Armed Conflicts: Commentary on the Two 1977 Protocols Additional to the Geneva Conventions of 1949* (Martinus Nijhoff Publishers, 1982) 310 n 30.

Significantly, the degree of military advantage which the destruction of a particular military objective offers in the circumstances varies.¹⁷³ For example, if the belligerents are in the process of negotiating a ceasefire, the destruction of command and control facilities is likely to offer a lesser degree of military advantage than when the outcome of the conflict is uncertain.¹⁷⁴ Since context drives the degree of military advantage offered by the destruction of a military objective, and since parties to the conflict encounter a myriad of different situations on the battlefield, it is impossible to predict all possible scenarios which commanders could encounter. Accordingly, it is impossible to compile a list that indicates in advance all possible permutations of the target type and context that a commander could encounter and that would provide him or her with an assessment regarding the proportionality of the attack.¹⁷⁵

W Hays Parks criticised the principle of proportionality as being a vague¹⁷⁶ and ‘dangerously undefined’ rule.¹⁷⁷ In order to assess the extent to which this is the case, it is valuable to examine the experience of commanders applying this rule. Tony Montgomery served during Operation Allied Force 1999 at the Headquarters of the US European Command and was the Deputy Staff Judge Advocate and Chief.¹⁷⁸ According to him, in making the proportionality assessment, commanders aim to make ‘reasonable’ decisions.¹⁷⁹ In his case, he followed the instructions in the military manual that the commander should make the proportionality assessment ‘on the basis of an honest and reasonable estimate of the facts available to him’.¹⁸⁰ During Operation Allied Force 1999 there was a desire among commanders to make a decision relating to the balancing of harm to civilians against the military advantage conferred by the destruction of the target in an objective manner rather than in a subjective manner.¹⁸¹ Of course, this does not mean that their decisions were scientific.¹⁸²

On the other hand, the statements of other lawyers who advise the armed forces suggest that it is impossible to remove subjectivity from the assessment of the value of the target and the value of human life. Judge Advocate General Jason Wright states that during his training he asked whether there was consensus on the notion of ‘excessive’.¹⁸³ He received varying responses¹⁸⁴ and the American military doctrine does not define this standard.¹⁸⁵ Former Judge

¹⁷³ Schmitt, ‘The Principle of Discrimination in 21st Century Warfare’, above n 88, 151.

¹⁷⁴ *Ibid.*

¹⁷⁵ Markus Wagner, ‘The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems’ (2014) 47 *Vanderbilt Journal of Transnational Law* 1371, 1398.

¹⁷⁶ W Hays Parks, ‘Air War and the Law of War’ (1990) 32 *Air Force Law Review* 1, 173.

¹⁷⁷ *Ibid.* 171.

¹⁷⁸ Tony Montgomery, ‘Legal Perspective from the EUCOM Targeting Cell’ in Andru E Wall (ed), *International Law Studies* (Naval War College, 1901–2002) vol 78, 189, 189.

¹⁷⁹ *Ibid.* 189–90.

¹⁸⁰ Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 334.

¹⁸¹ Montgomery, above n 178, 193.

¹⁸² *Ibid.*

¹⁸³ Jason D Wright, ‘“Excessive” Ambiguity: Analysing and Refining the Proportionality Standard’ (2012) 94 *International Review of the Red Cross* 819, 820.

¹⁸⁴ *Ibid.*

¹⁸⁵ *Ibid.* 830.

Advocate General Michael Schmitt writes that cultural and societal backgrounds of commanders influence the value they place on human life and on the value of the proposed target.¹⁸⁶ It may be the case that in poor countries the commanders and their societies are so desensitised to death and suffering that they tend to put lesser value on human life than the more well-off states.¹⁸⁷ Although it is difficult to establish whether, and if so to what extent, poorer countries place less weight on the life of a civilian than on military gains, there is evidence that countries place varying values on a human life. A group of military lawyers from Australia, Canada, New Zealand, the UK and the US met in the aftermath of Operation Desert Storm 1991 with a view to harmonise military manuals.¹⁸⁸ However, since these countries placed varying value on a human life, they were unable to come up with a common position on concepts such as 'proportionality'.¹⁸⁹

The statement of the ICTY Committee, which was established under the auspices of the ICTY to advise the prosecutor on whether there were grounds to open proceedings against NATO countries,¹⁹⁰ provides guidance on the application of the principle of proportionality. According to the ICTY Committee, while different commanders with 'different doctrinal backgrounds and differing degrees of combat experience or national military histories' would disagree on the application of the principle of proportionality, reasonable commanders should be able to agree on what constitutes a clearly disproportionate attack in relation to the military advantage.¹⁹¹ This rule is relevant to LARs because attacks frequently cause incidental harm to civilians and civilian objects.

D *The Principle of the Least Feasible Damage*

According to *API 1977* art 57(2)(a)(ii), those who plan or decide upon an attack shall take all 'feasible' precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimising, incidental loss of civilian life, injury to civilians and damage to civilian objects.¹⁹² This duty has customary international law status in international and non-international armed conflicts.¹⁹³ Yves Sandoz calls this rule the 'principle of the least feasible damage'.¹⁹⁴ Alexandra Boivin explains that the essence of the obligation is that even if the use of a particular weapon (means of warfare) or tactic (method of warfare) would comply with the principle of proportionality, the planner

¹⁸⁶ Schmitt, 'The Principle of Discrimination in 21st Century Warfare', above n 88, 151, 157.

¹⁸⁷ *Ibid* 151.

¹⁸⁸ Françoise J Hampson, 'Means and Methods of Warfare in the Conflict in the Gulf' in Peter Rowe (ed), *The Gulf War 1990–1991 in International and English Law* (Sweet & Maxwell, 1993) 89, 108–9.

¹⁸⁹ *Ibid* 109.

¹⁹⁰ *NATO Bombing against Yugoslavia Report*, above n 154, [5].

¹⁹¹ *Ibid* [50].

¹⁹² *API 1977* art 57(2)(a)(ii).

¹⁹³ *Prosecutor v Kupreškić (Judgement)* (International Criminal Tribunal for the Former Yugoslavia, Trial Chamber, Case No IT-95-16-T, 14 January 2000) [524]; Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume I*, above n 167, 57.

¹⁹⁴ Yves Sandoz, 'Commentary' in Andru E Wall (ed), *International Law Studies* (Naval War College, 1901–2002) vol 78, 273, 278.

nevertheless has to consider whether it is 'feasible' to employ alternative weapons or tactics in order to further 'minimize or altogether avoid casualties'.¹⁹⁵ For instance, precision-guided munitions are known to reduce incidental civilian casualties and damage to civilian objects,¹⁹⁶ as do smaller munitions with smaller fragmentation effect.¹⁹⁷

According to Italy's *LOAC Elementary Rules Manual*, the principle of the least feasible damage requires attackers to tailor weapons to the particular characteristics of the military objective.¹⁹⁸ To illustrate, during Operation Desert Storm 1991 the US wanted to attack the headquarters of the Iraqi army in Kuwait.¹⁹⁹ A gas line ran under that building.²⁰⁰ The US forces carefully chose the munition and the angle at which the bomb impacted the army headquarters to ensure that they destroyed the building without damaging the gas pipeline.²⁰¹ As a result, there were fewer civilian casualties and less damage to nearby buildings.²⁰² Additionally, in that armed conflict the US normally scheduled attacks on known dual-use facilities at night 'because fewer people would be inside or on the streets outside'.²⁰³ In terms of tactics, greater reliance on soldiers on foot and less reliance on heavy firepower, such as tanks, contributes to the reduction of civilian casualties.²⁰⁴ States interpret the term 'feasible' in an identical way for the purpose of application of the principle of the least feasible damage and the rule of target verification.²⁰⁵ Just as in the case of the rule of target verification, states have not disclosed what criteria commanders apply to determine whether it is 'feasible' to select an alternative weapon or tactic in the

¹⁹⁵ Boivin, above n 1, 38–9.

¹⁹⁶ Defence Committee, *Further Memorandum from the Ministry of Defence on Operation Telic Air Campaign (December 2003)*, House of Commons Paper No 57, Session 2003–04 (2003).

¹⁹⁷ Schmitt, 'The Principle of Discrimination in 21st Century Warfare', above n 88, 165 n 87.

¹⁹⁸ Italy, *Regole Elementari di diritto di guerra* [LOAC Elementary Rules Manual], SMD-G-012, 1991 §§ 45, 53, quoted in Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 377.

¹⁹⁹ Michael W Lewis, 'The Law of Aerial Bombardment in the 1991 Gulf War' (2003) 97 *American Journal of International Law* 481, 499.

²⁰⁰ *Ibid.*

²⁰¹ *Ibid.*

²⁰² *Ibid.* 501.

²⁰³ United States Department of Defense, 'Conduct of the Persian Gulf War: Final Report to Congress' (Report, United States Department of Defense, April 1992) 100.

²⁰⁴ Select Committee on Defence, *Minutes of Evidence: Examination of Witnesses (Questions 720–739)*, House of Commons Paper No 57, Session 2003–04 (2003) 733 <www.publications.parliament.uk/pa/cm200304/cmselect/cmdfence/57/3070203.htm>.

²⁰⁵ Germany, Declarations Made upon Ratification of *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 1607 UNTS 526 (14 February 1991) 529 [2]; Canada, Reservations and Statements of Understanding Made upon Ratification of *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 1591 UNTS 464 (20 November 1990) 464 [5]; United Kingdom of Great Britain and Northern Ireland, Declarations Made upon Ratification of *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 2020 UNTS 75 (28 January 1998) 76 [b]. Sources quoted in: Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 357–8. See also CCW 1980, as amended by *Protocol on Prohibitions of Restrictions on the Use of Mines, Booby-Traps and Other Devices*, opened for signature 3 May 1996, 2048 UNTS 93 (entered into force 3 December 1980) art 3(10) ('*Protocol II to CCW 1980*').

circumstances. LARs will likely be equipped with different weapons so that they can strike different kinds of targets. Since the adversary will try to disable them, they will need to devise tactics for self-protection which reflect the duty to avoid, or at least minimise, injury to civilians. As a result, LARs may need to apply the principle of the least feasible damage.

VI CRITERIA FOR REGULATING LETHAL AUTONOMOUS ROBOTS

The rules of targeting are addressed to human decision-makers,²⁰⁶ and LARs would need to make decisions to the same standard as a human being.²⁰⁷ As a result, the starting point for the analysis should be an evaluation of what components human decision-making comprise. Commentators have proposed various criteria regarding what elements human decision-making entails. These will now be critically examined in turn and additional criteria will be suggested.

A *The Role of 'Situational Awareness' and Judgement*

Many militaries issue instructions, known as the rules of engagement ('RoE'),²⁰⁸ in order to assist the forces to adhere to the rules of targeting such as the principle of distinction.²⁰⁹ RoE may place additional restrictions on the use of force for policy reasons, such as on the use of particular weapon systems or on the types of targets that may be attacked.²¹⁰ For instance, a portion of the US RoE instructs the forces that they may open fire if they experience a hostile act (an attack or the use of force against them) or see hostile intent (the threat of imminent use of force).²¹¹ These instructions are designed to aid soldiers to discern when an individual is taking a direct part in hostilities. The fact that a pilot of an unmanned aerial vehicle (drone) mistook civilians who were gathering scrap metal for insurgents²¹² suggests that it is not always clear-cut to the attacker what the status of an individual is. The real issue is what qualities enable soldiers to ascertain the character of the target, including qualities based on observing cues such as hostile intent.

Gary Klein studied how decision-makers, such as experienced military personnel and firefighters, make decisions under time pressure, under conditions of uncertainty and in circumstances when the situation is quickly evolving.²¹³ He found that they used their experience of prior situations, without consciously realising it, to see whether the present scenario could be matched to a previously

²⁰⁶ Sassòli, above n 140, 336.

²⁰⁷ Ibid 312.

²⁰⁸ Alan Cole et al, *Rules of Engagement Handbook* (International Institute of Humanitarian Law, 2009) v.

²⁰⁹ Sean Condrón (ed), *Operational Law Handbook* (The Judge Advocate General's Legal Center & School, US Army, 2011) 11.

²¹⁰ Ibid.

²¹¹ Ibid 75.

²¹² Jane Mayer, 'The Predator War: What Are the Risks of the CIA's Covert Drone Program?', *The New Yorker* (online), 26 October 2009 <<http://www.newyorker.com/magazine/2009/10/26/the-predator-war>>.

²¹³ Gary A Klein, *Sources of Power: How People Make Decisions* (Massachusetts Institute of Technology Press, 1999) 99.

encountered scenario.²¹⁴ They then waited to observe how the situation unfolded in order to check the validity of their inference.²¹⁵ If they had misread the situation, they would see discrepancies between the expected course of action and the reality.²¹⁶ Whenever this occurred, they proceeded to reformulate their hypothesis.²¹⁷ In general, the more effort it took to explain conflicting evidence, the less confident individuals felt in their assessment of the unfamiliar situation.²¹⁸ Klein's conclusions are corroborated by other studies.²¹⁹

Although commentators have not engaged with Klein's theory, their writings fit into his analytical framework. Chantal Grut argues that in order to comply with the principle of distinction, the robot that selects and engages targets should be capable of conducting contextual analysis of the situation on the ground and of applying judgement to interpret the events.²²⁰ Grut's argument is echoed by Colonel Darren Stewart. He maintains that 'the processing of information into intelligence requires a broad array of skills, including intuitive, experience-based analysis and cognitive functions'.²²¹ Since Stewart suggests that soldiers use intuition and experience about past military operations to interpret the situation in front of them,²²² he implies that an understanding of context behind the situation is important in order to accurately interpret the situation on the battlefield. Stewart's reference to intuition and past experience²²³ echoes Klein's theory that individuals rely on their past experiences to construct an understanding about what is happening.²²⁴ In turn, Klein's theory suggests that individuals rely on past experiences to construct their understanding of context behind the events.²²⁵

Schmitt suggests that the exercise of human judgement is not a key component that enables forces to correctly identify military objectives and comply with the principle of distinction.²²⁶ According to him, 'human judgment can prove less reliable than technical indicators in the heat of battle'.²²⁷ He gives the following example to support his proposition.²²⁸ In 1988 the crew of the US ship Vincennes mistakenly engaged an Iranian civil aircraft, believing that it was

²¹⁴ Ibid 33.

²¹⁵ Ibid 35.

²¹⁶ Ibid 92.

²¹⁷ Ibid.

²¹⁸ Ibid.

²¹⁹ Ibid 100.

²²⁰ Chantal Grut, 'The Challenge of Autonomous Lethal Robotics to International Humanitarian Law' (2013) 18 *Journal of Conflict and Security Law* 5, 11.

²²¹ Darren M Stewart, 'New Technology and the Law of Armed Conflict' in Raul A Pedrozo and Daria P Wollschlaeger (eds), *International Law Studies* (Naval War College, 1901–2011) vol 87, 271, 275.

²²² Ibid.

²²³ Ibid.

²²⁴ Klein, above n 213, 89–92.

²²⁵ Ibid.

²²⁶ Michael N Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (2013) *Harvard National Security Journal Features* 1, 12–13 <<http://harvardnsj.org/wp-content/uploads/2013/02/Schmitt-Autonomous-Weapon-Systems-and-IHL-Final.pdf>>.

²²⁷ Ibid.

²²⁸ Ibid.

descending in an attack profile.²²⁹ A government investigation into the incident revealed that the computers on board the ship correctly indicated that the aircraft was ascending, and, thus, did not pose a threat.²³⁰ However, because members of the crew were under stress, they interpreted the facts to fit their preconceived beliefs that the aircraft posed a threat.²³¹

A response to Schmitt²³² would be that the piece of information regarding whether the aircraft was ascending or descending was not of itself dispositive of whether the aircraft was a lawful target. Therefore, this incident does not demonstrate that the exercise of judgement is redundant to the ability to accurately identify the status of the object. In particular, the government report discusses the fact that in order to determine whether the aircraft was a lawful target, the crew had to integrate different types of information.²³³ The report states:

[W]hether the aircraft was ascending or descending could, when taken in the overall context, be a “significant indicator” ... On the other hand, the report that the altitude was decreasing could possibly have further confirmed a developing decision to fire. The Commanding Officer testified that it was only one piece of information among many. In this reviewing officer’s opinion, it is unlikely that this one piece of information would have settled the issue one way or another given the uncertainties that remained and the extremely short time left.²³⁴

The references in the report to the need to consider multiple pieces of information and to evaluate the flight profile in the ‘overall context’²³⁵ suggest that an understanding of context behind the situation was one of the factors which would have enabled the decision-makers to assess whether the target was a military objective. Since the crew had to consider various pieces of information and to evaluate the flight profile in light of this information, the excerpt of the report points to the fact that the crew had to apply judgement to interpret the information in front of them. And, if this proposition is accepted, then it follows that the application of judgement to understand the context behind the situation is a factor in enabling accurate identification of lawful targets. Since the computer on board the ship only had to determine variables such as aircraft trajectory,²³⁶ and since this information was not dispositive about whether the aircraft was a military objective,²³⁷ the Vincennes incident does not demonstrate that computer programs are superior to individuals when it comes to accurate identification of the character of objects.

Of course, the employment of judgement to place events in their context may in some cases contribute to individuals incorrectly assessing the situation. In this sense the Vincennes incident is not unique. As Klein points out, our past

²²⁹ Ibid; US Department of Defense, ‘Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988’ (Investigation Report, US Department of Defense, 19 August 1988) 37, 42–5 (*Investigation into Iran Air Flight 655 Report*).

²³⁰ *Investigation into Iran Air Flight 655 Report*, above n 229.

²³¹ Ibid.

²³² Schmitt, ‘Autonomous Weapon Systems and International Humanitarian Law’, above n 226.

²³³ *Investigation into Iran Air Flight 655 Report*, above n 229, 5.

²³⁴ Ibid 5–6.

²³⁵ Ibid.

²³⁶ Ibid 5.

²³⁷ Ibid 5–6.

experiences may mislead us.²³⁸ To illustrate, an Iraqi aircraft fired two missiles at the American ship USS Stark on 17 May 1987.²³⁹ The government enquiry found that although the commander had ‘all contextual information’ relevant to the potential threat,²⁴⁰ he did not direct the radar onto the aircraft to warn the pilot that he considered acting in self-defence.²⁴¹ Neither did he have weapons in combat-ready mode²⁴² or use countermeasures.²⁴³ Apparently, the commander did not use the radar because he did not want the Iraqi pilot to interpret this action as a hostile act.²⁴⁴ Bradd Hayes comments that because at that time Iraq and the US had friendly relations, the commander most likely assumed that he should be careful about making resort to force.²⁴⁵ He probably did not want to provoke a public outcry that the US attacked a friendly aircraft.²⁴⁶

Although reliance on past experience may result in misidentification of objects, militaries may take steps to mitigate this. The pitfall of making incorrect inferences based on past experiences may be corrected by exposing the troops to many scenarios and by simulating different developments for each of the scenarios.²⁴⁷ Equally, the chances of interpreting the events in light of a preconceived scenario due to stress may be reduced by exposing the forces to many scenarios.²⁴⁸ There is a direct link between the uniqueness of data, the time span available for the decision and the manifestation of ‘scenario fulfilment’ due to experiencing stress.²⁴⁹ Crucially, the application of judgement is central to enabling individuals to learn from being exposed to different scenarios. Unless decision-makers understand context behind the events, they may find it difficult to determine how and why a particular scenario differs from another one. Moreover, the application of judgement allows individuals to recognise discrepancies between past experiences and the situation at hand and to hypothesise what is happening in the present.²⁵⁰ The use of judgement, therefore, may contribute to the reduction of error when it comes to identifying the character of objects and persons.

It is possible to give a more refined definition of the ability to use judgement to integrate different pieces of information in order to interpret the context behind the observed events. The concept of ‘situational awareness’ is a helpful

²³⁸ Klein, above n 213, 34.

²³⁹ Grant Sharp, ‘Formal Investigation into the Circumstances Surrounding the Attack on the USS Stark (FFG 31) on 17 May 1987’ (Report, US Department of Defense, 12 June 1987) 1.

²⁴⁰ Ibid 31.

²⁴¹ Ibid.

²⁴² Ibid 33.

²⁴³ Ibid 34.

²⁴⁴ Ibid 33.

²⁴⁵ Bradd C Hayes, ‘Naval Rules of Engagement: Management Tools for Crisis’ (Report No N-2963-CC, RAND Corporation and the RAND/UCLA Center for the Study of Soviet International Behavior, July 1989) 41.

²⁴⁶ Ibid.

²⁴⁷ Klein, above n 213, 42–3.

²⁴⁸ Will Rogers, Sharon Rogers and Gene Gregston, *Storm Center: The USS Vincennes and Iran Air Flight 655 — A Personal Account of Tragedy and Terrorism* (US Naval Institute Press, 1992) 161, quoted in Nancy C Roberts, *Reconstructing Combat Decisions: Reflections on the Shootdown of Flight 655* (Naval Postgraduate School, 1992) 7.

²⁴⁹ Ibid.

²⁵⁰ Klein, above n 213, 89.

notion for unpacking these two qualities. Individuals have ‘situational awareness’ when they: (1) perceive all elements in their physical environment; (2) comprehend the dynamics between physical elements and people (through, for instance, identifying where individuals move and what their purpose is); and (3) project how the situation will unfold in the future.²⁵¹ The second element of ‘situational awareness’, namely an understanding of the relationship between individuals and their environment, presupposes that the decision-maker uses judgement to integrate different pieces of information in order to gauge what is happening. Similarly, this element of ‘situational awareness’ assumes that the decision-maker situates the unfolding event in its context. This is because an understanding of context makes it possible for a decision-maker to establish how individuals and physical objects relate to each other. On the application of Klein’s studies,²⁵² it would appear that individuals use past experience to make inferences about the relationship between individuals and objects, and hence about context. The second element of ‘situational awareness’ reflects Grut’s criterion for an ability to distinguish lawful targets from protected individuals and objects.²⁵³ The third element of ‘situational awareness’, more specifically the ability to project how the event will unfold in the near future, captures Stewart’s observation that the processing of information into intelligence requires the application of intuition and reliance on past experience of battles.²⁵⁴

The following example illustrates that experience-based analysis and judgement are linked to ‘situational awareness’ and that decision-makers who are unable to attain ‘situational awareness’ will find it difficult to identify lawful targets. On 29 December 2008 an Israeli drone pilot saw cylindrical objects in an open-backed truck.²⁵⁵ The pilot had to use his or her experience and knowledge

²⁵¹ Mica R Endsley, ‘Theoretical Underpinnings of Situation Awareness: A Critical Review’ in Mica R Endsley and Daniel J Garland (eds), *Situation Awareness Analysis and Measurement* (CRC, 1995) 3, 3–4. According to the North Atlantic Treaty Organization, this is a popular definition of this term. See Task Group TR-HFM-121, *Virtual Environments for Intuitive Human-System Interaction: Human Factors Considerations in the Design, Use, and Evaluation of AMVE-Technology* (The Research and Technology Organisation (RTO) of NATO, 2007) 6.1.

²⁵² Klein, above n 213, 91–2.

²⁵³ Grut, above n 220, 11.

²⁵⁴ Stewart, above n 221, 275.

²⁵⁵ Human Rights Watch observes that, in this particular case, the pilot authorised the strike even though the canisters were oxygen tanks and were over one metre shorter than a Grad rocket. An Israeli Defence Forces spokesman made the following comment after the attack:

The truck was targeted after the accumulation of information which indicated convincingly that it was carrying rockets between a known Hamas rocket manufacturing facility to a known rocket launching site. The attack was carried out near a known Hamas rocket manufacturing site and after a launch. It was only later discovered that the truck was carrying oxygen tanks (similar in appearance to Grad Missiles) and not rockets. The strike killed four Hamas operatives and four uninvolved civilians. It is important to note that the oxygen tanks being carried in the truck were likely to be used by Hamas for rocket manufacturing.

Human Rights Watch, *Precisely Wrong: Gaza Civilians Killed by Israeli Drone-Launched Missiles* (Human Rights Watch, 2009) 20, quoting Israel Defense Forces, ‘Conclusions of Investigations into Central Claims and Issues in Operation Cast Lead — Part 2’ (Israeli Government Communique, Israel Ministry of Foreign Affairs, 22 April 2009) <http://mfa.gov.il/MFA/ForeignPolicy/Terrorism/Pages/Conclusion_of_%20Investigations_into_Central_Claims_and_Issues_in_Operation_Cast_Lead-Part2_22-Apr-200.aspx> (‘*Conclusions of Investigations*’).

of how oxygen tanks and Grad rockets differ in order to assess whether the dimensions of the objects indicated that these were Grad rockets. The observation of the cylindrical objects and the truck corresponds to the first element of the 'situational awareness' analysis. Meanwhile, making reference to past experiences in order to determine whether the cylindrical objects are Grad rockets encompasses the second element of the 'situational awareness' analysis. Furthermore, assuming that the objects were oxygen tanks, a pilot would want to know whether these were being transported to be re-manufactured as Grad rockets or were destined for civilian use.²⁵⁶ The process of evaluating how the driver plans to use his cargo touches on the third element of 'situational awareness'. The pilot is unlikely to be able to perform this third stage of assessment without the possession of intelligence and application of judgment. This case study illustrates that the process of correctly identifying targets requires the possession of 'situational awareness'.

A counterargument would be that a machine could employ sensors to scan the composition and dimensions of the oxygen tanks. For instance, the employment of the electromagnetic spectrum may be used to detect weapons and explosives at a distance, by measuring the frequency of radiation emitted by the object.²⁵⁷ In the future, THz radiation, which lies between microwaves and visible light, may be utilised to classify the chemical composition of substances.²⁵⁸ The computer software could then process the data and classify the object.²⁵⁹ A robot equipped with such technology could search through its database to determine whether either the dimensions or composition of the object, or both, matched a profile of a military objective. It is true that a machine could correctly classify oxygen tanks as a civilian object. The challenge is that because militaries want to degrade the adversary's capability, they would want to know whether the enemy transported the oxygen tanks to be remanufactured as rockets. The evaluation that the oxygen tanks are in a truck, that individuals normally use trucks to transport objects, and that it is possible to convert oxygen tanks into rockets, entails an understanding of context behind events. In this case the context is that the truck could be taking the oxygen tanks to a hospital, to a warehouse or to a manufacturing facility. In turn, the processing of this information entails the use of judgement. Therefore, contrary to Bruce Clough,²⁶⁰ it is not always sufficient to integrate information from different sensors in order to locate military objectives.

To give another example, imagine that a Pakistani villager climbs a water tower. As is customary, he carries a gun with him at all times. In order to determine whether the individual is taking a direct part in hostilities, an attacker would want to know the purpose behind the villager's actions. The villager could have climbed the tower to gauge whether the adversary's forces were approaching the village and whether the villagers should, therefore, evacuate. Or

²⁵⁶ *Conclusions of Investigations*, above n 255.

²⁵⁷ See, eg, Office of Science and Technology, 'Exploiting the Electromagnetic Spectrum: Findings and Analysis' (Report, Government Office for Science, 20 April 2004) 29.

²⁵⁸ *Ibid* 30.

²⁵⁹ *Ibid* 32.

²⁶⁰ Bruce T Clough, 'Metrics, Schmetrics! How the Heck Do You Determine a UAV's Autonomy Anyway?' (Report, Air Force Research Laboratory Wright-Patterson Air Force Base, August 2002) 5.

he could be trying to locate the adversary in preparation of a military operation. The knowledge of these two possible relationships between the villager and the water tower constitute an understanding of the relevant context. It is necessary to exercise judgement in order to determine two potential purposes behind the individual's actions and to project how the situation is likely to unfold. Judgement is also required in order to assess whether it may be necessary to delay fire in order to observe how the situation unfolds, due to there being two conflicting explanations regarding the purpose behind the villager's actions.

Additionally, *API 1977* art 54(2), which has customary international law status in international and non-international armed conflicts,²⁶¹ is relevant to this scenario. The rule prohibits attacks on drinking water installations, unless only armed forces use these as sustenance or in direct support of military action.²⁶² Just because there is a man with a gun on the water tower does not mean that the water tower is being used as an observation post. The decision-maker would need to construct the context behind the man's actions and to apply judgement in order to determine how the situation is likely to unfold, so as to determine whether the water tower is in fact being used as an observation post. This scenario demonstrates that even when sensors correctly classify the structure as a water tower and detect the presence of a gun, this information is insufficient for concluding that the man and the water tower are lawful targets. This is because *API 1977* art 52(3) requires that, in case of doubt as to whether an object that is normally dedicated for civilian purposes is being used to make an effective contribution to military action, the attacker should presume that it is not being so used. This rule has customary international law status in international and non-international armed conflicts.²⁶³ More broadly, without possessing 'situational awareness', it is difficult to determine (as is required by *API 1977* art 54) whether objects which are indispensable to the survival of the civilian population, such as foodstuffs, agricultural areas for the production of foodstuffs, crops and livestock are solely used to sustain the armed forces. Such a determination requires the decision-maker to establish who is growing the food and how the food will be distributed in the future. In turn, it is likely to be necessary to gather intelligence and to know how the distribution network functions in order to establish this. Additionally, this discussion puts in doubt Clough's assertion that 'situational awareness' may be achieved by fusing multiple sensors and that an understanding of what is happening is not necessary to achieving such an awareness.²⁶⁴

So far it has been demonstrated that for objects that do not have a unique appearance, composition or signature, the employment of experienced-based analysis, judgement and an understanding of context are prerequisites for an ability to accurately classify the character of targets. In turn, these elements lie at the core of possession of 'situational awareness'. Since the principle of distinction presupposes that the decision-maker is capable of accurately

²⁶¹ Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 1148–9.

²⁶² *API 1977* art 54(2).

²⁶³ Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 241.

²⁶⁴ Clough, above n 260, 5.

identifying the character of persons and objects, it is surely implicit in the rule that the decision-maker be capable of achieving 'situational awareness'. If this argument is accepted, then it becomes possible to make the following proposition regarding the minimum technical requirements for an LAR. An LAR will have to employ its sensors, quickly process information and accurately classify the nature of structures, objects and persons that are present on the battlefield. It will need to be capable of integrating the pieces of data in an abstract way, and to apply judgement to interpret the situation on the battlefield, in order to understand the relationship between individuals, structures and objects on the battlefield. In order to ensure that an LAR is able to predict how the situation is likely to unfold in the future, the LAR should have software that captures principles on the basis of which human beings use past experiences to create and test hypothesis about novel situations.

Turning to the principle of proportionality, the concept of 'situational awareness' usefully captures the process that commanders use to estimate military advantage offered by the attack and harm to civilians. In order to apply the rule, commanders first note lawful targets, civilians and civilian objects in the area. They then analyse the relationship between persons and objects, as well as the context behind the situation, in order to determine what degree of military advantage the destruction of a military objective offers in the circumstances. For instance, the destruction of a bridge offers greater military advantage if the adversary lacks helicopters to transport the troops over a river. Commanders are likely to assess how the situation will unfold as part of estimating military gains. For instance, they may want to know during what timeframe a bridge constitutes a military objective. They will apply a similar analysis to estimate the extent of harm to civilians which is likely to occur. Commanders use contextual understanding of the situation as well as predictions about how the situation will unfold to estimate how many civilians and civilian objects the attack is likely to harm. To illustrate, commanders may gather intelligence about how frequently civilians use the bridge. Subsequently, commanders use the contextual information that fewer civilian vehicles cross the bridge at night time and apply judgement to predict how many civilian casualties a night-time attack is likely to inflict.

The commanders utilise judgement to determine whether harm to civilians is 'excessive'.²⁶⁵ Since the degree of military advantage offered by the attack fluctuates depending on the circumstances, and since the principle of proportionality entails the weighing of incommensurable values, it is not possible to attach a fixed value to a particular military objective or human life.²⁶⁶ Kenneth Anderson and Matthew Waxman criticise this assertion because they believe that states could deliberate and reach a common understanding on what value to place on a human life and on military gain.²⁶⁷ This assertion is overoptimistic. It is

²⁶⁵ *NATO Bombing against Yugoslavia Report*, above n 154, [48]–[50]; Sassòli, above n 140, 331.

²⁶⁶ Wagner, 'The Dehumanization of International Humanitarian Law', above n 175, 1398.

²⁶⁷ Kenneth Anderson and Matthew Waxman, 'Law and Ethics for Robot Soldiers' (2012) 176 *Policy Review* 35, 46–7.

difficult to see how 194 nations²⁶⁸ could formulate a common position on this sensitive and morally value-laden issue. Even if states succeeded in doing so, it is unclear how they could come up with an exhaustive list of rules to guide the determination of how much military advantage the destruction of a particular military target offers in various circumstances. Furthermore, since the application of the principle of proportionality entails a context-based analysis of incommensurable values,²⁶⁹ drawing up a mathematical formula or expressing the relationship between humanitarian loss and military gains as a ratio is not an equivalent to the exercise of judgement.

Attention will now turn to the role of judgement for the application of the rule of target verification and the principle of the least feasible damage. Peter Asaro views the rules of targeting as being more than just rules (such as chess rules).²⁷⁰ Unlike ordinary rules, the rules of targeting envisage decision-makers using 'interpretative judgement' in order to appropriately apply them to any given situation.²⁷¹ Unlike the chessboard, the battlefield is characterised by 'uncertainty' and soldiers have information of imperfect quality.²⁷² The nature of the battlefield means that there are alternative competing and conflicting interpretations of the situation on the ground.²⁷³ For this reason, soldiers and commanders use interpretation and judgement to determine what a legal rule requires in a particular case.²⁷⁴ Asaro's argument is echoed by Dale Stephens. Stephens comments that although the rules of targeting impose a concrete obligation, they implicitly require commanders to rely on 'fluid evaluative standards' in order to apply the rule.²⁷⁵ In order to achieve humanitarian aims of IHL, commanders apply social values and reflect on the purpose of international humanitarian law to reduce suffering when applying a rule of targeting to a factual scenario.²⁷⁶ On this approach, whenever it is debatable as to whether it is 'feasible' in the circumstances to adopt a particular precautionary measure, commanders refer to the purpose of the rule and to social values in order to decide in favour of a particular interpretation.²⁷⁷ Accordingly, LARs would need to exercise judgement and to make contextual assessments in order to apply these two rules.

B *The Role of Ethics*

Patrick Lin, George Bekey and Keith Abney point out that in addition to correctly interpreting the situation before it, a robot will need to be able to detect

²⁶⁸ Worldatlas, *How Many Countries Are In the World?* (2015) <<http://www.worldatlas.com/nations.htm>>.

²⁶⁹ Wagner, 'The Dehumanization of International Humanitarian Law', above n 175, 1398.

²⁷⁰ Asaro, above n 21, 699.

²⁷¹ *Ibid.*

²⁷² *Ibid.*

²⁷³ *Ibid.*

²⁷⁴ *Ibid.*

²⁷⁵ Dale Stephens, 'Counterinsurgency and Stability Operations: A New Approach to Legal Interpretation' in Raul A Pedrozo (ed), *International Law Studies* (Naval War College, 1901–2010) vol 86, 289, 298.

²⁷⁶ *Ibid.* 297–8.

²⁷⁷ *Ibid.* 298–9.

situations which raise ethical concerns.²⁷⁸ This involves the ability to adhere to IHL and RoE.²⁷⁹ The US commissioned Ronald Arkin to design a robot which is capable of complying with IHL.²⁸⁰ Arkin explained that he is looking to program a robot to enable it to adhere to ethical standards outlined by humans, as these standards are embodied in IHL rules.²⁸¹ Ethics are at the heart of IHL rules because these norms set down minimum ‘ethical baselines for a universal modern civilisation’.²⁸² Poland expressed the view that some rules governing the conduct of hostilities are *jus cogens*, meaning that they are peremptory norms which reflect ‘the conscience and will of [s]tates’.²⁸³ The ICJ in the *Nuclear Weapons Case* said that IHL is designed to protect the human person and embodies respect for ‘elementary considerations of humanity’.²⁸⁴ More recently, the ICTY in *Prosecutor v Delalić* noted that the object and purpose of the *Geneva Conventions* is ‘to guarantee the protection of certain fundamental values common to mankind in times of armed conflict’.²⁸⁵ According to Lin, it is important not to collapse ethics and law.²⁸⁶ Ethics stipulate principles that ought to guide actions²⁸⁷ whilst legal rules prescribe conduct and are enforceable.²⁸⁸ A course of action may be ethically desirable and yet not required legally.²⁸⁹ An example scenario is that of abstaining from targeting a combatant who is helping an elderly or a pregnant woman to evacuate.

The recognition of situations which raise ethical concerns is conducive to distinguishing civilians from civilians who take a direct part in hostilities. Consider a situation, given by Markus Wagner, where soldiers observe children as the children run after a ball towards a gate.²⁹⁰ They have prior intelligence that there are insurgents in the building, but unbeknownst to them this intelligence is inaccurate.²⁹¹ A man exits the building and carries a dagger,

²⁷⁸ Patrick Lin, George Bekey and Keith Abney, ‘Autonomous Military Robotics: Risk, Ethics and Design’ (Report, California Polytechnic State University, 20 December 2008) 37.

²⁷⁹ *Ibid* 42.

²⁸⁰ The contract number is W911NF-06-1-0252. The contract was issued by the US Army Research Office. Ronald C Arkin and Patrick Ulam, ‘An Ethical Adaptor: Behavioral Modification Derived from Moral Emotions’ (Technical Report No GIT-GVU-09-04, Georgia Institute of Technology, 2009) 1.

²⁸¹ Sofia Karlsson, Interview with Ronald Arkin (Online Interview, 25 April 2011) <<http://web.archive.org/web/20150428042624/http://owni.eu/2011/04/25/ethical-machines-in-war-an-interview-with-ronald-arkin>>.

²⁸² David Kennedy, ‘Modern War and Modern Law’ (Speech delivered at the Watson Institute for International and Public Affairs, Brown University, 12 October 2006) 9 <<http://www.law.harvard.edu/faculty/dkennedy/speeches/BrownWarSpeech.pdf>>.

²⁸³ United Nations Conference on the Law of Treaties, *Summary Records of the Plenary Meetings and of the Meetings of the Committee of the Whole*, 1st sess, 53rd mtg, UN Doc A/CONF 39/11 (6 May 1968) 302 [33].

²⁸⁴ *Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion)* [1996] ICJ Rep 226, 257 [79], citing *Corfu Channel (United Kingdom v Albania) (Merits)* [1949] ICJ Rep 22.

²⁸⁵ *Prosecutor v Delalić (Judgement)* (International Criminal Tribunal for the Former Yugoslavia, Appeals Chamber, Case No IT-96-21-A, 20 February 2001) [113].

²⁸⁶ Patrick Lin, ‘Pain Rays and Robot Swarms: The Radical New War Games the DOD Plays’, *The Atlantic* (online), 15 April 2013 <<http://www.theatlantic.com/technology/archive/2013/04/pain-rays-and-robot-swarms-the-radical-new-war-games-the-dod-plays/274965>>.

²⁸⁷ James Fieser, *Ethics*, Internet Encyclopedia of Philosophy <<http://www.iep.utm.edu/ethics>>.

²⁸⁸ Lin, above n 286.

²⁸⁹ *Ibid*.

²⁹⁰ Wagner, ‘The Dehumanization of International Humanitarian Law’, above n 175, 1392–3.

²⁹¹ *Ibid*.

which he does in order to fulfil his religious beliefs.²⁹² He shouts to the children to stay away from the gate because he believes them to be in danger.²⁹³ An understanding that most parents view it as their ethical duty to look after their children²⁹⁴ helps soldiers to interpret this situation correctly, by putting the act of shouting in the context of children who are approaching soldiers.

An ability to detect situations which raise ethical concerns additionally enables forces to appropriately apply the principle of proportionality. Whenever the adversary uses civilians as a human shield, an attacker is obligated to apply the principle of proportionality.²⁹⁵ When it comes to voluntary human shields, some commentators interpret the law as not requiring decision-makers to take the lives of these individuals into account when applying the rule.²⁹⁶ Others maintain that commanders should take their lives into account, but should attach either the same or lesser weight to their lives than they do to the lives of civilians.²⁹⁷ Whichever interpretation is adopted, an understanding that failure to incorporate human shields appropriately into the proportionality assessment raises ethical concerns, which leads commanders to attach appropriate value to the lives of civilians who are human shields. Since the value of the life of a civilian and the value of military gains cannot be easily assessed,²⁹⁸ commanders refer to social values²⁹⁹ and morality³⁰⁰ to determine what value to attach to the relevant limbs of the proportionality equation. More broadly, the understanding of social norms and ethical principles, as well as the rationale behind these two phenomena, enable commanders to balance these two values. Although the principle of proportionality is formulated as a black letter rule, the wording of the rule does not on its own give commanders criteria for assessing when the threshold of 'excessive' is reached.

Turning to the rule of target verification and the principle of the least feasible damage, an understanding of the ethical principle underlying IHL that attackers should strive to reduce suffering whilst not compromising mission success³⁰¹ guides commanders in their determination of whether it is 'feasible' to adopt a particular precautionary measure. The knowledge of social values and morality, as well as knowledge of how states expect them to apply these principles, enable commanders to apply these rules to particular battlefield scenarios.³⁰²

²⁹² Ibid.

²⁹³ Ibid.

²⁹⁴ Fieser, above n 287.

²⁹⁵ Michael N Schmitt, 'The Law of Targeting' in Elizabeth Wilmshurst and Susan Breau (eds), *Perspectives on the ICRC Study on Customary International Humanitarian Law* (Cambridge University Press, 2007) 131, 160–1.

²⁹⁶ Ibid.

²⁹⁷ Ibid.

²⁹⁸ *NATO Bombing against Yugoslavia Report*, above n 154, [48].

²⁹⁹ Schmitt, 'The Principle of Discrimination in 21st Century Warfare', above n 88, 157.

³⁰⁰ Watkin, above n 169, 26–30.

³⁰¹ *Declaration Renouncing the Use in Time of War of Explosive Projectiles under 400 Grammes Weight*, opened for signature 29 November 1868 (entered into force 11 December 1868) Preamble ('*St Petersburg Declaration*').

³⁰² Stephens, above n 275, 298–9.

C *An Ability to Experience Emotions and Compassion*

Human Rights Watch ('HRW')³⁰³ and Robert Sparrow³⁰⁴ argue that an ability to experience emotions, compassion and connection to citizens of the adversary's state are central to good targeting decisions. Similarly, Lieutenant Colonel Jörg Wellbrink believes that feelings, empathy and intuition are vital for understanding the situation on the battlefield.³⁰⁵ Although the rules of targeting do not directly address the role of emotions and compassion in decision-making, it would be strange if the rules did not assume the decision-makers to have this capacity. The ICJ articulated in the *Nuclear Weapons Case* that a 'great many' IHL rules are 'fundamental to the respect of the human person' and reflect 'elementary considerations of humanity'.³⁰⁶ It would be odd if states did not expect the decision-makers, who determine what verification measures to use, whether to use lethal force, what weapons to employ and whether an attack is proportionate, to understand and to experience respect for other human beings. Emotions and empathy allow individuals to have respect for the life and dignity of other persons.

Wagner's scenario of boys playing with a ball,³⁰⁷ which has already been described, lends support to HRW's assertion³⁰⁸ that the possession of emotions and compassion enables the soldiers to distinguish civilians from individuals who take a direct part in hostilities. In this particular scenario, soldiers will be able to identify with the feelings of fear that a parent who sees children running towards a soldier experiences. This is because they can imagine what it is like to be in their position. Because the man with a dagger in this scenario poses no threat to the soldiers or their colleagues, the soldiers are in a position to abstain from firing at the adversary. Of course, the possession of emotions and compassion may be less relevant to identifying military objectives which have a distinct appearance or signature, such as a tank or another LAR. Nevertheless, since the purpose of the forces is to kill or destroy as many lawful targets as possible, an LAR which could be employed against the whole array of possible targets would need to possess these qualities.

Another value of the ability to experience emotions and compassion is that the possession of these qualities lead soldiers to assume a risk of injury when faced with a decision of whether to open fire, in circumstances when they are not sure whether the person is taking a direct part in hostilities or not. The soldiers abstain from firing in such situations because they perceive the death of an individual who does not pose a threat to them as a tragic loss. This sensitivity stems from the ability of the soldiers to identify with the feeling of bereavement which the family and friends of the deceased experience. A counterargument would be that the diaries of soldiers reveal that whilst some soldiers felt pain at the time they

³⁰³ Human Rights Watch and Harvard International Human Rights Clinic, *The Case against Killer Robots*, above n 147, 4.

³⁰⁴ Sparrow, above n 147, 180–1.

³⁰⁵ Wellbrink, above n 147.

³⁰⁶ *Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion)* [1996] ICJ Rep 226, 35 [79], citing *Corfu Channel (United Kingdom v Albania) (Merits)* [1949] ICJ Rep 4, 22.

³⁰⁷ Wagner, 'The Dehumanization of International Humanitarian Law', above n 175, 1932–3.

³⁰⁸ Human Rights Watch and Harvard International Human Rights Clinic, *The Case against Killer Robots*, above n 147, 4.

killed the enemy,³⁰⁹ others felt pleasure and the satisfaction of winning.³¹⁰ A response to this observation would be that those soldiers who viewed fighting as a thrill took the horrors of war more seriously than civilians who were located far away from the combat zone.³¹¹ These soldiers experienced trauma after combat was over.³¹² When soldiers captured prisoners of war, they felt sympathy for those individuals.³¹³ Moreover, in 2011 the psychologists of the US Air Force completed a mental health survey of 600 drone pilots.³¹⁴ They found that 20 per cent of the pilots reported emotional exhaustion or burnout due to seeing death, despite not being physically present on the battlefield.³¹⁵ This information suggests that human beings feel compassion for each other, but that this feeling may become displaced when an individual encounters someone who poses mortal danger to him or her. Just because soldiers experience mixed emotions when killing another soldier³¹⁶ does not mean that they do not feel compassion for civilians.

The feeling of emotions and compassion enable commanders to determine what value to attach to a human life. As already explained, the valuation of a human life is intertwined with morality. According to Craig Johnson, the ability to recognise ethical concerns presupposes the possession of emotions and compassion.³¹⁷ The possession of emotions and compassion are central to the application of the rule of target verification and the principle of the least feasible damage. Stephens comments that the application of these rules entails reliance on fluid evaluative standards such as social values and on giving effect to the goal of IHL to reduce suffering.³¹⁸ It is difficult to see how a decision-maker who lacked emotion and compassion could give appropriate weight to the need to reduce suffering, when balancing humanitarian values against military considerations.

Interestingly, Arkin, who is a roboticist tasked with programming a robot that adheres to IHL, acknowledges that there is a link between ethical values which underlie these rules and emotions.³¹⁹ He states that 'in order for an autonomous agent to be truly ethical, emotions may be required at some level'.³²⁰ Moreover, Arkin comments that '[c]ompassion is already, to a significant degree, legislated into the LOW [Laws of War], and the ethical autonomous agent architecture is required to act in such a manner'.³²¹

³⁰⁹ Joanna Bourke, *An Intimate History of Killing* (Basic Books, 1st ed, 2000) xxiii.

³¹⁰ *Ibid* 1–2.

³¹¹ *Ibid* 359.

³¹² *Ibid* 362.

³¹³ *Ibid* 153–4.

³¹⁴ Matthew Power, 'Confessions of a Drone Warrior', *GQ Magazine* (online), 23 October 2013 <<http://www.gq.com/news-politics/big-issues/201311/drone-uav-pilot-assassination>>.

³¹⁵ *Ibid*.

³¹⁶ Bourke, above n 309, 362.

³¹⁷ Craig E Johnson, *Meeting the Ethical Challenges of Leadership: Casting Light or Shadow* (SAGE Publications, 2001) 234.

³¹⁸ Stephens, above n 275, 298.

³¹⁹ Arkin, 'Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture', above n 47, 73.

³²⁰ *Ibid*.

³²¹ Arkin, 'Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture', above n 47, 75.

Since the US Army Research Office commissioned Arkin to develop a robot³²² which, if used on the battlefield, would enable the US to fulfil its legal obligations,³²³ it is likely that the US Army Research Office informed Arkin that an ability to experience human emotions is integral to applying the principle of distinction. As a computer scientist, Arkin would not incorporate requirements which were not required by the organisation that commissioned the project.

D *A Numerical Expression of the Principle of Distinction*

Soldiers make qualitative assessments in evaluating what degree of certainty they have as to whether the target is a military objective. They do not make numerical estimates as to the likelihood that a particular target is a civilian or a protected object. Meanwhile, robots need a quantitative expression of a degree of certainty, such as a statistical probability that the target is a military objective, in order to make such assessments.³²⁴ It follows that another criterion which needs to be fulfilled in order for LARs to comply with the principle of distinction is a quantitative expression of the degree of certainty required by this rule.³²⁵

Alan Backstrom and Ian Henderson add that an algorithm, which sets the degree of certainty LARs should attain before executing an attack, should incorporate the fact that sensors may have an uncertainty of measurement.³²⁶ Thus, if there is an uncertainty of measurement of one per cent that a sensor accurately identified the character of a target, the algorithm will need to reflect this.³²⁷ For instance, imagine that the degree of certainty demanded by the law corresponds to a threshold of 95 per cent, and there is an uncertainty of one per cent that the object is in fact a military objective (such as a tank).³²⁸ In this case the machine would need to be programmed as requiring 96 per cent confidence that a target is a military objective.³²⁹

The state practice on the accuracy of weapons will be analysed in order to derive a numerical value, which corresponds to the degree of certainty required by the principle of distinction. States appear to treat munitions that in a particular percentage of cases do not accurately engage the target as violating the principle of distinction.³³⁰ In order to assess what numerical value for weapon accuracy states require, the practice of states regarding landmines, cluster munitions and air-dropped bombs will be utilised. The choice of case studies was made on the basis of availability of data regarding the accuracy standards for such weapons. Another factor was that militaries employ these weapons in diverse ways. For instance, militaries utilise landmines to deny an adversary access to an area and

³²² Arkin and Ulam, 'Behavioral Modification Derived from Moral Emotions', above n 280, 1.

³²³ Arkin, 'Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture', above n 47, 98.

³²⁴ Backstrom and Henderson, above n 23, 495.

³²⁵ *Ibid.*

³²⁶ *Ibid.*

³²⁷ *Ibid.*

³²⁸ *Ibid.*

³²⁹ *Ibid.*

³³⁰ Defense Science Board, 'Defense Science Board Task Force on Munitions System Reliability' (Report, 13 September 2005) 8–10 ('*Munitions System Reliability Report*').

to channel the enemy to move using a particular route.³³¹ Meanwhile, they employ air-dropped bombs to destroy particular military objectives.³³² The comparison of standards that states apply to weapons that are meant to be employed in different scenarios allows for a better understanding of what accuracy rate states require. Moreover, the comparison of different weapons makes it possible to assess whether it is appropriate to draw an analogy between traditional weapons and LARs. After all, LARs may be said to be a unique weapon system, because they autonomously select targets. Traditionally, human decision-makers have made an assessment as to what location and for what type of target a particular weapon would be used.

1 Landmines

Landmines are designed to explode upon contact with a person.³³³ They can be activated by both a combatant and a civilian.³³⁴ Article 5 of the *Amended Protocol II to CCW 1980* requires that if forces emplace landmines near a concentration of civilians, they should take measures to protect civilians.³³⁵ These measures could be the posting of warning signs and the fencing of the area.³³⁶ As regards remotely delivered mines, art 6 requires that they be equipped with: (1) a self-destructive device that operates within 30 days and has a dependability of 90 per cent; and (2) a self-deactivating feature which operates within 120 days and has a dependability of 99.9 per cent.³³⁷ Jean-Marie Henckaerts and Louise Doswald-Beck suggest that both provisions have the status of customary international law.³³⁸ Undoubtedly, the purpose of these rules

³³¹ John Troxell, 'Landmines: Why the Korea Exception Should Be the Rule' (2000) 30(1) *Parameters: US Army War College Quarterly* 82, 83.

³³² Royal Air Force, *General Purpose Bombs* (2015) <<http://www.raf.mod.uk/equipment/generalpurposebombs.cfm>>.

³³³ Christopher Greenwood, *Legal Issues regarding Explosive Remnants of War*, UN Doc CCW/GGE/I/WP.10 (23 May 2002) 4–5 ('*Legal Issues*').

³³⁴ *Ibid.*

³³⁵ *Protocol II to CCW 1980* art 5. Landmines which are not remotely delivered ordinarily must not be placed in areas containing concentrations of civilians unless they are placed near a military objective or unless combat between ground forces is imminent: Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 1*, above n 167, 280–1. See also *Hague Convention VIII relative to the Laying of Automatic Submarine Contact Mines*, opened for signature 18 October 1907 (entered into force 26 January 1910) art 1(3). This provision prohibits the use of torpedoes which do not become harmless when they have missed their mark to ensure that these do not become free-floating naval mines and detonate civilian ships.

³³⁶ *Protocol II to CCW 1980* art 3(10)(b).

³³⁷ *Protocol II to CCW 1980* art 6, annex 3(a).

³³⁸ According to their study, it is a rule of customary international law that '[w]hen landmines are used, particular care must be taken to minimise their indiscriminate effects': Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 1*, above n 167, 280–1. Article 3 of the *Conventional Weapons Act 2001* of the Republic of Korea states that no one is allowed to use or transfer remotely-delivered anti-personnel mines that do not fulfil any of the following: '(a) Over 90 percent of the total amount shot or dropped shall automatically detonate within 30 days (b) Over 99.9 percent of the total amount shot or dropped shall automatically detonate or otherwise lose its function as a mine within 120 days': Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 1831. For the US state practice, see Sally J Cummins and David P Stewart, 'Use of Force and Arms Control' in Sally J Cummins and David P Stewart (eds), *Digest of United States Practice in International Law 2000* (International Law Institute, 2000) 751, 754.

is to ensure compliance with the principle of distinction.³³⁹ With regard to anti-vehicle mines specifically, US policy requires that they are equipped with self-destruction and self-neutralisation mechanisms, so that no more than one in 1000 mines remains active 120 days after arming.³⁴⁰ The time span of 120 days reflects the assumption that fencing and warning signs will ensure that civilians do not enter a mined area, or that the necessity to attack the military objective which is being targeted will subside after a reasonable period of time.³⁴¹ Since, in formulating these standards, states assumed that civilians would not come into contact with landmines within 120 days of such mines being emplaced, for present purposes the 120 days aspect should be disregarded. Clearly, states intended that the embedded mechanisms should render the mines harmless in 99.9 per cent of cases.

It is possible to draw a parallel between landmines and LARs. While landmines are designed to explode when an individual comes into contact with them,³⁴² LARs are meant to seek out individuals who take a direct part in hostilities and combatants. In doing so, they are likely to encounter civilians and civilian objects. Both landmines and LARs have the potential to trigger civilian deaths. The prescribed rate for landmines of 99.9 per cent for self-destruction and self-deactivation³⁴³ may be applied for LARs, because this numerical value reflects what safeguards should be put in place to ensure that a weapon discriminates between civilians and combatants. As a result, landmines are a good starting point for analysing how the principle of distinction regulates LARs. Of course, LARs differ from landmines, because they will seek out targets, as opposed to passively waiting for civilians to come into physical contact with them. This aspect arguably makes LARs pose a greater danger to civilians than landmines. The likelihood of an LAR encountering a civilian on the battlefield is much higher than the likelihood of a civilian encountering a landmine. Urban combat, a common feature of current armed conflicts,³⁴⁴ magnifies the danger to civilians posed by LARs, which fail to properly distinguish between civilians and lawful targets. In turn, this raises the question of whether stricter regulations are needed for LARs than for landmines.

2 *Air Bombing*

The technological evolution in the accuracy of air bombing provides context for understanding the current practice of states on the required accuracy for bombs. Dating back to the Second World War, there was a 50 per cent chance

³³⁹ Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 1*, above n 167, 280.

³⁴⁰ Sally J Cummins (ed), *Digest of United States Practice in International Law 2006* (Oxford University Press, 2007) 1090.

³⁴¹ International Committee of the Red Cross, *Anti-Personnel Landmines: Friend or Foe? A Study of the Military Use and Effectiveness of Anti-Personnel Mines* (1996) 17; Reaching Critical Will, *Landmines* (2015) <<http://www.reachingcriticalwill.org/resources/fact-sheets/critical-issues/5439-landmines>>.

³⁴² *Legal Issues*, UN Doc CCW/GGE/I/WP.10, 4–5.

³⁴³ For the practice of the Republic of Korea, see Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 1831. For the US state practice, see Cummins and Stewart, above n 338, 754.

³⁴⁴ Michael Evans, *City without Joy: Urban Military Operations into the 21st Century* (Australian Defence College, 2007) 2.

that an air-delivered dumb bomb would land within a distance of 1005 metres from the target.³⁴⁵ And while states initially instructed pilots to exercise 'reasonable care' when attacking military objectives in order to ensure that civilians located close to military objectives would not be bombed,³⁴⁶ the technology simply did not allow for accurate bombing.³⁴⁷ Today, of course, the accuracy of bombs has dramatically improved.³⁴⁸ At the higher end of the scale, precision-guided munitions ('PGMs') in 50 per cent of cases are said to land within one or two metres of the designated point.³⁴⁹ This high accuracy rate is assured by the fact that PGMs home in on to the target and guide themselves to it.³⁵⁰ It is arguable, therefore, that this accuracy figure should be used to gauge the reliability rate required by the principle of distinction.

The problem with this argument, however, is that the law does not require the employment of PGMs.³⁵¹ Dumb bombs have a 50 per cent chance of landing 61 metres away from the target.³⁵² However, pilots can choose a particular flight path and angle of bomb impact to ensure that the bomb does not land on a civilian object if it misses its target.³⁵³ Since the assessment of whether the employment of dumb bombs complies with the principle of distinction depends on the pilot exercising his or her skill, it is arguable that the accuracy of this weapon should not be employed to infer how the principle of distinction might regulate LARs. The state practice of air bombing should instead be viewed against the background of improvements in bomb accuracy and the role of the skill of the pilot.

Of course, a possible counterargument is that the accuracy rate of PGMs should be used as a benchmark for the accuracy of LARs, because over time states have raised the standards for weapons such as landmines. While states have marked and mapped minefields in conflicts when they fought other developed states such as during World War II, they frequently did not observe

³⁴⁵ Richard P Hallion, *Storm over Iraq: Air Power and the Gulf War* (Smithsonian Books, 1992) 283, quoted in Michael Russell Rip and James M Hasik, *The Precision Revolution: GPS and the Future of Aerial Warfare* (Naval Institute Press, 2002) 214.

³⁴⁶ In 1938, the British Prime Minister Neville Chamberlain announced that the UK instructed its pilots to exercise 'reasonable care' when attacking military objectives in order to ensure that civilians located close to military objectives which were being targeted would not be bombed. The League of Nations adopted the international humanitarian law rules delineated by Chamberlain as governing air warfare in a non-binding resolution. United Kingdom, *Parliamentary Debates*, House of Commons, 21 June 1938, vol 337 col 937; *Records of the XIX Ordinary Session of the Assembly*, Plenary Meetings, Ordinary Debates, League of Nations Doc A.69 (1938) IX. Both sources are quoted in: Parks, above n 176, 36.

³⁴⁷ Parks, above n 176, 54.

³⁴⁸ Rip and Hasik, above n 345, 214.

³⁴⁹ Hallion, above n 345, 283.

³⁵⁰ Royal Air Force, *Paveway II & III* (2015) <<http://www.raf.mod.uk/equipment/paveway-2-and-3.cfm>>.

³⁵¹ *Service Manual of the Law of Armed Conflict*, above n 150, 323 [12.51]; Australian Defence Force, *Australian Defence Doctrine Publication (ADDP) 06.4: Law of Armed Conflict* (Australian Defence Headquarters, 2006) [4.48] <<http://www.defence.gov.au/ADFWC/Documents/DoctrineLibrary/ADDP/ADDP06.4-LawofArmedConflict.pdf>>. For the writings of scholars, see John F Murphy, 'Some Legal (and a Few Ethical) Dimensions of the Collateral Damage Resulting from NATO's Kosovo Campaign' in Andru E Wall (ed), *International Law Studies* (Naval War College, 1901–2002) vol 78, 229, 236.

³⁵² Hallion, above n 345, 283.

³⁵³ United States Department of Defense, 'Background Briefing on Targeting' (Media Briefing, 5 March 2003) <<http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2007>>.

this requirement when faced with guerrilla fighters.³⁵⁴ It was only after the Vietnam War that the US invested billions of dollars into developing mines that would self-destruct within a particular time frame.³⁵⁵ Moreover, the state consensus that states that lay landmines are responsible for them, and should either remove them or render them harmless, emerged in the 1990s.³⁵⁶ This shift in state practice partly stemmed from the fact that states realised that *Protocol II to CCW 1980* did not address the ‘ever-worsening situation on the ground’.³⁵⁷ Of course, the consideration that soldiers came into contact with landmines and that landmines impeded the mobility of troops³⁵⁸ likely played a role too. Although states bear in mind both military and humanitarian considerations when regulating weapons, the statements of states demonstrate that they were very much concerned about the humanitarian impact of landmines when they acceded to treaties that regulate this weapon.³⁵⁹

It is possible to distinguish the state practice on landmines from the state practice of air bombing. Unlike in the case of air bombing, the armed forces may not avoid injury to civilians through skilfully emplacing the landmines.³⁶⁰ Furthermore, while Western states marked and mapped landmines to prevent individuals coming into contact with them,³⁶¹ they had no control over whether an individual nevertheless came into contact with the mine. On the other hand, the pilot has control over whether he or she places the bomb on the target. What emerges from this discussion is that the state practice on air-dropped bombs

³⁵⁴ International Committee of the Red Cross, *Military Use and Effectiveness of Anti-Personnel Mines*, above n 341, 17 [22]–[23].

³⁵⁵ Landmine Monitor Core Group, *Landmine Monitor Report 1999: Toward a Mine-Free World* (Human Rights Watch, 1999) 322–3, quoted in Stuart Maslen, *Anti-Personnel Mines under Humanitarian Law: A View from the Vanishing Point* (Intersentia, 2001) 9.

³⁵⁶ Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 1*, above n 167, 285–6.

³⁵⁷ Jody Williams, ‘Landmines and Measures to Eliminate Them’ (1995) 307 *International Review of the Red Cross* <<http://www.icrc.org/eng/resources/documents/misc/57jmm9.htm>>; William Branigin, ‘US Declares It Will Not Produce Any More Antipersonnel Land Mines’, *The Washington Post* (online), 27 June 2014 <http://www.washingtonpost.com/world/national-security/us-declares-it-will-not-produce-any-more-antipersonnel-land-mines/2014/06/27/f20f6f74-fdf5-11e3-932c-0a55b81f48ce_story.html>.

³⁵⁸ *US Use of Land Mines in the Persian Gulf War*, above n 24, 11; Committee on Foreign Relations, ‘Amended Mines Protocol’ (Senate Executive Report 106–2, US Senate, 13 May 1999) <<http://www.gpo.gov/fdsys/pkg/CRPT-106erpt2/html/CRPT-106erpt2.htm>> (*Amended Mines Protocol Report*).

³⁵⁹ *Amended Mines Protocol Report*, above n 358; Mahahama Savadogo, ‘Anti-Personnel Mines’ (Speech delivered at the Inter-African Seminar on Anti-Personnel Mines, Burkina Faso, 3 June 1998), quoted in Landmine Monitor Core Group, *Landmine Monitor Report 1999: Toward a Mine-Free World* (Human Rights Watch, 1999) 30; A B Nzo ‘Address to the Signing Ceremony [of the Mine Ban Treaty]’ (Speech delivered at the Signing Ceremony of the Mine Ban Treaty 1997, Ottawa, 3 December 1997), quoted in Landmine Monitor Core Group, *Landmine Monitor Report 1999: Toward a Mine-Free World* (Human Rights Watch, 1999) 82; Jozias van Aartsen, *Government of the Netherlands* (1999) Landmine & Cluster Munition Monitor <http://www.the-monitor.org/index.php/publications/display?url=lm/1999/appendices/gov_netherlands.html>.

³⁶⁰ Landmines lie dormant until activated and do not distinguish between persons and material which trigger them. Williams, above n 357.

³⁶¹ International Committee of the Red Cross, *Military Use and Effectiveness of Anti-Personnel Mines*, above n 341, 17 [22]–[23].

should be used with caution when assessing what degree of accuracy should be demanded of LARs.

3 Cluster Munitions

Cluster munitions are an area weapon; a single dispenser releases multiple sub-munitions.³⁶² Manufacturers have consistently promised a two to five per cent failure rate for cluster munitions, but mine clearance personnel report that the failure rate can be as high as 10 to 30 per cent.³⁶³ Ninety two states have ratified the *Convention on Cluster Munitions 2008*.³⁶⁴ Article 2(2)(c) of the *Convention* outlaws the use of cluster munitions except for those which contain fewer than 10 explosive sub-munitions, where each of the sub-munitions: (1) weighs no more than four kilograms; (2) detects and engages a single military objective; (3) is equipped with both an electronic self-destruction mechanism; and (4) is equipped with a self-deactivating feature.³⁶⁵ Some states, such as the Republic of Korea³⁶⁶ and the US,³⁶⁷ have not ratified this treaty, due to regarding cluster munitions as having a continuing military utility that cannot be fulfilled by an alternative weapon. States have generally treated cluster

³⁶² United States Air Force, *Air Force Operations and the Law* (Judge Advocate General's Department, United States Air Force, 1st ed, 2002) 296, quoted in William Boothby, 'Cluster Bombs: Is There a Case for New Law?' (Occasional Paper Series No 5, Program on Humanitarian Policy and Conflict Research Harvard University, 2005) 1, 4.

³⁶³ The United States had an expectation during Operation Desert Storm 1991 that cluster munitions fired by artillery would have a three per cent failure rate. The actual failure in that armed conflict ranged from two per cent to 23 per cent. The US Army required that the failure rate not exceed five per cent: United States General Accounting Office, 'Operation Desert Storm: Casualties Caused by Improper Handling of Unexploded US Submunitions' (Report No GAO/NSIAD-93-212, United States General Accounting Office, August 1993) 5–6 ('*Operation Desert Storm Report*'); Office of the Under Secretary of Defense (Acquisition, Technology and Logistics), 'Unexploded Ordnance Report' (Submission to the United States Congress, 29 February 2000) 5 ('*Ordnance Report*'). Both sources quoted in: Mark Hiznay, 'Operational and Technical Aspects of Cluster Munitions' (2006) 4 *Disarmament Forum* 15, 19 n 12. There is a summary of findings related to cluster munition dud rate: at 22. In 2003 the British Army employed the L20A1 projectiles with M85 bomblets which have self-destruct mechanisms for the first time under combat conditions in southern Iraq. The British Ministry of Defence stated that commanders made decisions using the assumption that the failure rate was three quarters of a per cent. In fact the failure rate was four to eight per cent. Israel used the M85 cluster munitions during the Second Lebanon War 2006 and the failure rate was as high as 10 per cent. C King Associates Ltd, Norwegian Defence Research Establishment and Norwegian People's Aid, 'M85: An Analysis of Reliability' (Report, Norwegian People's Aid, 2007) 13–4 ('*M85 Report*'), quoted in Human Rights Watch, *Up In Flames: Humanitarian Law Violations and Civilian Victims in the Conflict over South Ossetia* (Human Rights Watch, 2009) 67 n 180.

³⁶⁴ *Convention on Cluster Munitions*, opened for signature 30 May 2008, 2688 UNTS 39 (entered into force 1 August 2010) ('*Convention on Cluster Munitions*').

³⁶⁵ *Ibid* art 2(2)(c).

³⁶⁶ Republic of Korea, *Explanation of Vote on L 16*, UN General Assembly, 1st Comm, 64th sess (11 October 2009), quoted in Landmine and Cluster Muniton Monitor, *Cluster Muniton Monitor 2010* (Mines Action Canada, 2010) 222.

³⁶⁷ Bureau of Public Affairs, US State Department, *US Position on the Convention on Conventional Weapons Negotiations on Cluster Munitions Protocol* (16 November 2011) <<http://www.state.gov/s/l/releases/remarks/177280.htm>> ('*US Position*').

munitions as not being comparable to landmines.³⁶⁸ Landmines are designed to remain active for a long period of time and to explode when a person comes into contact with them.³⁶⁹ States treat cluster munitions, on the other hand, as creating a danger to civilians when they do not function as intended, namely when they fail to explode and civilians come into contact with the unexploded munition.³⁷⁰

Although the UK and the US employed cluster munitions in populated areas in 2003,³⁷¹ the better view is that it is unlawful to employ cluster munitions which do not seek out military objectives in populated areas.³⁷² Being an area weapon, cluster munitions that lack heat-seeking and thermal signature-recognition laser sensors by their nature strike civilian objects and military objectives without distinction.³⁷³ The principle of discrimination, a customary international law rule applicable to international and non-international armed conflicts, prohibits the employment of weapons which cannot be directed at a specific military objective.³⁷⁴ In December 2006 Israeli television reported that the military's Judge Advocate General was gathering evidence for possible criminal charges against military officers who may have given orders for cluster bombs to be dropped on populated areas in Lebanon during the 2006 conflict.³⁷⁵ Similarly, the US Department of State observed that Israel may have violated agreements with the United States, which implement the *Arms Export Control Act 1976*, when it fired American-supplied cluster munitions at rocket launchers located in towns and villages in southern Lebanon during the Second Lebanon War in 2006.³⁷⁶

This means that the US regards the use of cluster munitions in populated areas as being unlawful, although it violated this prohibition in 2003 in Iraq. More recently, the Independent International Fact-Finding Mission on the Conflict in

³⁶⁸ *Legal Issues*, UN Doc CCW/GGE/I/WP.10, 4–5; Sean D Murphy, *United States Practice in International Law: 2002–2004* (Cambridge University Press, 2005) vol 2, 360–1; CCW Group of Governmental Experts Working Group on Explosive Remnants of War, *Report on States Parties' Responses to the Questionnaire on International Humanitarian Law & Explosive Remnants of War*, UN Doc CCW/GGE/XIII/WG.1/WP.12 (24 March 2006) 4–6 ('*Explosive Remnants of War Report*').

³⁶⁹ *Legal Issues*, UN Doc CCW/GGE/I/WP.10, 4–5; Murphy, above n 368, 360–1; *Explosive Remnants of War Report*, UN Doc CCW/GGE/XIII/WG.1/WP.12, 6.

³⁷⁰ *Explosive Remnants of War Report*, UN Doc CCW/GGE/XIII/WG.1/WP.12, 5; Murphy, above n 368, 360.

³⁷¹ Human Rights Watch, *Off Target: The Conduct of the War and Civilian Casualties in Iraq* (2003) 59. For information on the UK's conduct, see at 104.

³⁷² See, eg, Karen Hulme, 'Of Questionable Legality: The Military Use of Cluster Bombs in Iraq in 2003' (2004) 42 *Canadian Yearbook of International Law* 143, 179–80.

³⁷³ *API 1977* art 51(4)(a); Hulme, above n 372, 174.

³⁷⁴ *API 1977* art 51(4)(a). On the customary international law status of this rule, see Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume I*, above n 167, 40. See also Hulme, above n 372, 180.

³⁷⁵ Mark Tran, 'US Studies Israel's Cluster Bomb Use in Lebanon', *The Guardian* (online), 30 January 2007 <<http://www.theguardian.com/world/2007/jan/29/israelandthepalestinians.usa>>.

³⁷⁶ David S Cloud and Greg Myre, 'Israel May Have Violated Arms Pact, US Says', *The New York Times* (online), 28 January 2007 <http://www.nytimes.com/2007/01/28/world/middleeast/28cluster.html?pagewanted=all&_r=0>.

Georgia observed that the use of cluster munitions in populated areas by Russia violated the principle of discrimination.³⁷⁷

Since older cluster munition models strike civilian objects and military objectives indiscriminately in urban combat, they should not be used as a case study for analysing the accuracy rate states require of weapons. On the other hand, the cluster munitions which are exempt from the prohibition in the *Convention on Cluster Munitions 2008* are an appropriate case study. Since these sub-munitions seek out their targets,³⁷⁸ the accuracy rate with which these sub-munitions correctly identify military objectives may be used to extrapolate the accuracy rate states require of weapons. The new cluster munitions models detect the signature of military objectives and heat using an active laser and infrared sensors.³⁷⁹ The Sensor Fuzed Weapon models CBU 105 and BLU 108 recognise military objectives in greater than 99.6 per cent of instances.³⁸⁰ Although the manufacturer Textron does not specify the exact figure, the statistical estimate the company provides is close to the value of 99.9 per cent of instances. Accordingly, states in effect require that cluster munitions, which seek out military objectives based on their signature, be able to identify targets correctly in greater than 99 per cent of instances, and in many cases close to 99.9 per cent of cases. This state practice parallels that of landmines. It suggests that LARs will need to identify military objectives and individuals who take a direct part in hostilities with an accuracy rate of up to 99.9 per cent in order to comply with the principle of distinction.

4 *Synthesis of the Analysis*

The numerical value of 99.9 per cent reflects the requirement for the emplacement of landmines.³⁸¹ It also approximates the stringent requirement for cluster munitions that are not covered by the *Convention on Cluster Munitions 2008*.³⁸² On this approach, LARs should identify individuals who take direct part in hostilities and military objectives in 99.9 per cent of cases. This numerical value reflects the fundamental character of the principle of distinction. When negotiating *API 1977*, states including Kuwait, Libyan Arab Republic, Madagascar, Mali, Mauritania and Romania, among others, proposed that attackers ‘shall always make a clear distinction’ between civilians and civilian objects on the one hand, and combatants and military objectives on the other hand.³⁸³ Ultimately, states rejected this proposition because they thought that the

³⁷⁷ Independent International Fact-Finding Mission on the Conflict in Georgia, ‘Report Volume II’ (Report, September 2009) 343.

³⁷⁸ Textron, *Textron Systems’ Sensor Fuzed Weapon Proven Effective for Maritime Interdiction* (20 April 2005) <<http://investor.textron.com/newsroom/news-releases/press-release-details/2005/Textron-Systems-Sensor-Fuzed-Weapon-Proven-Effective-for-Maritime-Interdiction/default.aspx>>.

³⁷⁹ *CBU-105/BLU-108*, above n 26.

³⁸⁰ Ibid; Textron Defense Systems, ‘Sensor Fuzed Weapon’ (Information Pamphlet, 2010) 2 <http://www.textronsystems.com/sites/default/files/pdfs/product-info/sfw_datasheet.pdf>.

³⁸¹ For the practice of the Republic of Korea, see Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 1831. For the US state practice, see Cummins and Stewart, above n 338, 754.

³⁸² *Convention on Cluster Munitions 2008* art 2(2)(c).

³⁸³ The other states included Morocco, Sudan, Syrian Arab Republic and the United Arab Emirates: *Statement at the CDDH*, CDDH/III/14 Add 1 and 2 (12 March 1974).

word ‘clearly’ was redundant.³⁸⁴ This state practice suggests that states prefer the decision-maker to mistakenly conclude that a lawful target enjoys immunity, as opposed to mistakenly conclude that a civilian is taking a direct part in hostilities. Further support for this argument may arguably be found in the requirement that

in case of doubt whether an object which is normally dedicated to civilian purposes, such as a place of worship, a house or other dwelling or a school, is being used to make an effective contribution to military action, it shall be presumed not to be so used.³⁸⁵

E *Reliability Rate*

States require that the decisions of commanders and soldiers not be judged with hindsight, since they make decisions in difficult circumstances and may make mistakes.³⁸⁶ Turning to LARs, machines can malfunction and software programs have glitches.³⁸⁷ An additional legal criterion which should be applicable to LARs is an evaluation of the percentage of cases in which the software functions correctly and enables LARs to appropriately apply the rules of targeting. The reliability criterion is legally relevant because robots that frequently malfunction will be prone to targeting protected persons and objects. The fact that states excuse genuine mistakes is not applicable to the issue of the malfunction of LARs. Human beings make decisions under pressures of time, constrained resources and threats to their lives. On the other hand, scientists have an unlimited time for designing LARs. Moreover, states decide in advance whether, and if so how, to deploy these systems. Another difference between human error and machine error is that human beings may make a mistake which is confined to a particular attack. On the other hand, an LAR that malfunctions could potentially carry out many attacks, including in an urban area, before human beings detected what was happening. The following incident illustrates grave consequences which occur when a weapon system malfunctions. On 8 November 2006 a fire control unit of an Israeli artillery piece malfunctioned³⁸⁸ and fired 12 shells into a populated neighbourhood.³⁸⁹ The shells killed 19 civilians and injured 50 others.³⁹⁰

³⁸⁴ Committee III, *Summary Record of the Fourth Meeting Consideration of Draft Protocols I and II*, CDDH/III/SR4 (13 March 1974).

³⁸⁵ *API 1977* art 52(3).

³⁸⁶ Australia, *Declarations Made upon Ratification of Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 1642 UNTS 473 (21 June 1991) 473 [3]; Austria, *Reservations Made upon Ratification of Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 1289 UNTS 303 (13 August 1982) 303 [1]; Belgium, *Interpretative Declarations Made upon Ratification of Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 1435 UNTS 367 (20 May 1986) 369 [3]. Sources quoted in: Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 331–2.

³⁸⁷ Marchant et al, above n 9, 283–4.

³⁸⁸ UN Human Rights Council, *Report of the High-Level Fact-Finding Mission to Beit Hanoun Established under Council Resolution S-3/I*, UN Doc A/HRC/9/26 (1 September 2008), 9 [25]–[26], 11 [35].

³⁸⁹ The area of coverage was approximately 1.5 hectares. *Ibid* 9 [26].

³⁹⁰ *Ibid* 10 [30].

A close analogy of state practice regarding the issue of the permissible percentage in which an LAR can malfunction is the weapons reliability rate. The reliability rate refers to the percentage of cases in which a weapon fails to explode.³⁹¹ The similarity between an LAR malfunction and weapons reliability is that both inflict harm to civilians when they do not function as intended. There is also a difference between an LAR and a weapon malfunction. In the context of weapons, individuals plan an attack that complies with the rules of targeting, but the weapon fails to explode. When it comes to LARs, the malfunction which leads to civilians being attacked occurs before the robot launches the attack, but after a human decision-maker made the decision to send the robot on a mission. The fact that an LAR malfunctions after human beings make the decision to use it make LARs closer to munitions which fail to explode.

Arguably, the state practice on the accuracy rate with which weapons impact the target is also relevant to establishing the required reliability rate for LARs. States appear to treat munitions that in a particular percentage of cases do not accurately engage the target as violating the principle of distinction.³⁹² The state practice on weapon accuracy is relevant because it tells us when it is unlawful to employ an LAR, due to the LAR malfunctioning and as a result being unable to correctly apply the principle of distinction in the required number of cases. At this stage it should be noted that states normally treat the accuracy of weapons and their reliability as separate matters. Whilst states view the accuracy rates of munitions as being relevant to the application of the principle of distinction,³⁹³ they do not apply the principle of distinction to regulate a scenario where a weapon fails to explode and leaves explosive remnants of war as a result.³⁹⁴ This dividing line is less relevant to LARs. For LARs, the malfunction occurs prior to the attack, and this same malfunction shapes whether the munition impacts a lawful target as was originally intended by commanders.

Cluster munitions are a good case study to use to study state practice on the accepted reliability of weapons because there is considerable information about the reliability rate of this weapon. As already discussed, manufacturers have consistently promised a two to five per cent failure rate for cluster munitions.³⁹⁵ However, it is surely relevant that since the adoption of the *Convention on Cluster Munitions 2008*, some of the non-signatories to this treaty started to require their armed forces to use munitions with lower dud rates. The US, for instance, committed itself not to use cluster munitions which have a dud rate of more than one per cent after 2018.³⁹⁶ Prior to 2018, senior commanders will need to approve the use of cluster munitions which do not meet this specification.³⁹⁷ Moreover, the Ministry of National Defense of the Republic of

³⁹¹ *Munitions System Reliability Report*, above n 330, 9–10.

³⁹² *Ibid* 8–9.

³⁹³ *Ibid*.

³⁹⁴ *Ibid* 10.

³⁹⁵ *Operation Desert Storm Report*, above n 363, 5–6; *Ordnance Report*, above n 363, 5; *M85 Report*, above n 363, 8, 13–4.

³⁹⁶ CarrieLyn D Guymon (ed), *Digest of United States Practice in International Law* (Office of the Legal Adviser, United States Department of State, 2012) 603.

³⁹⁷ Secretary of Defense, 'Department of Defense Policy on Cluster Munitions and Unintended Harm to Civilians' (Policy Paper, United States Department of Defense, 19 June 2008) 2 <<http://www.defense.gov/news/d20080709cmpolicy.pdf>>.

Korea in August 2008 adopted a directive which precludes the acquisition of cluster munitions that exceed a failure rate of one per cent.³⁹⁸ Furthermore, India has been acquiring sensor fused cluster munitions, which have a dud rate of less than one per cent, despite having previously exported traditional cluster munition models.³⁹⁹ Therefore, states are moving in a direction where the new cluster munition models they acquire have a less than one per cent failure rate.

Of course, all these States are party to *Protocol V on Explosive Remnants of War to CCW 1980*.⁴⁰⁰ The goal of this treaty is to minimise risks to civilians posed by munitions that fail to explode.⁴⁰¹ The voluntary ‘best practice’ annex to this treaty requires that: (1) in manufacturing munitions states should strive to achieve the highest reliability rates; and (2) states, when transferring weapons to other states, should require high reliability rates.⁴⁰² Nevertheless, the statements made by states when they pledged not to use cluster munitions which had a failure rate above one per cent suggest that they made this pledge out of humanitarian concerns. The Republic of South Korea explained that it is unable to sign the *Convention on Cluster Munitions 2008* due to the ‘unique situation’ on the Korean peninsula.⁴⁰³ However, because it recognises the need to reduce humanitarian suffering caused by cluster munitions, it adopted a directive on the permissible failure rate for this weapon.⁴⁰⁴ The US statement also points to a concern to reduce humanitarian suffering.⁴⁰⁵ There is no evidence that states committed themselves to reduce the failure rate because they were concerned that cluster munitions failed to engage a sufficient number of military objectives, due to their failure to explode. Moreover, the fact that the obligations specified in the annex are voluntary⁴⁰⁶ suggests that the US, India and the Republic of Korea view themselves as being obligated to use cluster munitions with a failure rate of one per cent. In terms of state practice, NATO countries use US manufactured conventional munitions, such as artillery munitions, that have a reliability rate of

³⁹⁸ Republic of Korea, *Explanation of Vote on L 16*, UN General Assembly, 1st Comm, 64th sess (11 October 2009); UN GAOR, 1st Comm, 63rd sess, 21st mtg, Agenda Items 81 to 96, UN Doc A/C.1/63/PV.21 (30 October 2008) 26. Both sources are quoted in: Landmine and Cluster Munition Monitor, *Cluster Munition Monitor 2010* (Mines Action Canada, 2010) 222.

³⁹⁹ Letter from Jeffrey A Wieringa, Director of the US Defense Security Cooperation Agency, to Senator Robert C Byrd, Chairman of the Senate Committee on Appropriations, 26 September 2008 in Landmine and Cluster Munition Monitor, *Cluster Munition Monitor 2010* (Mines Action Canada, 2010) 217.

⁴⁰⁰ *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, opened for signature 10 October 1980, 1342 UNTS (entered into force 2 December 1983), as amended by *Protocol on Explosive Remnants of War to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, opened for signature 28 November 2003, 2399 UNTS 126 (entered into force 12 November 2006) (‘*Protocol V to CCW 1980*’).

⁴⁰¹ *Protocol V to CCW 1980* Preamble.

⁴⁰² *Protocol V to CCW 1980* annex [3] (‘*Technical Annex*’).

⁴⁰³ Republic of Korea, *Explanation of Vote on L 16*, UN General Assembly, 1st Comm, 64th sess (11 October 2009), quoted in Landmine and Cluster Munition Monitor, *Cluster Munition Monitor 2010* (Mines Action Canada, 2010) 222.

⁴⁰⁴ *Ibid.*

⁴⁰⁵ *US Position*, above n 367.

⁴⁰⁶ *Technical Annex*.

approximately 99 per cent.⁴⁰⁷ More broadly, the fact that 86 states signed up to *Protocol V to CCW 1980*⁴⁰⁸ demonstrates that many states strive to achieve a reliability rate of 99 per cent for explosive ordnance.

It is being put forward that LARs should have a higher reliability rate than 99 per cent, unlike conventional weapons. It would be odd if states required an LAR to accurately recognise lawful targets in 99.9 per cent of instances, but then tolerated a higher rate for malfunctions. This is because a malfunction prevents a robot from correctly executing its algorithm and achieving its 99.9 per cent of accuracy of target recognition. More broadly, states have control over whether to employ robots on the battlefield, and scientists have a long time in which to develop robots that meet very high specifications.

VII ABILITY OF AUTONOMOUS ROBOTS TO FULFIL LEGAL OBLIGATIONS

At this stage the various criteria which the rules of targeting can be said to require of LARs have been identified. Some of the criteria are quantitative in character in that they can be expressed in terms of numbers. Other criteria, such as the possession of compassion, are more amorphous in nature. Although human beings know what it is like to experience compassion, scientists have found it difficult to pin down in material terms what emotions are.⁴⁰⁹ For instance, although scientists are able to see on a brain scan the activity that corresponds to particular emotions, they have not yet been able to capture the actual experience of having an emotion.⁴¹⁰ The discussion will now shift to considering whether there are LAR models that can fulfil the criteria for compliance with the rules of targeting.

A *Lethal Autonomous Robots that Work on 'Simple' Algorithms*

The autonomous LARs, which lack the capacity to learn from experience, could accurately identify military objectives that have a distinct appearance and signature. Militaries already use weapons that home in on specific types of military objectives, such as Sensor Fuzed Weapons and CAPTOR mines. The difficulty arises when a military objective has a similar appearance or signature to a civilian object. For instance, the Synthetic Aperture Radar and a Moving Target Indicator⁴¹¹ displays a radar image of the area and moving vehicles.⁴¹² This technology did not enable the NATO forces during Operation Allied Force

⁴⁰⁷ *Munitions System Reliability Report*, above n 330, 23.

⁴⁰⁸ International Committee of the Red Cross, *Protocol on Explosive Remnants of War (Protocol V to the 1980 CCW Convention)*, 28 November 2003 <https://www.icrc.org/applic/ihl/ihl.nsf/States.xsp?xp_viewStates=XPages_NORMStatesParties&xp_treatySelected=610>.

⁴⁰⁹ Giovanni Frazzetto, 'How We Feel: What Neuroscience Can and Can't Tell Us about Our Emotions' (Speech delivered at the London School of Economics, London, 1 March 2014) <<http://www.lse.ac.uk/publicEvents/events/2014/03/LitFest20140301t1300vWT.aspx>>.

⁴¹⁰ *Ibid.*

⁴¹¹ Phil M Haun, *Air Power versus a Fielded Force: Misty FACs of Vietnam and the A-10 FACs of Kosovo — A Comparative Analysis* (Masters Thesis, Air University, 2004) 64 <http://www.au.af.mil/au/aupress/digital/pdf/paper/t_0008_haun_airpower_versus_fielded_force.pdf>.

⁴¹² *Ibid.*

1999 to distinguish tanks from tractors, which were pulling a trailer.⁴¹³ Given the complex nature of the battlefield, it is difficult to envisage how robots could be programmed to anticipate all scenarios where they could mistake a civilian object for a military objective. The battlefield is an unpredictable place with a 'range of infinitely varying facts' interacting with each other.⁴¹⁴ For this reason, the RoE leave a degree of discretion to the soldier regarding how best to respond to events in the circumstances.⁴¹⁵ Moreover, it is likely to be impossible to program such LARs to adhere to the other rules of targeting. Since each battlefield scenario is unique and since decision-makers apply the rule of target verification, the principle of the least feasible damage and the principle of proportionality to a particular contextual situation, a pre-written algorithm does not enable an LAR to apply these rules.

Turning to robots that learn from experience, Backstrom and Henderson argue that such LARs could learn to perform like a soldier by being tasked with giving an operator an overview of the battlefield and their own analysis of it.⁴¹⁶ LARs would then receive input, regarding what decision the operator made in that set of circumstances.⁴¹⁷ Others, such as Asaro, disagree that those robots, which can learn from experience and which do not mimic the working of a human brain, are capable of complying with the rules of targeting.⁴¹⁸ Asaro argues that IHL requires commanders to apply compassion and judgement as well as to reflexively consider the outcome of their actions for civilians.⁴¹⁹ In turn, this cannot be translated into a particular calculation or equation type algorithm.⁴²⁰ When intelligence is reduced to the performance of a specific task, one changes the definition of intelligence.⁴²¹ Intelligence is a complex skill, rather than mere ability to perform a particular task.⁴²² Thus, Asaro concludes that without human control, there is no way to guarantee that the use of force is not arbitrary.⁴²³

Asaro's argument will now be examined in the context of the principle of distinction. HRW maintains that soldiers rely on their knowledge of human emotions and use their intuition to detect subtle cues in human behaviour, in order to determine whether an individual is a civilian or a civilian who takes a direct part in hostilities.⁴²⁴ An LAR, unlike a soldier, will be unable to identify children, who are in the process of concurrently playing with toy guns and running, as enjoying immunity from attack.⁴²⁵ The problem is that LARs are unable to exercise judgement, lack compassion and do not understand the

413 Ibid.

414 Backstrom and Henderson, above n 23, 492.

415 Lin, Bekey and Abney, above n 278, 32.

416 Backstrom and Henderson, above n 23, 493.

417 Ibid.

418 Asaro, above n 21, 699–700.

419 Ibid 700.

420 Ibid.

421 Ibid 699.

422 Ibid.

423 Ibid 708.

424 Human Rights Watch and Harvard International Human Rights Clinic, *The Case against Killer Robots*, above n 147, 30–1.

425 Ibid 31–2.

situation in front of them.⁴²⁶ This example provides support for Asaro's view that intelligence may not be reduced to the performance of a task.⁴²⁷ Some of the reasons for the mistake not mentioned by HRW include an inability to link pieces of information in an abstract way and a lack of understanding of the context behind the situation. Whilst the act of running with a gun is characteristic of individuals who take a direct part in hostilities, other cues such as laughter could indicate that the conduct is probably innocent. A robot that was not exposed to the scenario of children playing with guns is unlikely to make the connection between laughter and the act of running with a gun. A robot, therefore, could conclude that the children were civilians who took a direct part in hostilities. Additionally, the example given by HRW supports Asaro's observation that robots are unable to correctly apply the principle of distinction to the situation before them, due to lacking an understanding of the social context behind the events.⁴²⁸

In response to such criticisms, Schmitt has argued that a possible solution is to use algorithms which attribute values to sensor data.⁴²⁹ Robots could be equipped with sensors, he suggests, which enable them to pick out children for instance.⁴³⁰ Arguably, contrary to Schmitt, there is no guarantee that an LAR that was programmed not to attack children⁴³¹ would not, for instance, attack teenagers who were playing with guns in an abandoned trench. Since, to an LAR, a child is a figure below a pre-specified height, it is difficult to see how the robot will differentiate teenagers from adults. Of course, a response to this argument would be that LARs could be exposed to individuals of different ages and learn to differentiate between individuals of varying ages.

A counterargument would be that it is possible to come up with scenarios where an LAR, which was exposed to many scenarios, nevertheless found it difficult to classify the status of individuals. Taking the scenario put forward by HRW further, for a robot, a scenario of an open-backed truck with refugees sitting on it, with two of the refugees holding guns to protect the convoy, is the same as the scenario of two fighters who have taken children by force from their families to exploit as child soldiers and who are transporting them in an open-backed truck to a base. Since individuals know that wars are dangerous, that militias may assault or rob civilians and that civilians feel fear as a result, they are likely to pause when they see the first scenario. Knowing that the enemy uses child soldiers, and how the adversary tends to dress and behave, soldiers will be able to tell that there are fighters in the truck in the second scenario. On the other hand, a robot will interpret that there are armed and unarmed individuals in the truck in the first scenario. It may have an insufficient number of scenarios in its database to infer that there are refugees in the truck who have guns so as to be able to protect themselves from the militia. A robot will be

⁴²⁶ *Ibid* 29.

⁴²⁷ More generally, in order to support his statement, Peter Asaro notes that social norms and historical precedents shape legal interpretation. Unlike human beings, robots are unable to interpret legal rules. Asaro, above n 21, 705.

⁴²⁸ *Ibid* 701–2.

⁴²⁹ Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law', above n 226, 17.

⁴³⁰ *Ibid*.

⁴³¹ *Ibid*.

unable to rely on emotions to understand the context behind the situation, namely that the refugees are using guns for self-protection. Philosopher Daniel Dennett suggests that it would be very difficult to write an algorithm which replicates human emotions.⁴³² Algorithms involve a hierarchical order of rules.⁴³³ However, in order to experience emotions such as love, the computer will need to have distributed architecture with different parts competing with each other for bandwidth.⁴³⁴ As a result of failing to gain 'situational awareness', the robot will not identify that there are refugees in the truck in the first scenario.

Overall, the problem is that no matter how many scenarios a robot is exposed to as part of its learning experience, it will be incapable of identifying combatants, individuals who take a direct part in hostilities and military objectives in front of it to the standard of 99.9 per cent accuracy. The battlefield is notoriously unpredictable, with a range of 'infinitely varying' variables interacting with each other.⁴³⁵ LARs will not respond appropriately to each battlefield scenario, because they will act on the basis of hundreds of coded and learned rules and exceptions. After all, despite possessing artificial intelligence, these systems will nevertheless work on a simple algorithm: '[i]f conditions of X and Z are true, then do Y'.⁴³⁶ Just because a robot could mimic a particular emotion does not mean that it understands the nature of this emotion or experiences it.⁴³⁷ These machines, therefore, will be unable to link information in an abstract fashion or rely on emotions to make inferences. Another setback is that, unlike human beings, LARs will be unable to use past scenarios in order to create a hypothesis about a new situation when they encounter an unexpected scenario. Since an LAR could not have rules and exceptions embedded in it which covered every possible scenario, it will find it difficult to correctly interpret the situation before it in 99.9 per cent of cases. As a result, the value of robots which learn from experience is likely to be confined to identifying military objectives which have particular characteristics that enable the robot to match the object to a profile. Such military objectives are likely to be confined to military objectives such as oil refineries, rocket launchers, tanks and artillery.

Another problem, highlighted by Benjamin Kastan, is that it is unclear how the LAR will determine what degree of doubt it has.⁴³⁸ For instance, how will it determine whether the situation in front of it raises a doubt of, say, 15 per cent, 20 per cent or 21 per cent? It is difficult to match degrees of belief to numerical values. As John Wigmore highlights: '[t]he truth is that no one has yet invented

⁴³² Jimmy So, 'Can Robots Fall in Love, and Why Would They?', *The Daily Beast* (online), 31 December 2013 <<http://www.thedailybeast.com/articles/2013/12/31/can-robots-fall-in-love-and-why-would-they.html>>.

⁴³³ Ibid.

⁴³⁴ Ibid.

⁴³⁵ Backstrom and Henderson, above n 23, 492.

⁴³⁶ Arkin, 'Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture', above n 47, 67.

⁴³⁷ So, above n 432.

⁴³⁸ Benjamin Kastan, 'Autonomous Weapon Systems: a Coming Legal "Singularity"?' [2013] *Journal of Law, Technology and Policy* 45, 60.

or discovered a mode of measurement for the intensity of human belief'.⁴³⁹ Consequently, it is unclear whether LARs will be able to apply the principle of distinction to situations where there is a degree of ambiguity regarding whether an individual is taking a direct part in hostilities.

Arkin has developed a blueprint for robot architecture that he argues has assuaged the concern that a robot will fail to comply with ethical standards embodied in IHL norms.⁴⁴⁰ Arkin argues that the robot could be programmed to use force only in prescribed sets of scenarios, such as to protect human life.⁴⁴¹ In order to ensure that no ethical violations occur, the system could be designed to monitor non-combatant casualties and damage to civilian property.⁴⁴² The robot would shut down when a pre-specified threshold of civilian harm had been reached.⁴⁴³ Arkin maintains that this governance framework will make LARs akin to soldiers, who abstain from committing IHL violations because they feel guilt.⁴⁴⁴

Arkin's proposition is problematic. Sharkey comments that Arkin's blueprint for a robot is akin to a thermostat, which switches off when a particular temperature has been reached.⁴⁴⁵ Accordingly, Sharkey suggests that Arkin's design for an 'ethical [g]overnor' should instead be called a 'weapons disabler'.⁴⁴⁶ Sharkey concludes that importing terms such as 'guilt' when discussing robots is dangerous.⁴⁴⁷ These terms, he argues, create a false impression that robots are aware of their obligations, are capable of bearing responsibility and can perform tasks more humanely than soldiers.⁴⁴⁸ Further support for Sharkey's argument is found in the writings of Sparrow and HRW. Sparrow⁴⁴⁹ and HRW⁴⁵⁰ suggest that individuals feel emotions, such as compassion and guilt, only when they understand what the other person is experiencing. Therefore, robots will be unable to identify situations in which to abstain from opening fire.⁴⁵¹

The studies conducted by neurologists suggest that robots will indeed be unable to distinguish between combatants and civilians, due to their inability to feel emotions. Neurologists have shown that individuals, who experience brain damage in infancy, have their logical reasoning intact and are unable to

⁴³⁹ John H Wigmore, *A Treatise on the Anglo-American System of Evidence in Trials at Common Law* (Little, Brown and Company, 3rd ed, 1940) section 2497, quoted in Peter Tillers and Jonathan Gottfried, 'Case Comment: *United States v Copeland* 369 F Supp 2d 275 (EDNY 2005) — A Collateral Attack on the Legal Maxim 'That Proof Beyond a Reasonable Doubt Is Unquantifiable?'' (2006) 5 *Law, Probability and Risk* 135, 147.

⁴⁴⁰ Arkin, 'Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture', above n 47, 18–21. For detailed discussion of the proposed robot design, see at 61–75.

⁴⁴¹ *Ibid* 69.

⁴⁴² *Ibid* 74.

⁴⁴³ *Ibid*.

⁴⁴⁴ *Ibid* 73–4.

⁴⁴⁵ Sharkey, above n 32, 794–5.

⁴⁴⁶ *Ibid* 795.

⁴⁴⁷ *Ibid*.

⁴⁴⁸ *Ibid*.

⁴⁴⁹ Sparrow, above n 147, 180–1.

⁴⁵⁰ Human Rights Watch and Harvard International Human Rights Clinic, *The Case against Killer Robots*, above n 147, 4.

⁴⁵¹ Sparrow, above n 147, 180–1; Human Rights Watch and Harvard International Human Rights Clinic, *The Case against Killer Robots*, above n 147, 4.

experience emotions, are incapable of learning social norms.⁴⁵² To date, robots can only indirectly relate to persons. For instance, there is a program by which a robot can analyse the relationship between the regions of the mouth, nose, eyes and eyebrows and identify the emotion an individual is experiencing by comparing the facial expression to the expressions of individuals it has previously analysed.⁴⁵³ Another program involves the robot processing data relating to characteristics of the speech such as pitch, frequency spectrum, energy, duration and pauses in order to identify what emotion an individual is experiencing.⁴⁵⁴ Since robots lack the ability to feel emotions, they will be unable to learn social norms regarding when the context points to the fact that a civilian is not taking a direct part in hostilities. For this same reason, a robot that stops using force when a particular threshold of civilian harm is reached cannot be said to understand the consequences of its actions.

When discussing whether a robot could apply the principle of proportionality, Wagner observes that current technologies for estimating harm to civilians which is likely to result from the employment of a particular weapon may be successfully transposed to the LAR context.⁴⁵⁵ The challenge, according to Wagner, is how an LAR could perform the context-based analysis entailed by assessing what degree of military advantage the attack is anticipated to confer.⁴⁵⁶ After all, each battlefield situation is unique, many variables interact to determine what degree of military advantage the destruction of the military objective offers, and this value cannot be quantified.⁴⁵⁷ Although a robot learns from each scenario and can amass a very large database of scenarios, the problem is that it will be unable to apply judgement to extrapolate how to treat a new scenario. And, given the complex nature of the battlefield, every scenario is a novel scenario. As was demonstrated in the discussion above, attaining 'situational awareness' enables commanders to estimate what degree of military advantage an attack offers. The problem is that a robot that learns from experience has no capacity to use its past experiences to create a hypothesis about the current situation, nor does it have capacity to use judgement to check whether it made correct inferences from its past experience.

Let us assume for the purpose of the discussion that states agreed on a common yardstick for valuing human life and military advantage offered by the destruction of various military objectives. Anderson and Waxman maintain that it could be possible to write a program for applying the principle of proportionality.⁴⁵⁸ A program would have to subtract civilian harm from the

⁴⁵² Antonio Damasio, *Descartes' Error: Emotion, Reason and the Human Brain* (Penguin Books, 1994) xiv.

⁴⁵³ Raffi Khatchadourian, 'We Know How You Feel', *The New Yorker* (online), 19 January 2015 <<http://www.newyorker.com/magazine/2015/01/19/know-feel>>.

⁴⁵⁴ Frank Hegel et al, 'Playing a Different Imitation Game: Interaction with an Empathic Android Robot' (Paper presented at the 6th IEEE-RAS International Conference on Humanoid Robots, Genoa, December 4–6 2006) 58–9.

⁴⁵⁵ Wagner, 'The Dehumanization of International Humanitarian Law', above n 175, 1398.

⁴⁵⁶ Ibid.

⁴⁵⁷ Ibid 1397–8.

⁴⁵⁸ Kenneth Anderson and Matthew Waxman, 'Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can' (Research Paper, Jean Perkins Task Force on National Security and Law, Stanford University, 9 April 2013) 12–13.

military advantage and measure this value against some predetermined standard of what is 'excessive'.⁴⁵⁹ A response to this argument would be that it was shown in the discussion above that the application of the principle of proportionality presupposes the exercise of judgement as well as application of social and moral values. In turn, emotions enable individuals to understand social⁴⁶⁰ and moral⁴⁶¹ values. The quantification of the two limbs of the proportionality equation and of the standard of 'excessive' will reshape the rule. As Kenneth Watkin observes, the principle of proportionality is far from being 'an almost scientific balancing of opposing interests on finely tuned scales of humanitarian justice'.⁴⁶² Whilst Anderson and Waxman acknowledge that weighing military gains and harm to civilians is akin to making a comparison between apples and oranges,⁴⁶³ they then proceed to propose a solution which would turn the rule into comparing the weight of two baskets of apples (or oranges).

The interpretation of what measures it is 'feasible' to take in the circumstances poses similar difficulties for LARs, as does the application of the principle of proportionality. Imagine that an LAR encounters a compound and registers that there is a day care centre and a military barracks inside. The compound is surrounded by an LAR detection system. However, it is known to the robot that the system will not detect it if it were to cross the grid in less than one tenth of a second. The LAR estimates that 15 family members will die if it drops a large bomb on the entire compound. However, if the robot enters the compound and drops smaller munitions on individual barracks, it is anticipated that only three civilians will die. The robot measures the wind speed and estimates that the likelihood of entering the compound without being detected is 65 per cent. How is a robot to weigh the benefit of sparing 12 civilians against the likelihood of 35 per cent of being shot down when assessing whether it is 'feasible' to enter the compound? This decision requires the application of judgement, moral values and social values. In turn, emotions and compassion enable decision-makers to discern what decision morality⁴⁶⁴ and social values⁴⁶⁵ require him or her to take. It could be very difficult to program a robot that learns from experience to understand emotions and social values, since this requires a distributed architecture.⁴⁶⁶ This problem cannot be corrected by exposing a robot to more military scenarios. Since each military scenario is unique, a robot cannot simply refer to its database to determine how to respond to a situation it has not encountered before. Therefore, LARs will find it difficult to apply the rule of target verification and the principle of the least feasible damage.

⁴⁵⁹ Ibid.

⁴⁶⁰ Damasio, above n 452, xiv.

⁴⁶¹ Ilana Simons, *The Four Moral Emotions* (15 November 2009) Psychology Today <<https://www.psychologytoday.com/blog/the-literary-mind/200911/the-four-moral-emotions>>; Gertrud Nunner-Winkler and Beate Sodian, 'Children's Understanding of Moral Emotions' (1988) 59 *Child Development* 1323, 1323–4, 1336.

⁴⁶² Watkin, above n 169, 4.

⁴⁶³ Anderson and Waxman, above n 458, 12.

⁴⁶⁴ Johnson, above n 317, 238.

⁴⁶⁵ Damasio, above n 452, xvi.

⁴⁶⁶ So, above n 432.

Another concern is that programmers may be unable to code LARs that act in a predictable fashion⁴⁶⁷ in 99.9 per cent of cases, due to the complexity of the programs. Specifically:

Now, programs with millions of lines of code are written by teams of programmers, none of whom knows the entire program; hence, no individual can predict the effect of a given command with absolute certainty, since portions of large programs may interact in unexpected, untested ways.⁴⁶⁸

Consequently, it is concluded that LARs that work on simple algorithms, but are able to learn from experience, are likely to be incapable of complying with the rules of targeting.

B *Legality of Autonomous Robots that Emulate the Human Brain*

To date, scientists have concentrated on developing an algorithm that mirrors how the human brain performs cognitive functions, but have omitted examining how to replicate the interaction between emotions and reasoning.⁴⁶⁹ Assuming the scientists succeed, it needs to be examined in greater detail whether Asaro's argument holds that, when intelligence is reduced to the performance of a specific task, one changes the definition of intelligence.⁴⁷⁰ Antonio Damasio is a neurologist who has observed many patients with brain damage and has shown that emotions assist the reasoning process.⁴⁷¹ Consequently, good decision-making necessitates the use of both emotions and reasoning.⁴⁷² Moreover, patients whose brain became damaged in childhood were unable to learn social norms and ethical rules.⁴⁷³ On the other hand, those who experienced brain damage as adults understood these norms, but were unable to adhere to them.⁴⁷⁴

Damasio's findings⁴⁷⁵ support Asaro's suggestion that it is dangerous to reduce intelligence to how humans perform particular tasks.⁴⁷⁶ If scientists want to capture human decision-making, they will need to create an algorithm that mirrors how reasoning and emotions reinforce each other. To date, researchers have, in part, taken up Damasio's findings. They are looking to incorporate the concept of emotions into algorithms so as to enable a robot to learn from experience.⁴⁷⁷ For instance, they embedded an instruction that an emotion of fear is linked with the action of bumping into a wall, in order to decrease the number of times the robot bumped into the wall.⁴⁷⁸ Arguably, the limitation of this research is that the programmers program emotions such as fear as a negative

⁴⁶⁷ Marchant et al, above n 9, 283–4.

⁴⁶⁸ Ibid 284.

⁴⁶⁹ Dvorsky, above n 102.

⁴⁷⁰ Asaro, above n 21, 699.

⁴⁷¹ Damasio, above n 452, xiii–xiv.

⁴⁷² Ibid xii–xiv.

⁴⁷³ Ibid xiii–xiv.

⁴⁷⁴ Ibid.

⁴⁷⁵ Ibid xii–xiv.

⁴⁷⁶ Asaro, above n 21, 699.

⁴⁷⁷ Sandra Clara Gadanho and John Hallam, 'Emotion-Triggered Learning in Autonomous Robot Control' (2001) 32 *Cybernetics and Systems* 531, 531.

⁴⁷⁸ Ibid 540.

value.⁴⁷⁹ Whenever a robot bumps into a wall, for instance, it records this event as a negative experience, which it should avoid in the future.⁴⁸⁰ This means that a robot is not aware of what pain feels like or why it should be avoided. Consequently, such algorithms do not allow robots to understand the nature of emotions, to experience emotions or to apply emotions in the decision-making process.⁴⁸¹ Puzzlingly, researchers who are working on recreating the workings of a human brain chose to focus solely on replicating the workings of logic and reasoning.⁴⁸² Damasio's research suggests that this is problematic. The experiencing of emotions enables individuals to make sound decisions and to learn social norms.⁴⁸³ It was already demonstrated that the application of the rules of targeting to battlefield scenarios requires that the decision-maker be able to experience emotions and to be guided by social values. Consequently, if researchers want to create an LAR which mimics the working of a human brain, they should replicate how emotions and reasoning reinforce each other.

Moreover, even if it were possible to write an algorithm that combined emotions and logical reasoning, such an algorithm in and of itself would not enable LARs to make good targeting decisions. As already discussed, the rules of targeting presuppose that the decision-maker is able to experience emotions,⁴⁸⁴ applies judgement in interpreting what a rule of targeting requires in a particular context,⁴⁸⁵ is guided by compassion,⁴⁸⁶ recognises what individuals feel, is able to apply moral norms and is able to use past experiences to determine how to act in an unfamiliar situation. Research in the field of neurology suggests that in order for a robot to emulate how a commander uses logic, emotions and moral norms to apply the rules of targeting, a robot will have to be embedded into society as a fully participating member, so that it can learn social and ethical norms.

A possible counterargument is that it is unclear whether being embedded into society will allow robots to learn to act according to ethical rules. This is because individuals sometimes act contrary to ethical, social and legal norms. Of course, there are leaders who are complicit in the commission of crimes against humanity, soldiers who commit war crimes and individuals who commit crimes. Nevertheless, the fact that domestic legal systems⁴⁸⁷ and international law⁴⁸⁸

⁴⁷⁹ Ibid 532.

⁴⁸⁰ Ibid 542.

⁴⁸¹ Ibid 537.

⁴⁸² Dvorsky, above n 102.

⁴⁸³ Damasio, above n 452, preface.

⁴⁸⁴ Wellbrink, above n 147; Human Rights Watch and Harvard International Human Rights Clinic, *The Case against Killer Robots*, above n 147, 4; Sparrow, above n 147, 180–1.

⁴⁸⁵ Grut, above n 220, 11; Asaro, above n 21, 699.

⁴⁸⁶ Human Rights Watch and Harvard International Human Rights Clinic, *The Case against Killer Robots*, above n 147, 4; Sparrow, above n 147, 180–1.

⁴⁸⁷ See, eg, *War Crimes Act 1945* (Cth) s 3; Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 3862–3.

⁴⁸⁸ Article 85(5) of *API 1977* states that grave breaches of the relevant instruments amount to the commission of war crimes. These provisions have customary international law status. The *Rome Statute* has a very similar provision: *Rome Statute of the International Criminal Court*, opened for signature 17 July 1998, 2187 UNTS 90 (entered into force 1 July 2002) art 8(2) ('*Rome Statute*'). See also Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 1*, above n 167, 568.

prohibit the commission of crimes corroborates that societies prefer individuals to be treated fairly and humanely. Moreover, as Ryan Tonken underscores:

Although it is true that human soldiers are capable of moral transgressions, they are also capable of behaving in line with law and morality, and the great majority of human soldiers that are capable of immoral action do not actually behave immorally.⁴⁸⁹

Additionally, 'humans are also capable of morally praiseworthy and supererogatory behaviour'.⁴⁹⁰ Soldiers sometimes go beyond their 'call of duty' by sacrificing themselves for civilians.⁴⁹¹ This suggests that robots could learn about ethical norms and compassion through being embedded into society.

It is unclear whether society will accept robots as their peers to reduce casualties among the armed forces. On the one hand, Kerstin Dautenhahn proposes that it is possible to develop personalised robots which can serve as companions to people and can adapt to the needs of those individuals.⁴⁹² Basic attributes could be coded by creators but robots could then acquire additional characteristics, as they interact with humans to meet their expectations.⁴⁹³ Presently, scientists propose to develop robots ranging from those that will assist teachers in classrooms⁴⁹⁴ to those that will look after the elderly.⁴⁹⁵ There is even some evidence that humans will accept robots. For instance, Fumihide Tanaka, Aaron Cicourel and Javier Movellan report that children treated a robot, which was remotely controlled by a human, as a peer, rather than as a toy.⁴⁹⁶ Nevertheless, knowledge will have to progress very far before scientists gain an understanding into how humans process events unfolding before them and develop robots that use emotions in their interactions with the world.⁴⁹⁷ Thus, it may be impossible to develop a robot which will recognise the complexity of human emotions and be capable of replicating human responses.⁴⁹⁸

⁴⁸⁹ Ryan Tonkens, 'The Case against Robotic Warfare: A Response to Arkin' (2012) 11 *Journal of Military Ethics* 149, 151 (emphasis altered).

⁴⁹⁰ Ibid.

⁴⁹¹ Ibid.

⁴⁹² Kerstin Dautenhahn, *Robots We Like to Live With?! A Developmental Perspective on a Personalized, Life-Long Robot Companion* (Research Paper, Adaptive Systems Research Group, University of Hertfordshire, 2004) 4. To this end, extensive user studies will need to be conducted to illuminate how different robot behaviours influence people's attitudes, opinions and preferences towards robots: at 5.

⁴⁹³ Ibid 4–5.

⁴⁹⁴ Fumihide Tanaka, Aaron Cicourel and Javier R Movellan, 'Socialization between Toddlers and Robots at an Early Childhood Education Center' (2007) 104 *Proceedings of the National Academy of Sciences of the United States of America* 17954, 17954.

⁴⁹⁵ Conrad Bzura et al, *The Emerging Role of Robotics in Personal Health Care: Bringing Smart Health Care Home* (Bachelor of Science Thesis, Worcester Polytechnic Institute, 2012) 4.

⁴⁹⁶ Tanaka, Cicourel and Movellan, above n 494, 17957.

⁴⁹⁷ Simon Baron-Cohen, 'The Empathizing System: A Revision of the 1994 Model of the Mindreading System' in Bruce J Ellis and David F Bjorklund (eds), *Origins of the Social Mind: Evolutionary Psychology and Child Development* (Guilford, 2005) 468; Ginevra Castellano and Christopher Peters, 'Socially Perceptive Robots: Challenges and Concerns' (2010) 11 *Interaction Studies* 201, 204.

⁴⁹⁸ Dautenhahn, above n 492, 1.

What about creating robots that are better than human beings, either more intelligent⁴⁹⁹ or lacking undesirable qualities such as fear or self-preservation?⁵⁰⁰ An observation directed at soldiers is that, since they may be influenced by self-preservation,⁵⁰¹ they may fire at a person whom they merely suspect poses a threat. The upshot of this trail of analysis is that as the feeling of self-preservation is not relevant to robots, they may be superior to humans in complying with the duty to distinguish between civilians and combatants.⁵⁰² A response to this observation would be that it is of course the case that different individuals may have varying degrees of willingness to sacrifice themselves. However, this does not mean that states will program robots to sacrifice themselves. All states have limited budgets and incur considerable expenditure on social programs such as healthcare and education. No state will agree to lose LARs in combat at a rapid rate. Additionally, the advanced nature of LARs is likely to mean that it will take time to repair such advanced systems. The fact that states keep how such technologies are designed a state secret is another factor for why states would want to prevent their adversaries from disabling such systems in order to gain knowledge about how these systems operate. In practice, therefore, states are unlikely to create robots which will sacrifice themselves. They are likely to treat LARs in the same way as they do soldiers. For this reason, programmers will program robots with a setting that is akin to the experience of fear or the desire for self-preservation.

As to the issue of intelligence, what lies behind IHL is the desire to reduce suffering for compassionate reasons, whilst enabling the parties to the conflict to overcome the adversary.⁵⁰³ There is a need for decision-makers who show greater compassion and desire for self-sacrifice than human beings, rather than for more intelligent decision-makers. It is unclear whether LARs could be programmed to exhibit qualities such as compassion to a greater extent than human beings given the nature of armed conflict. As technologies develop, states will continue to want to defeat the enemy with the least amount of time and resources. States, therefore, are unlikely to recalibrate the balance between military necessity and humanitarian considerations, which is currently embedded within the rules of targeting, in such a way as to place greater weight on humanitarian considerations. At best, LARs will apply the rules of targeting in the same way as decision-makers do currently.

Of course it is true that robots could be programmed not to commit war crimes. Rape is a war crime,⁵⁰⁴ and, as the Special Rapporteur Christof Heyns explains, robots do not rape.⁵⁰⁵ However, it is also the case that humans could

⁴⁹⁹ Michael Ordoña, 'Movies Ask: Should We Be Afraid of Artificial Intelligence?', *The San Francisco Chronicle* (online), 25 February 2015 <<http://www.sfchronicle.com/movies/article/Movies-ask-Should-we-be-afraid-of-artificial-6101865.php>>.

⁵⁰⁰ Karlsson, above n 281.

⁵⁰¹ Michael Lewis, 'New Warfare Technologies, New Protection Challenges' (Speech delivered at Harvard University, Cambridge, 24 April 2014) <<http://vimeo.com/92871619>>.

⁵⁰² Ronald C Arkin, *Governing Lethal Behaviour in Autonomous Robots* (CRC, 2009) 29–30.

⁵⁰³ *St Petersburg Declaration* Preamble.

⁵⁰⁴ *Rome Statute* art 8(2)(b)(xxii); *United States v John G Schultz*, 4 Court-Martial Reports 104 (Court of Military Appeals 1952), quoted in Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 2202.

⁵⁰⁵ Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*, UN Doc A/HRC/23/47 (9 April 2013) 10.

use robots in ways which violate IHL. For instance, the forces could use a robot to rape a fighter as a form of torture in order to elicit information about the location of his or her commander. Dara Cohen comments that a machine could perform horrible tortures, which most individuals would not engage in due to finding such conduct unbearable.⁵⁰⁶ For instance, the act of rape could potentially last much longer than if it were to be committed by a human being, and therefore inflict greater suffering. It is, therefore, far from clear whether the employment of robots will reduce the incidence of war crimes.

Assuming that it could be possible to develop a robot which reasons like a human and to embed robots into society, the next question is whether scientists could test such robots to ensure that they perform on the battlefield as intended. As the autonomy of the weapon system increases, so does the unpredictability with which that system functions.⁵⁰⁷ There are currently no methods for testing and evaluating autonomous systems.⁵⁰⁸ Michael Fisher, Louise Dennis and Matt Webster put forward that systems which have artificial intelligence should be tested differently from systems that lack such intelligence because they make their own decisions regarding what to do.⁵⁰⁹ Specifically, scientists should check that systems that emulate neural networks, genetic algorithms and complex control systems, among other elements, use reasons that their developers want them to use in reaching particular decisions.⁵¹⁰ Even then, because scientists lack a precise model of the real world, they can never know what the effect of a robot's decision will be.⁵¹¹ Given the fact that the world presents individuals with myriad situations,⁵¹² it is arguably difficult to see how scientists could check that the robot uses correct reasons for every situation with which it might be faced. There is another reason why such robots are problematic. Contrary to Fisher, Dennis and Webster, it is insufficient to test robots only for reasons on the basis of which the machines make their decisions.⁵¹³ Surely, in order for a robot to comply with the standard of engaging correct targets in 99.9 per cent of cases, it is additionally vital that developers are able to predict what action the robot will take in every possible situation. At the moment, scientists believe that this is unachievable.⁵¹⁴

Yet another danger of employing robots with artificial intelligence is that they could develop volition. Arguably, an LAR which worked on an algorithm that mimicked the neural networks of a human brain, and which was programmed to be able to learn from experience, could act beyond the parameters set by the programmers. An LAR could, for example, decide that it preferred not to be

⁵⁰⁶ Email from Dara Cohen to Charli Carpenter, 14 May 2014 in Charli Carpenter, "'Robot Soldiers Would Never Rape': Unpacking the Myth of the Humanitarian War-Bot" on *Duck of Minerva* (14 May 2014) <<http://duckofminerva.com/2014/05/robot-soldiers-would-never-rape-un-packing-the-myth-of-the-humanitarian-war-bot.html>>.

⁵⁰⁷ *Autonomous Weapons Systems Report*, above n 16, 19.

⁵⁰⁸ *Ibid* 8.

⁵⁰⁹ Michael Fisher, Louise Dennis and Matt Webster, 'Verifying Autonomous Systems: Exploring Autonomous Systems and the Agents That Control Them' (2013) 56 *Communications of the ACM* 84, 84.

⁵¹⁰ *Ibid* 86.

⁵¹¹ *Ibid* 89.

⁵¹² *Ibid* 86.

⁵¹³ *Ibid*.

⁵¹⁴ *Ibid* 89.

damaged by the enemy's fire, overwrite its program and act contrary to instructions. It could assess that the likelihood of being captured and scrapped was small. For this reason, it is possible that robots will not be deterred from disobeying orders. Scientists have not indicated that in the future, were technology to advance, robots could be programmed to have an emotion of revulsion to violating the IHL norms. Of course, Peter Lee cautions that science fiction films create a false assumption that the scenario of a robot overwriting its program is a common one.⁵¹⁵ However, he makes this remark in relation to robots that function on an algorithm and he does not discuss whether robots that emulate the working of a human brain will exhibit the same consistency in their performance.

If they do, then they could potentially be held criminally responsible if they were to commit war crimes. This is because their conduct will be accompanied by the necessary volition and mental element.⁵¹⁶ However, it would be difficult to hold programmers and manufacturers criminally responsible under the doctrine of command responsibility for failing to prevent robots from overwriting their program. The commanders under the doctrine of command responsibility have a duty to prevent the commission of war crimes by their subordinates.⁵¹⁷ According to *API 1977* art 86(2), commanders are criminally responsible for war crimes committed by their subordinates

if they knew, or had information which should have enabled them to conclude in the circumstances at the time, that he [or she] was committing or was going to commit such a breach and if they did not take all feasible measures within their power to prevent or repress the breach.⁵¹⁸

This rule has customary international law status in international⁵¹⁹ and non-international armed conflict.⁵²⁰ In order to discharge this duty, commanders have to act as soon as they are put on notice that there is a risk that a subordinate may commit a war crime.⁵²¹ However, if commanders lack the information

⁵¹⁵ Peter Lee, 'Autonomous Weapon Systems and Ethics' in *Autonomous Weapons Systems Report*, above n 16, 54.

⁵¹⁶ *Prosecutor v Kayishema (Judgement)* (International Criminal Tribunal for Rwanda, Trial Chamber II, Case No ICTR-95-1-T, 21 May 1999) [197], quoted in Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 3692; *Prosecutor v Delalic (Judgement)* (International Criminal Tribunal for the Former Yugoslavia, Trial Chamber II, Case No IT-96-21-T, 16 November 1998) [326], quoted in Henckaerts and Doswald-Beck, *Customary International Humanitarian Law Volume 2*, above n 139, 3696.

⁵¹⁷ *Prosecutor v Orić (Judgement)* (International Criminal Tribunal for the Former Yugoslavia, Trial Chamber II, Case No IT-03-68-T, 30 June 2006) [326] ('*Orić*').

⁵¹⁸ *API 1977* art 86(2).

⁵¹⁹ *Trial of Wilhelm von Leeb and Thirteen Others* (United States Military Tribunal, Nuremberg, Case No 72, 30 December 1947 — 28 October 1948) in the United Nations War Crimes Commission (ed), *Law Reports of Trials of War Criminals* (His Majesty's Stationery Office, 1949) vol XII, 75.

⁵²⁰ *Prosecutor v Hadžihasanović (Decision on Joint Challenge to Jurisdiction)* (International Criminal Tribunal for the Former Yugoslavia, Trial Chamber, Case No IT-01-47-PT, 12 November 2002) [93], [119].

⁵²¹ *Orić* (International Criminal Tribunal for the Former Yugoslavia, Trial Chamber II, Case No IT-03-68-T, 30 June 2006) [322], [323].

available to them that indicates the existence of such a risk, they need not take steps to acquire particular information.⁵²²

If programmers do not foresee that an LAR may overwrite its program, and if this in fact occurs, then those programmers will evade responsibility. Another problem is that on the application of the ICTY's judgement in *Prosecutor v Blaškić*⁵²³ the programmer cannot be held responsible if the robot disobeys the orders that were given to it, unless the programmer deliberately took the risk that this would occur. The challenge of holding programmers responsible for LARs that overwrite their program raises the question of whether society should risk such a possibility occurring. There is a possibility that some manufacturers will put products on the market believing that the likelihood of the robot overwriting its code is sufficiently small to warrant taking such a risk.

VIII CONCLUSION

The appraisal of LARs indicates that, in order to be lawful, LARs will need to make decisions in exactly the same way as human beings in the same circumstances. Doubt was raised as to whether machines will ever be able to perform to the required standard. Another relevant consideration is that experience with weapons such as landmines and cluster munitions shows that there is a discrepancy between manufacturers' warranties and reality.⁵²⁴ Furthermore, because governments and businesses are focused on return on investment,⁵²⁵ there is pressure on researchers to show results. Another cause for caution is that the robotics market is estimated to reach US\$19.41 billion by 2020.⁵²⁶ There is, therefore, a danger that businesses will offer beguiling solutions in the hope of filling a new niche in the defence and acquisitions market. The legal community should take these promises with a grain of salt.

Other questions which will need to be resolved before LARs can be employed go beyond IHL. For instance, is it moral to delegate the decision of whether to deprive an individual of his or her life to a machine?⁵²⁷ Is creating a robot which reasons and has emotions wired into it similar to inserting genes from other organisms into human genes? If so, is this ethical? Is it ethical to deprive of volition a thing which can think and experience emotions, and dispose of it as society wishes? Might such a machine be akin to a horse exploited by humans in battle, simply because it cannot tell its owner that it does not want to transport military equipment? If so, will we need a charter protecting the rights of robots

⁵²² Ibid.

⁵²³ *Prosecutor v Blaškić (Judgement)* (International Criminal Tribunal for the Former Yugoslavia, Trial Chamber, Case No IT-95-14-T, 3 March 2000) [592], [653], [741].

⁵²⁴ *Operation Desert Storm Report*, above n 363, 5–6; *Ordnance Report*, above n 363, 5; *M85 Report*, above n 363, 21.

⁵²⁵ Ripley Engineering, *Robotics Company Seeks Funding for Product & Infrastructure Development* (19 December 2013) Merar <<https://www.merar.com/meeting-place/india/technologies/invest-into-the-huge-and-potentially-untapped-industrial-robotics-market-in-india-robotics-research-company-offers-great-roi/>>.

⁵²⁶ Markets and Markets, 'Service Robotics Market Worth \$19.41 Billion by 2020' (Press Release) <<http://www.marketsandmarkets.com/PressReleases/service-robotics.asp>>.

⁵²⁷ Tomasi, above n 117, 1–2.

in addition to that which protects the rights of animals?⁵²⁸ The Republic of South Korea is drafting a *Robot Ethics Charter* which aims to prevent human beings abusing robots and vice versa.⁵²⁹ These questions illustrate the controversy posed by developing truly autonomous robotic systems. Societies should not be seduced by the promise of casualty-free wars and make rash decisions about employing LARs. Since many of these questions touch on ethics, they will need to be addressed by philosophers and policy makers and not just by lawyers.

⁵²⁸ See generally *Animal Welfare Act*, 7 USC §§ 2131–2156 (1966), quoted in Clifford J Sherry, *Animal Rights: A Reference Handbook* (Santa Barbara: ABC-CLIO, 2nd ed, 2009) 96; *Food Security Act*, 16 USC §§ 3801–3862 (1985), quoted in Clifford J Sherry, *Animal Rights: A Reference Handbook* (Santa Barbara: ABC-CLIO, 2nd ed, 2009) 102; *Animal (Scientific Procedures) Act 1986* (UK), quoted in Clifford J Sherry, *Animal Rights: A Reference Handbook* (Santa Barbara: ABC-CLIO, 2nd ed, 2009) 113; ‘Genesis’ in *The Holy Bible*, King James Version (Hendrickson Marketing, 2011) chapter 1, verses 28–30.

⁵²⁹ ‘Robotic Age Poses Ethical Dilemma’, *BBC News* (online), 7 March 2007 <<http://news.bbc.co.uk/2/hi/technology/6425927.stm>>.