Kent Academic Repository
Full text document (pdf)

Citation for published version

Huelss, Hendrik (2019) Deciding on appropriate use of force: human-machine interaction in weapons systems and emerging norms. Global Policy . ISSN 1758-5880.

DOI

https://doi.org/10.1111/1758-5899.12692

Link to record in KAR

https://kar.kent.ac.uk/73849/

Document Version

Publisher pdf

Deciding on Appropriate Use of Force: Human-machine Interaction in Weapons Systems and Emerging Norms

Hendrik Huelss [iD]

University of Kent

Abstract

This article considers the role of norms in the debate on autonomous weapons systems (AWS). It argues that the academic and political discussion is largely dominated by considerations of how AWS relate to norms institutionalised in international law. While this debate on AWS has produced insights on legal and ethical norms and sounded options of a possible regulation or ban, it neglects to investigate how complex human-machine interactions in weapons systems can set standards of appropriate use of force, which are politically normatively relevant but take place outside of formal, deliberative law-setting. While such procedural norms are already emerging in the practice of contemporary warfare, the increasing technological complexity of AI-driven weapons will add to their political-normative relevance. I argue that public deliberation about and political oversight and accountability of the use of force is at risk of being consumed and normalised by functional procedures and perceptions. This can have a profound impact on future of remote-warfare and security policy.

1. Autonomous weapons systems and the question of norms

The policy of extensive deployment of armed drones in the last 15 years, representing an era of remote-warfare featuring long-standing interventions involving only limited ground troop deployment, approaches the next stage: the development of autonomous weapons systems (AWS) and the political debate on their possible regulation and prohibition has raised attention in recent years in the academic and political community and even in the wider public. While state parties to the UN's Convention on Certain Conventional Weapons (UN-CCW) in Geneva discuss the case of AWS since 2014, their broader political implications are still understudied.

The political and academic debates on AWS focus predominantly on how AWS challenge international law (Asaro, 2012; Grut, 2013; Kastan, 2013; Noone and Noone, 2015; Sehrawat, 2017) as well as ethics (Heyns, 2016; Johnson and Axinn, 2013; Leveringhaus, 2016; Sharkey, 2008). While both dimensions are interrelated and arguments for why AWS are legally problematic in terms of International Humanitarian Law (IHL) are also motivated by ethical concerns, such as human dignity and the question of whether machines should ultimately have the decision-making power to end human life, the current debate clearly takes place in the margins of international law. Certainly, cases such as the Nuclear Test Ban Treaty, the Treaty on the Prohibition of Nuclear Weapons, the Mine Ban Treaty (Ottawa Convention), or the Protocol on Blinding Laser Weapons underline the importance of legal norms for defining when and how it is

appropriate to use force. They also prove the ability of the international community to have a significant impact on the trajectory of how force is used.

However, the international community has made slow progress in their consideration of what AWS are and do. This is widely criticised not only by NGOs, for example by the 'Campaign to Stop Killer Robots', but also by countries calling for a prohibition of AWS such as 28 UN-CCW state parties (Campaign To Stop Killer Robots, 2018). A main problem of the UN-CCW process is its inability to find a shared definition of autonomy as the crucial, qualifying feature of AWS in relation to human agency. Likewise, conceptualising 'meaningful human control' (MHC) (Crootof, 2016; Moyes, 2016; Roff and Moyes, 2016) – a term that gains currency at the moment – proves to be very difficult, not least because control and autonomy are interrelated and are multi-dimensional issues. The discussion is hence dominated by attempts to find a common ground for formulating and institutionalising norms governing the use of force by providing guidelines of when and how the use of force is appropriate, if specific weapons technologies are used. In contrast to the contestation of drone-warfare (Kaag and Kreps, 2014), technical aspects are therefore now in the centre of interest.

But the focus on the seeming uniqueness of a new generation of autonomous weapons and the resulting problem of human control overshadows the importance of (existing) practices of human-machine interaction in the use of force. The importance for the US military, for instance, of developing and testing artificial intelligence (AI) solutions funded by multi-billions of US$ in the 'AI Next Campaign' (see

DARPA, 2018) underlines the complexity of the issue at stake, which is not only about 'autonomy in the critical functions of weapons systems' (ICRC, 2016) but also about an extensive arrangement of actors and technologies influencing the use of force. For this reason, the narrow focus on AWS can be problematic: while the loss of MHC is arguably accelerating with the increasing role of AI-driven technologies, ensuring an acceptable extent of MHC is always important but often problematic in the practice of using force.

Complex human-technology interaction is a significant feature of fighter jets, drones, or cruise missiles, in which humans often fail to exert meaningful control because they rely on system information and lack an external basis to doubt the electronic representation of reality. Nevertheless, this compromised human control is hardly contested. If we understand norms as 'standards of appropriateness' (see Bode and Huelss, 2018), practices comprising the procedural interplay of different technologies, elements of weapons systems, and human actors can create norms that establish a standard of appropriate human control and hence acceptable use of force. These norms are procedural norms because they are the product of technical-functional processes in contrast to political deliberation.

While this article does not deliver an extensive review of the debate on law and ethics of AWS, it aims at discussing the role of norms outside of these deliberative settings. The article also contributes to emphasising the *political* relevance of AWS by outlining their implications for how force is used, which can have a structural impact on the development and conduct of security policy and warfare in the future. The article is organised as follows: first, it introduces the research dimension of norms and AWS, outlining the shortcomings of the main research focus. Second, it discusses the implications of autonomous features for the emergence of norms. Third, in concluding, the article outlines the political implications of the described technological development.

2. Conceptualising norms and the impact of AI

While the political community largely operates within the margins of international law-making, International Relations (IR) scholarship has diversified the perspective on what norms are and why they matter. For example, initial constructivist research stressed the simultaneously regulative and constitutive qualities of norms, which do not only prohibit but also shape opinions and identity (Checkel, 1998; Klotz, 1999; Kratochwil, 1989). Without providing a more detailed review of norms research in IR given the limitations of this article, it is important to note for the following discussion that there are different types of norms: I suggest differentiating between legal, ethical, and procedural norms. Considering the broad definition of norms as standards of appropriateness, it is apparent that appropriateness can have very different meanings to different actors in different contexts. Legal appropriateness codified in legal norms purports a formalised and (supposedly) fixed definition of what is prohibited. But international law actually leaves ample

room for interpretation and political decisions (see Koskenniemi, 2011). Law only offers a narrow understanding of what is permissible, instead of giving precise guidance on what is 'right' or legitimate in a moral sense. In contrast, ethical appropriateness is about providing ethical norms that can inform what is the morally 'right' thing to do. Ethical norms can overlap with legal norms such as basic human rights, but their formalisation and codification are less extensive.

In this regard, norms can be further differentiated by their function: legal norms (prohibiting) and ethical norms (prescribing) are conceptualised as shaping what is normal (the normality). They influence behaviour and are regulative but also constitutive for how social relations work. Procedural norms, in contrast, emerge in the normality of practices out of specific understandings of what is appropriate. This appropriateness might be in line with standards of what is prohibited or prescribed, but the role and impact of procedural norms is foremost functional. In this sense, we can further differentiate between a prohibiting and prescribing normativity, based on two different foundations for defining what is the (legal and ethical) right thing to do, and what is normal/normality. What is a normal action is defined by the average behaviour in the statistical sense, representative of how the majority of people act. This normality is also constitutive by setting a framework of appropriate action without resorting to legal or ethical norms.

The underlying theoretical assumption of procedural norms is that once certain practices as way of doing things are established, they manifest in an understanding of what is procedurally appropriate. In other words, in any organisational context such as the military or in bureaucracies, procedures and ways of operating exist that are established and correspond to certain aims and objectives. These procedures can be formalised in manuals and trainings, but more often they are established in diversion from or in the absence of concrete, pre-defined guidelines. Importantly, ways of how technologies are used shapes procedures and thereby contribute to defining how to proceed appropriately.

Conceptually, legal and ethical types of norms are thought to emerge in structured, deliberative, reflective and even formalised processes, in communication and negotiation with others. Taken together, these norms can establish a specific norm-based order. Examples are the regulation and normative rejection of chemical and nuclear weapons in terms of a 'weapons taboo' (Price, 1995; Tannenwald, 1999). These norms shape what is normal (use of force) in international relations.

But the effects that perceived standards of appropriateness in a procedural-functional sense can have for what passes for normal use of force and how such understandings can eventually also shape even legal and ethical norms requires further scrutiny. Do functional necessities and procedural structures largely replace political deliberation? In theoretical terms, we would expect that a pre-defined, politically deliberated norm is implemented in actions like the violent use of force. In this sense, normativity would guide

Global Policy (2019)

the ways in which actors try to shape what is 'normal' in international relations, security policy, or warfare. Reversing this perspective, understandings of what is normal in international relations or warfare could also inform and influence the normative dimension when practices informed by functional-procedural considerations become widespread. The availability of technology influences this – not in the sense of technological determination but rather by creating conditions of possibility for certain actions and practices.

Moreover, the increasing role of AI-driven solutions such as machine-learning challenges human control and the ability to 'doubt' (Amoore, 2018) in the context of human-machine interactions. In other words, reality is always represented or reconstructed in human-machine interfaces – more often in terms of a transformation than as a linear, direct translation. This refers again to the fundamental problem of MHC and how it could be realised in practice. In terms of current and future developments, the explicability of algorithms is a major concern, referring to the problem that it is no longer possible for humans to retrace and understand how an algorithm has 'decided', ultimately making it a 'black box' (see Zarsky, 2016). The aforementioned AI Next campaign of the US military, for example, 'aims to enable AI systems to explain their actions, and to acquire and reason with common sense knowledge' (DARPA, 2018). While explicability is one of the central problems of 'autonomy', my argument in this article is broader, emphasising the way technology influences human decisions. This is important in the context of my conceptualisation of procedural norms because the weakening of human control will also mean that norms are increasingly less the outcome of human deliberation. Furthermore, it is important to highlight how human decisions are compromised, underlining that the degree of autonomy is less important than the question how human-machine interactions take place.

3. Transmitting reality: functional appropriateness and procedural norms

Research on remote-warfare is dominated by the growing importance of armed drones, particularly in the US-led 'war on terror'. While questions of international law are also predominant in this research field, theoretically motivated studies have also investigated the effect of drone warfare on concepts of the use of force (Kaag and Kreps, 2014; Sauer and Schörnig, 2012; Warren and Bode, 2015) and assessed technological features as well as political and sociological effects (Schwarz, 2016; Walters, 2014). However, drones are only part of a broader security network, comprising highly interrelated actors and technologies. This is important because it shows to what extent extensive human-machine interactions are contributing to drone-strikes as the endpoint of using force. In the context of AWS and increasing autonomy, this raises the question of how human control relates to multi-dimensional practices but also how norms of procedural appropriateness emerge here that define what appropriate human control is, for instance. Research on the 'targeting cycle' of fighter jets, for example, suggests that

human control is deeply compromised and distributed in contemporary weapons systems (Ekelhof, 2018). The 'human in the loop' holding total control over machines is therefore arguably 'an impossible figure' (Amoore, 2018, p. 9).

With regard to the broad arrangement of using force, I emphasise the importance of data collection and target identification, which is at present already strongly shaped by the use of AI-driven features such as machine-learning. Examples are the US activities in installing an extensive surveillance regime, partly using drones, partly satellites and other relevant technologies. Revelations in recent years, for example on the National Security Agency (NSA) PRISM programme, point to the excessive information gathering of US intelligence services on millions of users worldwide (Sottek, 2013). In the military context, the collected 'metadata' is a key element of surveillance and target identification, used inter alia in 'signature strikes', that is, strikes that target people based on patterned of supposedly 'suspicious' behaviour rather than identified individuals. Former Assistant Secretary for Homeland Security Policy Stewart Baker emphasised that 'metadata absolutely tells you everything about somebody's life. If you have enough metadata, you do not really need content', while General Michael Hayden, former director of the NSA and the CIA underline that '[w]e kill people based on metadata' (Cole, 2014).

A concrete example is the US surveillance programme SKYNET that shows how target identification is delegated to machine-learning solutions able to process a vast amount of data and identifying unusual patterns, in this case of 55 million mobile phone users in Pakistan (The Intercept, 2015). The aim of this programme segment was to develop an understanding of what constitutes 'normal' behaviour and is based on machine-learning solutions. In this case, normality is defined by the behaviour of most people – or rather of data points showing specific movement patterns. SKYNET contributed to large-scale surveillance operations executed by drones and was meant to inform the US targeted-killing practice. In a widely reported case, SKYNET identified the Al Jazeera journalist Ahmad Zaidan as a possible member of al-Qaeda and the Muslim Brotherhood, based on his travel patterns and mobile phone call logs, which were, however, linked to his investigative work in Pakistan (Parvaz, 2017). The decisive point is that machine-learning solutions develop an ad hoc understanding of anomalies (see Aradau and Blanke, 2018), which can have so far neglected political and normative implications. What counts as a legitimate target is not predefined but delegated to an algorithm that scans for 'anomalous' human behaviour.

'The politics of the list' (de Goede et al., 2016), whether this refers to 'kill-lists' or targeted sanctions based on surveillance metadata, points to the importance of how processing information is part of coercion and the use of force. With regard to the topic of this article, the decisive question is to what extent political deliberation is increasingly influenced or even replaced by technical procedures comprising the development of technical benchmarks to ultimately decide about life and death.

This is not a completely novel problem: a 'friendly fire' incident during the Iraq war in 2003, involving a US Patriot missile defence battery and a British Tornado fighter jet, showed that MHC was dependent on the accuracy of electronically processed information. The official inquiry noted that 'Patriot crews are trained to react quickly, engage early and to trust the Patriot system (...) The crew had about one minute to decide whether to engage. The crew was fully trained, but their training had focused on recognising generic threats rather than on those that were specific to Iraq or on identifying false alarms' (Ministry of Defence, 2004, p. 3). However, this incident did not lead to abandoning Patriot systems nor to a general contestation of technological autonomy. Moreover, the Patriot crew was cleared of any wrongdoing because they followed procedures. This means that a certain procedural norm of what is acceptable/appropriate human control has emerged, although lethal incidents could not be ruled out completely.

The important point is that procedural standards of appropriate action – defining the normality of warfare – can diffuse into the dimension of legality and normativity. When calculations of statistical anomaly (like in the case of SKYNET) turn into definitions of normal and abnormal human behaviour, a flexible benchmark rather than a fixed political norm defines what the appropriate use of force is in practice. Human control presupposing a definition of appropriate action – for example identifying a potential target and how to engage this target – is increasingly under pressure when algorithmic parameters for selecting the 'best' target are unclear or speed and quantity of data surpasses the cognitive abilities of humans. The further advancement of autonomous features in weapons systems implies the risk that norms governing the use of force are not only executed but increasingly created by machines. The influence of the human actor responsible for implementing decisions taken at a different level slowly recedes in the use of force sequence. The inability and partly unwillingness of the political debate to define key aspects such as autonomy in weapons systems or MHC in the framework of the UN-CCW – besides the multi-dimensional character of 'using force' – makes the adoption of a comprehensive regulation or ban of AI-driven and informed security policy difficult. This underlines the utmost importance of focusing on the norm-setting effects of practices comprising human-machine interactions. The outlined procedural norms are powerful compared to vague, abstract and distant legal and ethical norms because they are the most important practical measure to ensure 'appropriate' actions.

4. Conclusion: AI in weapons technology and the use of force

The implications of the growing significance of AI-driven features in weapons system and in processes of using force are important but understudied. This paper focused on how human-machine interactions influence and limit human control, leading to functionally defined norms of what appropriate use of force is. Practices of using force are increasingly influenced by technologically operated processes that compromise meaningful human control. These processes are creating a normality of what it means to use force in specific contexts. While the direct interaction of humans and weapons systems already creates understandings of what acceptable human control is – for example in the case of Patriot systems or armed drones – the use of force is a broader process and the analysis should therefore consider different dimensions that are relevant for human control. Particularly, current practices of surveillance and data collection such as PRISM or SKYNET show to what extent AI-driven applications already interfere in decisions of what constitutes abnormal behaviour or a legitimate target.

If we consider meaningful human control as a norm to be defined on the level of political deliberation and meant to satisfy requirements of accountability and responsibility, the question arises to what extent this norm will by shaped by the normality of what meaningful human control means in the practice of using force. The normality of the use of force is and will be defined in the military context and technologies influence this definition to a large extent. This suggests that the political definition of the desirable normal risks being replaced by procedures, benchmarks, and machinic agency. The existing lack of knowledge and awareness contests the political control of weaponised AI, leaving it to technical procedures to provide the required levels of accountability and responsibility.

Being aware of the outlined implications of AI in security and defence systems is a first step towards a possible political response to these issues. The more control shifts from humans to machines in terms of algorithms or machine-learning, the less the definition of what use of force ought to be in practices is still subject to legal-political accountability and authority. In this regard, the perspective on norms presented in this article shows why the growing relevance of AI in weapons systems should be of great concern.

References

Amoore, L. (2018) Doubtful Algorithms: Of Machine Learning Truths and Partial Accounts. *Theory, Culture & Society*, pp. 1–25.

Aradau, C. and Blanke, T. (2018) 'Governing others: Anomaly and the Algorithmic Subject of Security', *European Journal of International Security*, 3 (1), pp. 1–21.

Asaro, P. (2012) 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making', *International Review of the Red Cross*, 94 (886), pp. 687–709.

Bode, I. and Huelss, H. (2018) 'Autonomous Weapons Systems and Changing Norms in International Relations', *Review of International Studies*, 44 (3), pp. 393–413.

Campaign To Stop Killer Robots (2018) Country Views on Killer Robots. 22 November 2018. Available from: https://www.stopkillerrobots. org/wpcontent/uploads/2018/11/KRC_CountryViews22Nov2018.pdf. [Accessed 10 January 2019].

Checkel, J. T. (1998) 'The Constructivist Turn in International Relations Theory', in. M. Finnemore, P. Katzenstein and A. Klotz (eds.), *World Politics* 50 (2), pp. 324–348.

Cole, D. (2014) 'We Kill People Based on Metadata'. In: *The New York Review of Books*. New york. Available from: https://www.nybooks.c

om/daily/2014/05/10/we-kill-people-based-metadata/ [Accessed 26 February 2019].

Crootof, R. (2016) 'A Meaningful Floor for "Meaningful Human Control"'. *Temple International & Comparative Law Journal*, 30 (1), pp. 53–62.

DARPA (2018) *AI Next Campaign*. Available at: https://www.darpa.mil/work-with-us/ai-next-campaign [accessed 17 December 2018].

Ekelhof, M. A. C. (2018) 'Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting'', *Naval War College Review*, 71 (6), pp. 1–34. Available from: http://digital-commons.usnwc.edu/nwc-review/vol71/iss3/6.

de Goede, M., Leander, A. and Sullivan, G. (2016) 'Introduction: The Politics of the List', *Environment and Planning D: Society and Space*, 34 (1), 3–13.

Grut, C. (2013) 'The Challenge of Autonomous Lethal Robotics to International Humanitarian Law', *Journal of Conflict and Security Law*, 18 (1), 5–23.

Heyns, C. (2016) 'Autonomous Weapons Systems: Living a Dignified Life and Dying a Dignified Death', in N. Bhuta, S. Beck and R. Geiss, H.-Y. Liu and C. Kress (eds) *Autonomous Weapons Systems. Law, Ethics, Policy*. Cambridge: Cambridge University Press, pp. 3–20.

ICRC (2016) Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons. Available from: https://www.icrc.org/en/publication/4283-autonomous-weapons-systems [accessed 22 February 2019].

Johnson, A. M. and Axinn, S. (2013) 'The Morality of Autonomous Robots', *Journal of Military Ethics*, 12 (2), 129–141.

Kaag, J. and Kreps, S. (2014) *Drone Warfare*. Cambridge: Polity Press.

Kastan, B. (2013) 'Autonomous Weapons Systems: A Coming Legal Singularity', *Journal of Law, Technology & Policy*, 45 (1), 45–82.

Klotz, A. (1999) *Norms in International Relations: The Struggle against Apartheid*. Ithaca, NY: Cornell University Press.

Koskenniemi, M. (2011) *The Politics of International Law*. Oxford: Hart.

Kratochwil, F. V. (1989) *Rules, Norms, and Decisions*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511559044.

Leveringhaus, A. (2016) *Ethics and Autonomous Weapons*. London: Palgrave Macmillan.

Ministry of Defence (2004) Aircraft Accident to Royal Air Force Tornado GR MK4A ZG710. Available from: https://www.gov.uk/government/publications/military-aircraft-accident-summary-aircraft-accident-to-raf-tornado-gr-mk4a-zg710 [Accessed 15 February 2019].

Moyes, R. (2016) 'Meaningful Human Control', in R. Geiss and H. Lahmann (eds.), *Lethal Autononous Weapons Systems. Technology, Definition, Ethics, Law & Security*. Berlin: Federal Foreign Office, pp. 239–249.

Noone, G. P. and Noone, D. C. (2015) The Debate Over Autonomous Weapons Systems., *Case Western Reserve Journal of International Law*, 47 (1), 25–35.

Parvaz, D. (2017) Journalists Allege Threat of Drone Execution by US. Available from: https://www.aljazeera.com/news/2017/03/journalists-allege-threat-drone-execution-170330205806045.html [Accessed 13 September 2018].

Price, R. (1995) 'A Genealogy of the Chemical Weapons Taboo', *International Organization*, 49 (1), 73–103.

Roff, H. M. and Moyes, R. (2016) Meaningful Human Control, Artificial Intelligence and Autonomous Weapons. Briefing paper prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems. UN Convention on Certain Conventional Weapons.

Sauer, F. and Schörnig, N. (2012) 'Killer Drones: The 'Silver Bullet' of Democratic Warfare?', *Security Dialogue*, 43 (4), 363–380.

Schwarz, E. (2016) 'Prescription Drones: On the Techno-biopolitical Regimes of Contemporary "ethical killing"', *Security Dialogue*, 47 (1), 59–75.

Sehrawat, V. (2017) 'Autonomous Weapon System: Law of Armed Conflict (LOAC) and Other Legal Challenges', *Computer Law & Security Review*, 33 (1), 38–56.

Sharkey, N. (2008) 'The Ethical Frontiers of Robotics', *Science*, 322 (5909), 1800–1801.

Sottek, T. C. (2013) Everything You Need to Know about PRISM. Available from: https://www.theverge.com/2013/7/17/4517480/nsa-spying-prism-surveillance-cheat-sheet [Accessed 26 February 2019].

Tannenwald, N. (1999) 'The Nuclear Taboo : The United States Basis of and the Normative Nuclear Non-use', *International Organization*, 53 (3), 433–468.

The Intercept (2015) SKYNET: Courier Detection via Machine Learning - The Intercept. Available from: https://theintercept.com/document/2015/05/08/skynet-courier/[Accessed 3 November 2017].

Walters, W. (2014) 'Drone Strikes, Dingpolitik and Beyond: Furthering the Debate on Materiality and Security', *Security Dialogue*, 45 (2), 101–118.

Warren, A. and Bode, I. (2015) 'Altering the Playing Field: The U.S. Redefinition of the Use-of-force', *Contemporary Security Policy*, 36 (2), 174–199.

Zarsky, T. (2016) 'The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making', *Science, Technology & Human Values*, 41 (1), 118–132.

Author Information

Hendrik Huelss is a postdoctoral researcher in the School of Politics and International Relations, University of Kent. He currently focuses on norms and the weaponization of AI at the intersection of International Relations theory and International Political Sociology.