

Data, Speed, and Know-How: Ethical and Philosophical Issues in Human-Autonomous Systems Cooperation in Military Contexts

Mark Coeckelbergh^(✉) and Michael Funk

Department of Philosophy, University of Vienna, Vienna, Austria
{mark.coeckelbergh,michael.funk}@univie.ac.at

Abstract. Human-Autonomous Systems Cooperation raises several ethical and philosophical issues that need to be addressed not only at the stage of implementation of the system but also preferably at the stage of development. This paper identifies and discusses some of these issues, with a specific focus on human-machine cooperation problems and chances, focusing usage of these systems in military contexts. It is argued that ethical, philosophical, and technical problems include (1) data security and monitoring/management, (2) agency, distancing and speed/time, and (3) cooperation, networks and knowledge. These issues need to be taken into account not only in the application but also in processes of research and development and legal regulation.

Keywords: Human-machine cooperation · Ethics · Drones · Cyberwar · HFT · Speed · Time · Knowledge · Networks

1 Introduction

When technologies are no longer mere tools or extensions of the human, but acquire a high degree of “agency” themselves, different concepts have to be used to describe the relation between humans and machines. Systems such as autopilots or autonomous robots are no longer simply tools or even extensions of the (human) body; instead there is a relation that may be described by using the terms “cyborg” (merging of human and machine) or “cooperation” between human and non-human agents (cooperation between different agents, no merging). These new configurations raise several ethical and philosophical questions, such as: “Who is responsible for unintended damage, caused by an autonomous system?” This paper mainly focuses its discussion on cooperation problems in a military context. However, self-driving cars and traffic guidance systems, financial technologies of high speed trading (HFT = “High Frequency Trading”), as well as autopilots and drones (UAVs = “Unmanned Aerial Vehicles”) will also be discussed as ethically challenging forms of autonomous systems which may lead to military applications.

2 Data Security and Monitoring/Management

A first ethical issue concerns the security of the data, given that autonomous systems are not only autonomous agents, but also smart devices – embedded into IT-networks, linked to computers, databases etc. Autonomous systems collect an enormous amount of data, which raises the question (firstly) who has access to it, or could potentially force access to it, and (secondly) what can be done with that information. Who is responsible for the interpretation of images or semantic information? E.g. in drone-operations target-profiles play a major role. Who creates such profiles and will be held responsible, if collateral damage or illegal actions are performed on the basis of deficient target-profiles? Aspects of dual use and collateral damage – even friendly fire – include a cyber-component. In military contexts there clearly is this danger. Consider for instance the case of a US-American RQ-170 Sentinel drone “hacked” down by Iranian organs in December 2011 [1]. It is important to deal with these data-vulnerabilities and risks, especially since human lives are at stake. How to protect data and defense infrastructures? The network-factor and connection of technologies (especially IT) causes strong challenges here. Should we disconnect critical infrastructure? What about care-robots or medical bots which collect a lot sensitive data (in order to function successfully)? In times of international terrorism and IT-based warfare, those “civil” structures might become targets of “military” hits as well.

Conceptual understandings of “new wars” vs. “old wars” or “symmetric warfare” vs. “asymmetric warfare” after the collapse of Soviet Union and 1990th Yugoslav Wars have been emphasized primarily in political and social terms by Kaldor [2] and Münkler [3]. But cyberwar [4, 5] once more became a new technological category of warfare and cyber can be seen as new strategic operation-sphere such as “land”, “air”, “water”, “underwater” or “space”. At the same time, “cyber” affects all classical spheres, as information technologies are used in nearly every situation. When the line between conventional and cyber warfare becomes blurred by means of these new technologies – and as even the classical categories of combatant and non-combatant become blurry as well – military organizations and politicians have to rethink how they invest and organize defense infrastructures and capacities – both in civil as well as in military situations. As one consequence of IT-developments, even the meaning of “military context”, becomes questionable as well, given the use and development of IT across several domains. The application of data and information (not only semantics but also images etc.) causes new forms of actions such as more sublime forms of propaganda or blackmailing. Hackers without any classical military education could perform cyber-attacks and cause economic, psychological (especially in case of terrorism) or material damage. Where is the borderline between combatant-hackers and non-combatant-hackers? In military contexts, privacy plays another role than in civil life. But if the categories civil vs. military are changing, the limits of privacy in certain contexts of usage might also change. Ethics of data-security and information-protection in military contexts involves the problem that the values of security and safety are in tension with the values of data-security and privacy. The more secure and safe a system or society is intended to be, the more information is needed in order to ensure success. There is thus an ethical conflict between public security on the one side and data security, data privacy and information protection on the other side, e.g.:

- protect society against attacks of other nations, including espionage,
- protect society against terrorist attacks, or
- protect society against crime.

Those cases include the protection against attacks which come out of the own society (whether terroristic, criminal etc. motivated). Here we might find arguments for supporting a more strict form of surveillance of public life. But on the other hand free access to information and privacy are important moral values, which not only cause liberal political life, but also economic success and creativity (in order to create innovations).

Note that the same dilemma is characteristic for usage of data for the monitoring and management of people from one's own organization as well. Does monitoring of personnel via these technologies enhance responsible and effective behavior, or does it potentially create unworkable situations for people who are already under pressure? Putting in other words: Efficiency or Burnout? Does the flow of data enhance the safety and security of people? Or could it lead to decisions taken at a distance on the basis of data that may not correspond to the local situated and experienced knowledge of the people who use the technology? This raises knowledge problems (see below, part four). Anyway historical experiences (such as the economic collapse of GDR in 1989/90) illustrated: STASI-like surveillance diminished economic and technological efficiency. Human creativity is not such easily accessible for data-monitoring.

3 Agency, Distancing, and Speed

A second ethical issue concerns the agency of technologies and, related to that in a military context, the problem of speed. If for instance a robot is given more autonomy, technically speaking (acting on its own), decisions have to be made with regard to the degree of supervision of the robot. There are good ethical reasons to keep humans in the loop if the machine has lethal capacities (see also [6]), to ensure human judgment and human responsibility. We like to support the argument, that even so called "autonomous" systems remain means for human ends: "people kill people". However, it is a realistic scenario that if, in a military context, many autonomous systems are "employed"/applied and there might be – metaphorically speaking – an "arms race", the speed of war will increase to such a point that perhaps only non-human, autonomous systems can process what is going on and therefore make decisions. This is ethically problematic if one assumes that human judgment should be involved in actions that may have lethal consequences (for people of the own organization or for others). A ban on such systems may prevent this, but if such a ban were not very effective (which is also a realistic scenario), how can and should military organizations and societies deal with this problem?

A related problem raised by the new information technologies concerns distance as well. For instance drones seem to make the decision of lethal actions easier, but at the same time also create new forms of proximity or even intimacy that may make such actions and decisions less easy [7]. The point is that drone-operators sit thousands of kilometers away from the target, only present on a medial level with real-time images.

That might make killing easier. But on the other hand, drone-operators observe their victims a long time before starting a lethal strike. Here, the real-time pictures somehow bridge the physical distance and enable something like proximity or intimacy. There are also similar problems with technologies in civil contexts. Financial technologies [8] and high-speed trading can be seen as pre-forms or already existing forms of autonomous systems and real-time decision making under conditions that nearly exclude humans given the speed. There are paradoxical effects that have to do with the relation between speed and distance. Pre-computer and historical distance weapons (such as archery, muskets or ballistic artillery) are bound to a physical sphere where speed, time and the location as well as distance have been linked to each other: distance means a temporal delay of information, speed means higher distance of a projectile etc. But with computers and drones (again information technologies) it became possible to generate real-time images at the scale of global distance. Cyber-attacks and high frequency algorithm based counter-attacks might also be realized in real-time on a transcontinental scale. A physical link between time and space is suspended: in consequence of IT-embedded warfare, physical distance no longer means safety [9].

Again ethical problems enter the stage when it comes to decision making: automatized cyber-attacks in real time are situated in a temporal microcosm, in which the human bodily-sensory temporal mesocosm of ethical reflection or legal judgment is suspended. How can the problem of this gap between two worlds be addressed? Thus, here is an ethical conflict between the fact that the IT-basis of drones or cyber-operations enables real time decision making, but that human assessment and responsibility requires keeping temporal distance for reflection and ethical, legal or political evaluation [10]. What can be the solution? Slowing down the speed of warfare? Mechanical typewriters instead of PCs?

To further reflect on this problem concerning speed and distance, we propose to use the work of Virillo [11, 12] who offers a theory of ‘dromology’ (societal impacts of speed and acceleration) and also presents a paradoxical interrelation between acceleration and deceleration at the same time. Inspired by his theory, once more we want to argue that information technological networks as basis for autonomous military technologies cause an accelerated cyber-sphere (temporal microcosm) which is separated from the temporal sphere of political, ethical or juridical decision making (temporal mesocosm).

To clarify more our understanding of “temporal microcosm”, we like to draw the analogy to technical mediations and visual perception. Microscopes are tools which reveal visual microworlds: e.g. perception of bacteria, cells or even smaller things. In 17th- and 18th-century sciences Antoni van Leeuwenhoek became one prominent pioneer in the field of microscopy. Even before that in the 16th- and 17th-century developments, scientists like Galileo Galilei, Johannes Kepler or Tycho Brahe started systematically using telescopes for empirical astronomic observations. Microscopes technologically reveal visual microworlds (very small), whereas telescopes enable perceptual access to visual macroworlds (very big). Glasses are examples for technologies applied in human mesocosm, because they should not translate very small worlds (microcosm) or very large worlds (macrocosm) into a human sensory scale. Instead glasses are used in order to mediate mesoworlds, which means not enhancing visual capabilities, but moreover compensate visual disabilities without leaving human

sensory range of perception. Now the argument in context of time and IT-embedded warfare is that information technologies are related to very small worlds, not on the visual, but on the temporal level. The difference to microscopes here is accessibility. Whereas microscopes make invisible things visible, high frequency IT-embedded processes disable access because they do not bridge but cause a gap between speed and human capabilities of decision making.

There is also a meta-ethical issue concerning temporal forms of future and present: every ethics is linked to a concrete future as ethics always emphasizes following human actions, their consequences, conditions and values. When high-speed-wars are functionally performed in a temporal microcosm that appears to humans as “present” (we are too slowly, we cannot perceive it), what does this mean for ethics and its innate link to a concrete “future”? How is future (and past) possible in real time-actions? Future is a human-pragmatic term, which loses its meaning within real-time warfare (or high-speed trading etc.). Insofar IT-high-speed warfare could lead to the paradoxical consequence that the classical Clausewitz slogan “War is the continuation of politics by other means” [13] becomes useless, as means do not longer support a political aim, but disable political decision making in an unintended way. Three temporal logics of warfare and politics can be distinguished for illustrating this development:

- 19th-century understanding of warfare following the theoretical approach of Carl Philipp Gottlieb von Clausewitz: no nukes or ICT (bodily-mechanic temporal mesocosm, physical and physiological linking between space and time).
- 20th-century understanding of warfare including capabilities of nuclear weapons, ICBM and submarines: if nuclear strike starts, there will be enough time (minutes) for a counterstrike (still temporal mesocosm), in relation to the danger of destroying whole mankind and thereby destroying any political future on the one hand, but nuclear standoff became an element of political practice on the other hand as well.
- 21st-century ICT-related warfare: a temporal microcosm logically suspends politics as high-speed processes do not allow human decision making, fully autonomous real-time warfare might not necessarily destroy mankind as such, but “war” loses its meaning as means for political ends (political instrumental rationality is suspended) [10].

Deceleration in military contexts is often related to defensive situations (castles, partisan tactics etc.) or caused by economic benefits of “new wars” [3]. We think in consequence of the temporal microcosm deceleration becomes a legal and ethical demand as well.

4 Human-Machine Collaboration and Knowledge

A third, related and ethically relevant but also technical and philosophical issue concerns problems regarding human-machine collaboration. Even if the above mentioned problems could be addressed in a satisfactory way, decisions have to be made concerning the division of work and the distribution of knowledge (and responsibility) in hybrid human/machine systems. Such problems are analogous to those with (civil and military) airplanes. For instance, if there is an emergency, should the autonomous

system be given priority, or should humans decide, even if they may be wrong? The same questions play a crucial role in the context of self-driving cars and traffic guidance systems, where similar decisions have to be made. This raises the question what kind of knowledge autonomous systems (computing systems, robots, etc.) have, as opposed to human beings, and in which way knowledge is distributed in cooperative contexts and embedded in networks. Knowledge is needed in a situation for decision making and for acting responsibly. But how exactly do the new technologies change the knowledge configuration? Human beings interpret data, but autonomous systems also do so – based on code written by humans which enable the system to model what is going on in its environment. However, the way humans arrive at a judgment tends to be different, since humans are embodied, have emotions, and so on. They can also improvise, have a different kind of know-how. Some argue that therefore the autonomous systems should be given the final decision (to avoid emotional judgment, war crime etc.), whereas others think that humans should be given control since the kind of knowledge autonomous systems have is more limited. But what if autonomous systems have access to much more data than humans (“big data”)? Or is this not the same as having knowledge? What is the difference between data, information, and knowledge?

Moreover, how universal is the way we arrive at knowledge? Does the development of autonomous systems currently take into account cultural differences? What is the knowledge-base for ethical judgments? And what kind of weapons give rise to what kind of knowledge? Compare for instance traditional Japanese or Chinese theories of war- and martial-arts with current IT-embedded warfare. Ancient sword-fights require deeply embodied knowledge, a high level of sensory skills, fast perception (into the human scale, temporal mesocosm), but also tactical knowledge. Japanese approaches of samurai sword fights [14, 15] include a strong ethos of social relations between samurai, sword, enemy, and master (something like a “feudal lord” who is the commander). Samurai sword fight theory is pretty individually oriented, because the warrior and its embodied tacit knowledge of using the sword are the conceptual epicenter of this kind of warfare. On the other hand tactical and strategic knowledge including a lot of bodily elements is integrated in the ancient Chinese approach as well [16]. But in this case not so much the individual swordfighter, but moreover the general and the way he leads troops in several situations is emphasized from a primary point of view. Again the tactical and strategic understanding of how to use geographical factors, time and rhythm of strike and counter-strike, defense and offense movements, or integration of spies is related to bodily and physiological temporal and spatial mesoworlds. The physical and bodily link between time and space remains an important axiom of those theories of warfare [10]. Insofar implicit knowledge or tacit knowledge, which needs to be generated within processes of trial and error, and is more than theoretical textbook-information, serves as the primary knowledgebase.

But what kind of knowledge is involved in (big) data analysis, computer modeling, etc.? Is this disembodied knowledge, and if so in what sense? If not, how does the embodied relation, mediated by the new technologies, differ from earlier technologically mediated ways of fighting? Also the issue of distance plays a crucial role here (see also [6, 7]). Can we wield a drone with a distance of 10.000 km such as a sword or gun in our hands? Can real-time computer pictures and joysticks in an isolated container replace both the whole-body-perception and -knowledge of a regular soldier in a

battlefield? What does this mean for the attribution of responsibility and the ways we know how to detect ethical conflicts and to deal with it? Fact is that skills, emotions, intuition or sensory perceptions belong to a different domain of knowledge than computer-algorithms [17–20].

As we have seen the relations and meaning of time and space are changing. Bodily spheres of spatial-temporal knowing are replaced by cyber spheres of real-time actions. Those actions are enabled by a temporal microcosm. It is both an ethical and epistemological task to understand the structures of this new spatial-temporal form of warfare. For ethics and political philosophy applied to this problem, it might be the primary task to identify the exact ethical limits of autonomous decision-making. A gap of knowledge and a lack of time for assessment could cause inhumane consequences, collateral-damage or maybe terminator-scenarios etc. There is a need for legal regulation of new warfare and its technological structures, and for societal scenarios that explore the many ways we use and not-use those new possibilities. And, finally, these new technologies remain challenging for thinking about ethics. What is ethical decision-making, and can and should it be delegated to machines?

It may also be helpful to try to answer these questions by engaging with frameworks that are being developed in computer science and engineering. For instance, Modelling and Simulation (M&S) might help to answer the previous questions regarding knowledge involved in computer modeling. And when thinking about ethics one could look at M&S experimental frameworks build upon common vocabularies in the form of Autonomous System and M&S ontologies that aim to cover the ethical aspects [21, 22].

5 Conclusion

This paper has identified a number of ethical and philosophical problems raised by new developments in the area of autonomous systems and their actual or potential use in a military context. Particular attention has been paid to issues related to security of data flows, agency and speed/distance, and knowledge in cooperative human/autonomous system configurations. We identified ethical problems and conflicts with regard to security, speed, and cooperation, and also pointed to cultural differences with regard to knowledge in ethics and related to military technologies. It is important to take these problems into account not only when using but also when developing autonomous systems for defense/warfare. More research is needed on the (potential) impact of these systems on ethics and responsibility, and ultimately on the forms of warfare and the kinds of societies we will have in the future.

References

1. Biermann, K., Wiegold, T.: *Drohnen. Chancen und Gefahren einer neuen Technik*. Ch. Links Verlag, Berlin (2015)
2. Kaldor, M.: *Neue und alte Kriege. Organisierte Gewalt im Zeitalter der Globalisierung*. Suhrkamp, Frankfurt a.M. (2007)

3. Münkler, H.: *Der Wandel des Krieges. Von der Symmetrie zur Asymmetrie*. 3. Auflage. Velbrück Wissenschaft, Weilerswist (2014)
4. Gaycken, S.: *Jenseits von 1984. Datenschutz und Überwachung in der fortgeschrittenen Informationsgesellschaft. Eine Versachlichung*. Transcript Verlag, Bielefeld (2012)
5. Gaycken, S.: *Cyberwar: Das Wettrüsten hat längst begonnen*. Goldmann Verlag, München (2012)
6. Coeckelbergh, M.: Drones, morality, and vulnerability: two arguments against automated killing. In: Custers, B. (ed.) *The Future of Drone Use. Opportunities and Threats from Ethical and Legal Perspectives*. T.M.C. Asser Press, Hague (2016, forthcoming)
7. Coeckelbergh, M.: Drones, information technology, and distance: mapping the moral epistemology of remote fighting. *Ethics Inf. Technol.* **15**(2), 87–98 (2013)
8. Coeckelbergh, M.: *Money Machines. Electronic Financial Technologies, Distancing, and Responsibility in Global Finance*. Ashgate, Farnham (2015)
9. Funk, M.: Drohnen und sogenannte ‘autonom-intelligente’ Technik im Kriegseinsatz. Philosophische und ethische Fragestellungen. In: Funk, M., Leuteritz, S., Irrgang, B. (eds.) *Cyberwar @ Drohnenkrieg. Neue Kriegstechnologien philosophisch betrachtet*. Königshausen & Neumann, Würzburg (2016, forthcoming)
10. Funk, M.: Zeit als Element technologischer Kriegsführung. In: Funk, M., Leuteritz, S., Irrgang, B. (eds.) *Cyberwar @ Drohnenkrieg. Neue Kriegstechnologien philosophisch betrachtet*. Königshausen & Neumann, Würzburg (2016, forthcoming)
11. Virillo, P.: *Speed and Politics: An Essay on Dromology*. Semiotext(e), New York (1977/1986)
12. Virillo, P.: *Negative Horizon: An Essay in Dromoscopy*. Continuum, London (1989)
13. Clausewitz, C.V.: *Vom Kriege*. Reclam, Stuttgart (1994)
14. Musashi, M.: *Fünf Ringe. Die Kunst des Samurai-Schwertweges*. Trans. by Siegfried Schaarschmidt. Nikol Verlag, Hamburg (2008)
15. Yamamoto, J.: *Hagakure*. Ed. by Tsuramoto Tashiro. Trans. by Max Seinsch. Reclam, Stuttgart (2009)
16. Sunzi: *Die Kunst des Krieges*. Trans. by Dr. Hannelore Eisenhofer. Nikol Verlag, Hamburg (2011)
17. Polanyi, M.: *Personal Knowledge: Towards a Post-Critical Philosophy*. University of Chicago Press, Chicago (1958)
18. Dreyfus, H.: *What Computers Can’t Do: The Limits of Artificial Intelligence*. MIT Press, New York (1972)
19. Ferguson, E.S.: *Engineering and the Mind’s Eye*. MIT Press, New York (1992)
20. Funk, M., Coeckelbergh, M.: Is gesture knowledge? A philosophical approach to the epistemology of musical gestures. In: De Preester, H. (ed.) *Moving Imagination – Explorations of Gesture and Inner Movement in the Arts*, pp. 113–132. John Benjamins Publishing Company, Amsterdam (2013)
21. Hodicky, J.: HLA as an experimental backbone for autonomous system integration into operational field. In: Hodicky, J. (ed.) *MESAS 2014. LNCS*, vol. 8906, pp. 121–126. Springer, Heidelberg (2014)
22. Hodicky, J.: Modelling and simulation in the autonomous systems’ domain-current status and way ahead. In: Hodicky, J. (ed.) *MESAS 2015. LNCS*, vol. 9055, pp. 17–23. Springer, Heidelberg (2015)

Modelling and Simulation for Autonomous Systems
Third International Workshop, MESAS 2016, Rome, Italy,
June 15-16, 2016, Revised Selected Papers
Hodicky, J. (Ed.)
2016, XVI, 408 p. 230 illus., Softcover
ISBN: 978-3-319-47604-9