Safeguarding moral perception and responsibility via a partnership approach

Tjerk DE GREEF [a], Alex LEVERINGHAUS [b]
[a] *Interactive Intelligence Group, Delft University of Technology*
[b] *Oxford Institute for Ethics Law and Armed Conflict & Oxford Martin School, University of Oxford*

ABSTRACT

Introduction: New Combat Technologies (NCTs) create unprecedented capabilities to control the delivery of military force. But dependence on them generates difficult moral challenges impacting the decision-making process, which are only beginning to be addressed. **Method**: Some welcome NCTs explicitly whereas others plea to abandon such technologies. In order to arrive at an informed opinion about the impact of NCTs on decision-making, we need to know more about what NCTs are and how they operate. We provide a short overview of the different types of NCTs and discuss the challenges posed to moral perception. **Results and discussion:** Before NCTs can be deployed, we need to rest assured that their usage enhances, rather than undermines, human decision-making capacities. We propose a design recommendation based on a partnership approach, ensuring that operators use appropriate information in the decision-making process. There are important choices to be made, and sound design is 'design for responsibility'.

KEYWORDS

Sensemaking; military; drones; unmanned systems; moral perception; responsibility; partnership.

INTRODUCTION

One issue within the discourse over military interventions (MI) concerns the impact of new combat technologies (NCTs) on the human decision-making process. NCTs include military robots, such as unmanned aerial vehicles (UAV), computer-based targeting systems and missile defence systems. Scholars from different disciplines expressed concerns regarding implication on the decision-making process due to the distance from the battlefield and increased autonomy (Sharkey, 2010; Sparrow, 2007), potentially leading to the abdication of responsibility (Matthias, 2004). Dependence on NCTs thus generates moral and legal challenges impacting the quality of the decision-making process, which are only beginning to be addressed.

NCTs have already been used during MI. Commenting on NATO's intervention in Kosovo in 1999, Ignatieff speaks of 'virtual war', where NATO forces, supported by NCTs and more traditional air power, did the fighting, but only 'Serbs and Kosovars did the dying' (Ignatieff, 2000). NATO's service personnel, Ignatieff shows, were removed from the actual combat zones, but with the help of technology could carry out military missions. More recently, warships shelled targets in Libya to assist rebel fighters in the overthrow of the regime. Like Kosovo, intervening forces relied heavily on NCTs. Modern military technologies thus render 'boots on the ground' unnecessary.

Some scholars have welcomed explicitly the use of NCTs during MI (Altman & Wellman, 2009; Strawser, 2013). Firstly, for reasons of proportionality, they argue, the use of NCTs is desirable. Using UAVs leads to less damage and destruction than a large-scale military operation with boots on the ground. Secondly, western states are under pressure to minimize casualties amongst their own service personnel. Indeed, one of the morally and politically attractive features of NCTs is their ability to protect the lives of service personnel (Strawser, 2010).

There have also been critical voices. Ignatieff seems sceptical about the prospects of 'virtual war', while Sparrow argues that NCTs must not be deployed because they undermine a commitment to responsibility in the armed forces (Sparrow, 2007). If Sparrow's criticism is true, NCTs must not be deployed in *any* conflict.

However, in order to arrive at an informed opinion about the impact of NCTs on decision-making, we need to know more about what NCTs are and how they operate. Research in this area is in its infancy, but is likely to become more prominent. In this paper, we discuss how NCTs impact on the moral perception of those who operate them. The issue of moral perception, in fact, is crucial for a commitment to responsibility. We begin with some brief comments on moral perception. Subsequently, we give a short overview of the different types of NCTs and discuss some of the challenges they pose to moral perception. We continue by making design recommendations for NCTs. There are important choices to be made, and sound design is always 'design for responsibility' – or so we shall argue.

H. Chaudet, L. Pellegrin & N. Bonnardel (Eds.). *Proceedings of the 11th International Conference on Naturalistic Decision Making (NDM 2013), Marseille, France, 21-24 May 2013.* Paris, France: Arpege Science Publishing. ISBN 979-10-92329-00-1

MORAL PERCEPTION

The concept of moral perception is central to hold individuals responsible for their actions. Since the Nuremberg Trials, it also plays a major role in international law. Roughly, the concept of moral perception refers to the knowledge of the relevant facts in a particular situation. In holding individuals responsible for their actions, we assume that they have acted with knowledge of morally relevant facts. Conversely, in order to be exculpated from wrongdoing, an individual has to prove that he could not have acquired knowledge of the relevant facts. Since Nuremberg, combatants need to at least meet the moral perception criterion in order to be exculpated from wrongdoing (May, 2005).

However, in light of the rise of NCTs, the way in which soldiers acquire knowledge of relevant moral facts is transformed. As Ignatieff's idea of 'virtual war' implies, NCTs introduce (amongst other things) an element of distance. By removing soldiers from the combat zone, NCTs may well affect their moral perception of a particular situation, making it more difficult to, say, discriminate combatants from non-combatants. Some UAVs, for instance, transmit images they record with their sensors to a video screen where operators view them. UAVs remove their pilots from the actual battlefield physically, but in doing so it may restrict their moral perception of the situation. Depending on the context, it might be more difficult to hold operators of UAVs responsible for, say, applying force to a target. One potential reason for this is that, in case force is applied to the wrong target, operators could argue that, due to the restrictions imposed by the technology, they did not have full situational awareness and should therefore be exculpated from any wrongdoing.

This possibility leads to two immediate requirements. Firstly and from a more technologically oriented perspective, engineers designing military equipment must be sensitive to how different types of technology impact on moral perception. That is to say, they must take into account how psychological factors impact information processing and shape the perception of morally relevant facts. Secondly and from a more legally and normatively oriented perspective, NCTs must be designed in order to minimize any distortions or unnecessary restrictions of their operators' moral perception.

Overall, sound design must always be design that enhances, rather than undermines, the preconditions for individual responsibility. Ensuring this is, in our view, one of the central moral obligations of engineers and designers. Before we provide an indication what design for responsibility may look like, we provide an overview of the different types of NCTs.

NEW COMBAT TECHNOLOGIES

Within the broad category of NCTs, it is important to distinguish between two main classes:

1. Remote-controlled systems are being deployed by militaries already. Originally designed for reconnaissance and surveillance missions, UAVs are today capable of carrying a payload, which is why they are used for the morally and legally controversial practice of targeted killings. UAVs are tele-operated (they are controlled by using information provided by a video-link) from a large distance while the vehicle itself is airborne in an area of interest in or near a conflict zone.
2. The most controversial dynamic within the development of NCTs relates to what we call operationally autonomous weapons delivery systems. We define operational autonomy as the capacity of a machine to carry out specific tasks without assistance from an operator. In the future, UAVs, for instance, are expected to become increasingly operationally autonomous. This reflects a wider trend within the military, namely to reduce interaction between humans and machines.

Both NCTs have in common that they increase the distance to the battlefield. This has benefits as well as downsides. First, as we already noted, the possibility of death or serious injury amongst service personnel decreases greatly. Secondly, given that they do not face an immediate threat to their safety, the stress soldiers experience in combat is diminished. Arguably, stress affects moral perception because it influences how humans interpret their environment and frame certain issues. To illustrate the point, consider the infamous My Lai massacre that occurred during the Vietnam War. Fearing that the inhabitants of the hamlet of My Lai were Vietcong guerrillas posing as civilians, American soldiers experienced high levels of stress and, as a result, failed to apply the discrimination criterion accurately. This led to one of the worst massacres in history. A decrease in stress, then, might lead to greater awareness as well as more accurate interpretations of morally relevant facts. If this is true, the impact of NCTs on moral perception seems positive, rather than negative. That said, the reduction of stress could have negative effects. While it is correct that too much stress diminishes decision-making capacities, low levels of stress can have positive effects on an operator's alertness. It has, for example, been demonstrated that boring work conditions impact negatively on decision-making and performance (Endsley & Kiris, 1995).

There is another worry about the impact of distance on the moral perception. The quality of information that operators receive from machines may be lower than a human 'first hand' account. Operators, for instance, may receive too little information. Machines might not transmit certain information that a human would have picked up. This problem is particularly acute in operationally autonomous machines where the interaction between operator and machines is decreased. Depending on their task, operationally autonomous machines may process and filter large amounts of information themselves before passing on information to an operator. On the other hand, NCTs might supply operators with too much information, which may be more acute in tele-operated than operationally autonomous NCTs. Tele-operated UAVs have many sensors and remain airborne for long periods. Processing the amount of information may be difficult for a single operator. In both cases – the undersupply and oversupply of information – it becomes difficult for operators to filter out morally relevant facts. As a result, the deployment of NCTs may undermine a commitment to individual responsibility.

If these observations are accurate, the introduction of NCTs could go either way. It could increase or decrease the ability of individuals to acquire morally relevant facts. This is why different design options for NCTs are important.

PARTNERSHIP

How can we ensure that individuals perceive the relevant facts in a give situation? Contemporary literature shows two approaches in the development of NCTs. The first, defended by roboticist Ronald Arkin (2010), contends that NCTs should be given full operational autonomy in order to prevent war crimes. According to Arkin, automated systems lack emotional components that prompt humans to commit war crimes. Although a fair argument, taking the human completely out of the loop leads to underutilization of human experiences and capabilities. It remains a fact that specific humans capabilities are, up to date, hard to replicate by artificial intelligence. While, for example, computers are really strong in executing many calculations on large datasets, they lack creativity or a capability to recognize patterns that humans easily recognize. To do so, humans use a variety of psychological tools such as the knowledge-based reasoning mechanism (Rasmussen, 1986), which allows the human to cope with novel and unexpected situations by using fundamental knowledge (e.g. principles, physical laws) that governs the specific domain. Today's artificial intelligence technologies fail to model precisely the set of tools that allows being creative and recognizing patterns, leading to a fundamental asymmetry between human and artificial agents. It is, for instance, difficult to see how a machine could interpret human behaviour to distinguish a combatant from non-combatant whereas military operators recognize patterns-of-life to make such discriminations. Arkin is thus right to point out that humans are bad at decision-making and interpreting complex information under stressful conditions but machines presently lack the reasoning capacities that allow interpreting complex situations.

Faced with this problem, it is a common sense response to 'team-up' humans and artificial agents by integrating them into a *joint cognitive system* (Hollnagel & Woods, 2005). The partnership approach represents the second approach in the development of NTCs, which allows utilizing the strengths of one to compensate for weaknesses of the other. This move potentially enhances situational understanding and subsequent decision-making, as well as moral perception in general. Just as human team members develop different perspectives on a situation, machines and humans may develop different perspectives on a situation. Operators can use the perspective provided by their machine to check if they are missing morally relevant facts. The machine may flag up aspects of a situation that the operator might have otherwise overlooked.

The partnership approach is not entirely new (cf. de Greef, Arciszewski, & Neerincx, 2010; Johnson et al., 2011). The latest generation of partnerships is facilitated by *working agreements* (Arciszewski, de Greef, & van Delft, 2009) leading to a fine-grained division of labour between the human and machine with regard to specific tasks; this allows the human to stay in firm control of those tasks and objects that are regarded important. A human may, for example, delegate identifying unambiguous airplanes to the artificial agent while remaining in control of the more cognitive demanding ambiguous objects that potentially are legitimate targets.

The concept of working agreements were validated recently (de Greef et al., 2010). The eight navy officers who participated highly appreciated the division of labour between human and artificial agents introduced by the working agreement, especially when decisions had to be made under pressure. In this experiment, a partnership prototype was compared to a more static version resembling today's combat management workstations aboard navy frigates. The officers liked the working agreements and were relieved that they could focus on the more demanding bits while having the machine do the regular 'low risk' easy bits. The performance effects revealed a clear preference towards the partnership approach in that the identification times (a measured variable) increased in general 60%.

In light of these findings, the effect of the partnership approach on moral perception is potentially positive. Firstly, as the study shows, the partnership increases efficiency via a better division of labour, lowering the stress while keeping the human in the decision-making process. Secondly, working agreements allow for better information management. If there's a danger that the operator will not be able to filter out morally relevant facts because he either receives too much or too little information, it needs to be ensured that he gets the right amount of information. As part of the working agreement, the operator determines which information is provided and how it is managed in the decision-making process.

The partnership approach protects a commitment to responsibility within the armed forces. First, operators will be responsible for their working agreements. This raises issues about foresight, negligence and so on that we cannot tackle here. For now, it suffices to note that the operator remains firmly control of his machine – even if there's a physical distance between them. Secondly, working agreements ensure that operators receive the morally relevant facts needed to make decisions that comply with international humanitarian law, as well as key moral principles.

CONCLUSION

We started by noting that some commentators have argued in favour of the deployment of NCTs during MI. Indeed, there are some benefits associated with NCTs. But the difficulties posed by these systems must not be neglected. Before NCTs can be deployed, we need to rest assured that their usage is safe and that they enhance, rather than undermine, human decision-making capacities. During MI, the operational requirements upon NCTs and those who operate them are therefore high. Ordinary operators and their superiors need reliable information about the complex environment they operate in, especially when they are not directly present. We propose a design recommendations based on the partnership analogy, ensuring that operators use appropriate information in the decision-making process. There are important choices to be made, and sound design is 'design for responsibility'. Admittedly, the partnership approach is no magic formula and we doubt that there is such a formula. However, partnerships are a promising way forward, especially when compared to proposals for fully operationally autonomous machines.

ACKNOWLEDGEMENTS

This work is part of the research project 313-99-260 'Military Human Enhancement: design for responsibility in combat systems' that is financed by the Netherlands Organization for Scientific Research (NWO) in the research program line Societal Responsible Innovation. Parts of this publication have been also been published in the Whitehall Report "Hitting the Target?" published by the Royal United Services Institute (RUSI).

REFERENCES

Altman, A., & Wellman, C. H. (2009). *A Liberal Theory of International Justice*. Oxford University Press.

Arciszewski, H. F. R., de Greef, T. E., & van Delft, J. H. (2009). Adaptive Automation in a Naval Combat Management System. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, *39*(6), 1188–1199.

Arkin, R. (2010). The Case for Ethical Autonomy in Unmanned Systems. *Journal of Military Ethics*, *9*(4), 332–341.

De Greef, T. E., Arciszewski, H. F. R., & Neerincx, M. A. (2010). Adaptive Automation Based on an Object-Oriented Task Model: Implementation and Evaluation in a Realistic C2 Environment. *Journal of Cognitive Engineering and Decision Making*, *31*, 152–182.

Endsley, M., & Kiris, E. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors*, 381–394.

Hollnagel, E., & Woods, D. D. (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton: Taylor and Francis.

Ignatieff, M. (2000). *Virtual Wars: Kosovo and beyond*. London: Vintage.

Johnson, M., Bradshaw, J. M., Feltovich, P., Jonker.C., Van Riemsdijk, B., & Sierhuis, M. (2011). The Fundamental Principle of Coactive Design: Interdependence Must Shape Autonomy. In M. De Vos, N. Fornara, J. Pitt, & G. Vouros (Eds.), *Coordination, Organizations, Institutions, and Norms in Agent Systems VI* (Vol. VI, pp. 172–191). Berlin / Heidelberg: Springer.

Matthias, A. (2004). The responsibility gap. Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, (6), 175–183.

May, L. (2005). *Crimes Against Humanity: A Normative Account*. *European Journal of International Law* (Vol. 18, p. 310). Cambridge University Press.

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. Amsterdam: North-Holland.

Sharkey, N. (2010). Saying "No!" to Lethal Autonomous Targeting. *Journal of Military Ethics*, *9*(4), 369–383.

Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, *24*(1), 62–77.

Strawser, B. J. (2010). Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles. *Journal of Military Ethics*, *9*(4), 342–368.

Strawser, B. J. (2013). *Killing by Remote: Ethics of an unmanned military*. Oxford: Oxford University Press.