Ethics and Information Technology (2023) 25:17
https://doi.org/10.1007/s10676-023-09695-w

ORIGINAL PAPER

Role of emotions in responsible military AI

Jurriaan van Diggelen[1] · Jason S. Metcalfe[2] · Karel van den Bosch[1] · Mark Neerincx[1] · José Kerstholt[1]

Introduction

Following the rapid rise of military Artificial Intelligence (AI), people have warned against mankind withdrawing from the immediate war-related events resulting in the "*dehumanization*" of war (Joerden, 2018). The premise that machines decide what is destroyed and what is spared, and when to take a human life, would be deeply morally objectionable. After all, a machine isn't a conscious being, doesn't have emotions like empathy, compassion, or remorse, and isn't subject to military law (Sparrow, 2007). This argument has sparked the call for *meaningful human control*, requiring moral decisions to be made by humans, not machines (Amoroso & Tamburrini, 2020). The United States have proposed a similar principle, named "appropriate levels of human judgment"[1]. Likewise, the NATO principles of responsible use of AI in Defence (NATO, 2021) state that "AI applications will be developed and used with appropriate levels of judgment and care".

While the definition of these principles, and how they should be operationalized, is controversial and under debate, one thing is commonly agreed upon: the human must play the role of moral agent in military decision-making (Scharre, 2018). To enable humans to fulfil this role, the human-AI interaction must provide adequate support for this. Frequently claimed support capabilities are: the AI system must be able to explain its reasoning to humans (NATO, 2021); the human must be informed sufficiently and timely by the AI (Boardman & Butcher, 2019); and the human must be able to inspect and intervene in the plans and decisions of the AI (Ekelhof, 2019).

Although useful, we contend that such support is insufficient because it only applies to the rational reasoning processes of moral decision-making. We argue that support should also involve the processing of emotions that the human experiences, as these intrinsically reflect the human's personal values towards the decision- making problem. This paper discusses the function of emotions in military moral decision-making and claims that an appropriate level of emotional involvement is required for all those in the decision-making chain. Finally, we provide suggestions for the design and implementation of such emotional support.

Human emotions matter in military decision-making

AI-systems are capable of analyzing situations and judging the value of possible actions in rational terms. As AI-systems do not have emotions, their decisions, or the decisions they propose to humans, will always follow logic and rationality. This is believed by some to bring about superior decisions (Haraburda, 2019). It is understandable why some regard emotions as detrimental to decision-making. Firstly, emotions are difficult to standardize and control. In a critical organization like the military, soldiers need to behave in a predictable and consistent manner, which should not be disturbed by their individual beliefs, momentary fears and desires. Second, emotions may be poor moral advisors. For example, emotions such as anger might induce feelings of revenge which may in a military context even lead to war crimes. Thirdly, strong emotions may lead to functional dropout, such as post-traumatic stress syndrome (PTSS),

[1] https://geneva.usmission.gov/2016/04/12/u-s-delegation-statement-on-appropriate-levels-of-human-judgment/.

✉ Jurriaan van Diggelen
jurriaan.vandiggelen@tno.nl

Jason S. Metcalfe
jason.s.metcalfe2.civ@army.mil

Karel van den Bosch
karel.vandenbosch@tno.nl

Mark Neerincx
mark.neerincx@tno.nl

José Kerstholt
jose.kerstholt@tno.nl

[1] TNO, Soesterberg, Netherlands

[2] Army Research Laboratory (ARL), Maryland, USA

Published online: 16 February 2023

or decision paralysis. Others however argue that emotions are important, as they shape the choices of a military decision maker, and help to make decisions (Zilincik, 2022; Desmet & Roeser, 2015). We too argue that the property of humans to experience emotions is critically important for the appraisal of decision problems, and that this property enables a human-AI team to make decisions that are aligned with the conception of morality as adopted by the individual, its organization, and society. We do not argue that analytical rules should not be used when making decisions in moral situations. In contrast, such rules are important, and should be used as guide in the decision-making process. We also do not imply that human emotions should always have the final word in making the decision, because it is evident that emotions, particularly if they are intense, may distort a person's judgment, causing biased and erroneous decisions (Williams, 2010). But we do argue that acknowledging and processing emotions is important as they enable humans to appreciate what applying a moral rule really means in a given situation; to feel the consequences of considered decisions. We assert that such emotion-induced feeling of anticipated outcomes is essential. It helps to feel committed to the upcoming decision, to feel responsible, to feel regret for likely consequences, and to accept accountability.

We will discuss two functions of emotions that are important for moral decision-making.

Emotions reflect values

A main function of emotional reactions is to provide the individual with information about the subjective value attached to the pros and cons of the set of options available (Hartley & Sokol-Hessner, 2018). These emotions can be directly experienced in the decision situation at hand, but can also play a role in anticipating a particular outcome (Loewenstein & Lerner, 2003). Anticipation means that the decision maker mentally simulates a particular outcome and the related feelings, for example regret. Whether anticipated or directly experienced, an emotion informs the decision maker that the situation is appraised as relevant to one's concerns. Beneficial outcomes lead to positive emotions, detrimental outcomes to negative ones. As argued by (Schwartz, 2016), both rational and emotional evaluations are needed for human moral reasoning.

To illustrate this, consider a real-life scenario described by (Scharre, 2018): "My sniper team had been sent to the Afghanistan-Pakistan border to scout infiltration routes where Taliban fighters were suspected of crossing back into Afghanistan … A young girl of maybe five or six headed out of the village and up our way … frequently glancing in our direction. It wasn't a very convincing ruse. She was spotting for Taliban fighters." Exposed to the risk of being attacked by the Taliban, the team brought itself to safety again by calling for exfiltration and aborting the mission. During the mission debrief, the team realized they had another, possibly more "mission-effective", option. The girl participated in hostile activities by doing the spotting for the enemy and would have been a lawful target for engagement. About that option, Scharre writes: "Of course, it would have been wrong. Morally, if not legally. In our discussion, no one needed to recite the laws of war or refer to abstract ethical principles. No one needed to appeal to empathy. The horrifying notion of shooting a child in that situation didn't even come up".

This example illustrates that human moral decision-making is the result of a complex interplay of many simultaneous factors, such as sensations, feelings, emotions, and thoughts. Feelings and emotions are often the first reactions to a situation. They occur automatically and form part of subsequent judgment processing, providing information about our main concerns, our core values.

Emotions drive motivations and behavior

As emotions reflect our core values, they are the main drivers of motivation and behavior (Zeelenberg et al., 2008). Emotions act as a spotlight that optimizes the usage of our scarce cognitive resources. They may indicate what aspects of a situation have our focus during the moral decision-making process. Examples are emotions such as fear or desire of the anticipated outcome, or compassion with or dislike of the persons affected by the moral decision. However, even though the emotion will trigger a behavioral tendency, emotion regulation processes may lead to different outcomes. A decision maker could, for example, make a reappraisal of the situation or refrain from performing the behavior on the basis of a more thorough risks analysis. Emotions continue to affect the decision maker after the moral activity has been conducted, i.e. retrospectively. For example, emotions such as satisfaction, joy, sadness and regret, facilitate the decision-maker to reflect upon the considerations and decision, learn from it, and use that on subsequent occasions. Eventually, this will lead to the learned behavior being ingrained in their intuition.

Emotional engagement in an AI-driven defense organization

Following our argument that moral judgment necessitates an appropriate level of emotional involvement, we will now discuss why this is becoming increasingly important as AI-systems become more widely used. Although it is unclear how AI will exactly change military practice, we

can identify a number of trends and predict how they will affect human emotional engagement.

AI-tools are very well equipped to adopt a purely rational and computational moral reasoning style. Given the appeal of emotionless moral reasoning (as argued in the preceding section), some researchers have argued that the rise of AI should be embraced as an opportunity to make warfare more moral (Arkin, 2010). There are two arguments against this viewpoint. Firstly, it ignores the previously mentioned function of emotions: acknowledging moral values and motivating behavior. Secondly, it is based on the flawed idea that AI would put the human out of the loop entirely (Johnson & Vera, 2019).

Although humans may not be present at the moment that an AI-based system autonomously executes a morally sensitive activity (e.g., due to required reaction time or a lack of connectivity), humans will unarguably be present in other phases of the operation, such as mission planning, or debriefing. Furthermore, during the development of an AI-system, human programmers were involved in designing the AI's moral behavior.

For example, consider a future minesweeping operation by Autonomous Underwater Vehicles (AUV's). The AUV's are equipped with preprogrammed behaviors that allow them to inspect and defuse naval mines. During the planning phase, human navy personnel are responsible for tasking the system. They specify the search pattern, the available time, and how to act when mines are positioned close to other objects, such as fisher boats. Weighing the risks of missing an enemy mine against the risk of unintentional damage to fisher boats is a moral consideration that is conducted during the planning phase. During the mission the AUV performs its actions fully autonomously, as in underwater operations there are no opportunities for *real-time* human-robot communication. During the debriefing phase, the AI informs and explains to the human operators on how the mission went and suggests potential points of improvement.

This future scenario illustrates that AI still requires human involvement, but that human control is limited to prior to, and after the operation (Diggelen et al., 2023). Clearly, these mission characteristics have implications for a human's emotional involvement. It may be expected that during the mission, operators experience low to moderate levels of emotion, as when the AI makes its critical decisions, the operators cannot monitor the situation, nor can they intervene at that point in time. However, the impact of the AI's decisions in critical situations is severe. For developers to properly design decision-making for the AI in advance of the actual operation, they need to properly appreciate the moral implications of potential decisions. We argue that this proper appreciation does not arise when the problem is treated as a rational calculation only. Designers need a proper emotional involvement to feel engaged in the considerations and feel committed and responsible for the decisions they eventually implement in the system. Without a proper emotional involvement, designers may run the risk of becoming indifferent to the moral consequences of decisions taken by *their* AI-system.

Likewise, developers that instruct an AI-based system to act on the battlefield are likely to be less emotionally involved in the system's decisions and the consequences thereof than soldiers who actually operate on the battlefield. However, the contrary could also be true. For example, a drone pilot may experience more intense emotions than a traditional fighter jet pilot. This is because drone operators often closely observe their assigned targets for an extended period of time using high fidelity video. Because of this, they are likely to see the target as a human who goes about normal life activities. It is likely that the drone-operator experiences strong emotions when anticipating the future outcomes of decisions, or when observing their consequences (Enemark, 2019).

Summarizing, deploying AI in military organizations will have severe consequences for how humans are involved in military missions, and will have disruptive effects on human emotional involvement. Given that emotions play such an important role in moral behavior, this aspect requires proper attention when developing responsible AI.

Designing for emotional involvement

The first step in the design process is to determine the roles of humans and AI in moral decision-making when they collaborate, either directly or indirectly. These humans could be programmers, planners, operators, or commanders. They should all be supported, not only in an analytical-rational manner, but also to appreciate the morality of the impending decisions. This requires the right level of emotional engagement. Note that higher emotional involvement is not necessarily better. As argued, emotional experiences have downsides and upsides for moral judgement. Therefore, determining the right level of emotional engagement is far from trivial and is currently poorly understood.

The next step should be to design human-AI interactions, which involve emotional appraisals that address the moral values at stake and can be related to the rational moral reasoning in the decision-making process. Abstract numbers and symbols may not trigger the required experience for such moral assessments. Dialogues with a conversational AI-agent (Hagemeijer, 2020) could provide and request concrete information of the situation to create appropriate engagement and understanding of the relevant moral aspects. Immersive displays may also support such

assessments, by controlling the perceptual richness of the situation (visuals, sounds, tactile, or smell), and by creating a narrative of the situation. In general, such human-AI interactions would help to sense and weight the moral aspects, accommodating appropriate emotional appraisals and value assessments.

Conclusion

Emotions play a crucial role in human moral decision-making. They reflect moral values in a person, and they establish the motivation and engagement required for appropriate moral judgement and care. Therefore, emotions cannot be ignored when designing responsible military artificial intelligence. Human-AI interaction relies on asynchronous control of technology, whereby the humans specify in advance what decisions the machines should take when anticipated events occur later in time. This asynchrony between decision-making and decision-execution raises novel challenges, such as how to assist decision-makers in comprehending and feeling the moral impact of potential decisions.

We propose the following research agenda to tackle these. Firstly, we must study which types and levels of emotion are most appropriate for moral decision-making in various situations, how they represent the values at stake, and how they interact with rational evaluations. Secondly, we must better understand how these emotions can be evoked in human-AI interactions, such as in dialogues with conversational agents and sensory and narratively immersive user interfaces. Overall, we argue that there is a fundamental need to more fully integrate humans into AI-driven organizations. Attempts to simplify human-AI interaction by parceling out or ignoring human emotional state run a considerable risk of omitting some of the most valuable information available from the human counterparts, i.e. their emotions. This, we argue, may lead to instantiating the exact problems that *meaningful human control* is meant to avoid.

References

Amoroso, D., & Tamburrini, G. (2020). Autonomous weapons systems and meaningful human control: ethical and legal issues. *Current Robotics Reports*, *1*(4), 187–194.

Arkin, R. C. (2010). The case for ethical autonomy in unmanned systems. *Journal of Military Ethics*, *9*(4), 332–341.

Boardman, M., & Butcher, F. (2019). An exploration of maintaining human control in AI enabled systems and the challenges of achieving it. In *Workshop on Big Data Challenge-Situation Awareness and Decision Support. Brussels: North Atlantic Treaty Organization Science and Technology Organization. Porton Down: Dstl Porton Down*.

Desmet, P. M., & Roeser, S. (2015). Emotions in design for values. Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains, 203–219.

van Diggelen, J., van den Bosch, K., Neerincx, M., & Steen, M. (2023). *Designing for Meaningful Human Control in Military Human-Machine Teams, in Research handbook on Meaningful Human Control of Artificial Intelligence Systems*. Edward Elgar Publishing.

Ekelhof, M. (2019). Moving beyond semantics on autonomous weapons: meaningful human control in operation. *Global Policy*, *10*(3), 343–348.

Enemark, C. (2019). Drones, risk, and moral injury. *Critical Military Studies*, *5*(2), 150–167.

Hartley, C., & Sokol-Hessner, P. (2018). Affect is the foundation of value. In A. S. Fox, R. C. Lapate, A. J. Shackman, & R. J. Davidson (Eds.), *The nature of emotion: fundamental questions*. 348 – 51, New York: Oxford University Press.

Hagemeijer, M. (2020). Affective Intelligent System Design for Empathy in Decision-making Unpublished Bachelor Thesis.

Haraburda, S. S. (2019). Benefits and Pitfalls of Data-Based Military Decisionmaking | Small Wars Journal. Retrieved January 16, 2023, from https://smallwarsjournal.com/jrnl/art/benefits-and-pitfalls-data-based-military-decisionmaking

Joerden, J. C. (2018). Dehumanization: the ethical perspective. *Dehumanization of Warfare* (pp. 55–73). Cham: Springer.

Johnson, M., & Vera, A. (2019). No AI is an island: the case for teaming intelligence. *AI magazine*, *40*(1), 16–28.

Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences*. New York: Oxford University Press.

NATO, NATO principles of responsible AI (2021). ; https://www.nato.int/cps/en/natohq/official_texts_187617.htm

Scharre, P. (2018). *Army of none: Autonomous weapons and the future of war*. WW Norton & Company.

Schwartz, M. S. (2016). Ethical decision-making theory: an integrated approach. *Journal of Business Ethics*, *139*(4), 755–776.

Sparrow, R. (2007). Killer robots. *Journal of applied philosophy*, *24*(1), 62–77.

Williams, B. S. (2010). Heuristics and biases in military decision-making.Military Review,40–52

Zeelenberg, M., Nelissen, R. M., Breugelmans, S. M., & Pieters, R. (2008). On emotion specificity in decision-making: why feeling is for doing. *Judgment and Decision-making*, *3*(1), 18.

Zilincik, S. (2022). The role of Emotions in Military Strategy. *Psychology*, *5*(2), 11–25.