Human Factors Considerations in Autonomous Lethal Unmanned Aerial Systems

Kristine M. Kiernan
Embry-Riddle Aeronautical University

ABSTRACT

The United States military is committed to the development of complete autonomy in unmanned vehicles, including armed unmanned aerial systems (UAS). The design and deployment of autonomous lethal UAS raises ethical issues that have implications for human factors. System design, procedures, and training will be impacted by the advent of autonomous lethal UAS. This paper will define relevant vocabulary, review the literature on robot ethics as it applies to the military setting, discuss various perspectives in the research community, address levels of UAS autonomy, and discuss implications for operator training, responsibility, and human-machine interaction. Familiarity with these ethical issues and their repercussions will prepare human factors practitioners for the challenges created by this developing technology.

INTRODUCTION

The use of unmanned aerial systems (UAS) has increased dramatically in the past ten years (McMahan & Strawser, 2013), and is forecast to grow (Clapper, Young, Cartwright, & Grimes, 2007). The military forces of the United States are committed to the development of complete autonomy in unmanned vehicles, including armed UAS (Sharkey, 2008). However, the development and deployment of autonomous lethal UAS raises questions about ethics that must be addressed. How will these autonomous systems make moral choices? What are the potential moral costs and benefits of autonomous lethal systems? If these systems are truly autonomous, who bears responsibility for their actions? In addition, the operation of autonomous lethal UAS will lead to changes in the human-machine interface that have far reaching consequences for human factors.

The primary purpose of this paper is to introduce human factors practitioners to the issues, arguments, and possibilities surrounding autonomous lethal UAS. The secondary purpose is to suggest some of the changes in human-machine interaction that will develop as a result of the implementation of autonomous lethal UAS.

REVIEW

Definitions

A discussion of the ethical considerations of autonomous lethal UAS must begin by establishing a common vocabulary. Familiar terms, such as morality, ethics, autonomy, and responsibility need to be defined, as well as less familiar concepts such as the Laws of War and Just War theory.

The terms "morality" and "ethics" are often used interchangeably, but a distinction is necessary in the context of autonomous machines. Morality consists of behaviors and beliefs about what is right and wrong (Gros, Tessier, & Pichevin, 2012). Ethics, on the other hand, can be defined as philosophical reflection on morality (Ethics, 2013). Morality, therefore, is concerned with right behavior, while ethics is concerned with systems of thought about right behavior.

Morality and ethics are only relevant when an agent possesses sufficient autonomy to make choices. While there is no universally accepted definition in the literature, a useful view of autonomy in the context of robots is "the capacity to operate in the real-world environment without any form of external control for extended periods of time" (Lin, Bekey, & Abney, 2008, p. 103). Inherent in this definition is the ability to make decisions independently from outside control. However, autonomy is not a binary concept. Rather, there are levels of autonomy for both humans and machines. At the highest level, human autonomy includes the Kantian notion of *autonomy of will*. That is, human beings have the capacity to think ethically, to reflect upon morality and formulate a system by which to make choices. This autonomy of will is not particularly desirable in automated systems, since the main purpose of using robots is to have them meet the goals set by the human operator (Gros et al., 2012, p. 2). Therefore, when we speak of autonomy in robots, we are talking about *autonomy of means* in how to accomplish a goal, not of end in choosing a goal. Creating a robot capable of moral behavior is operationally desirable; creating a robot with human-like autonomy of will that is capable of thinking ethically is not.

Although they share the same etymology, autonomy and automation must also be distinguished. Automation is a "system that accomplishes a function that was

previously carried out by a human operator" (Parasuraman, Sheridan, & Wickens, 2000). Automation, of itself, does not imply any ability to function independently or make decisions.

Possessing, at a minimum, autonomy of means is a necessary precondition for being assigned *moral responsibility*. In the context of lethal autonomous machines, moral responsibility is distinguished from causal responsibility. Causal responsibility is ascribed when an agent causes an outcome, while moral responsibility is ascribed only when an agent makes a decision that causes an outcome. For example, if a town is flooded because rain overwhelms the capacity of a dam, the rain has causal responsibility. If, however, the town is flooded because shortcuts were taken in building the dam, the builder bears moral responsibility. Clearly, the idea of moral responsibility is meaningless without the decision making power inherent in autonomy.

The most salient aspect of moral responsibility for an autonomous lethal machine is related to conduct during battle. Internationally accepted rules of behavior during wartime, known as Laws of Armed Conflict (LOAC) or Laws of War (LOW) are drawn from Just War theory. Just War theory usually consists of two elements: *jus ad bellum*, the issues involved in going to war; and *jus in bello*, the issues involved in conducting war (Schulzke, 2011). Even the most enthusiastic supporter of the role of autonomous lethal UAS would acknowledge that we are a long way from allowing a machine to decide when and why we wage war; therefore, discussion primarily involves *jus in bello*. The two generally accepted components of *jus in bello* are proportionality and discrimination. Proportionality requires that the use of force be at an appropriate level for the threat, while discrimination requires that force be applied only to the threat rather than to noncombatants. United States military forces are bound by law to abide by LOAC (Department of Defense, 2011).

Approaches to Creating Moral Machines

Three approaches are generally considered in the creation of moral reasoning in robots: top-down, bottom-up, and hybrid (Allen, Smit, & Wallach, 2005). The top-down approach involves a rule-based, approach to programming moral reasoning. Several rule-based approaches are part of the Western tradition, including consequentialist, in which the morality of an act is judged by its outcome, and Kantian, in which moral behavior consists of adhering to the Categorical Imperative that one should behave "only according to that maxim by which you can at the same time will that it be a universal law" (Gros et al., 2012). In the military setting, the Laws of War provide a set of rules which could be used to govern moral reasoning. Top-down approaches have the advantage of clarity and simplicity, but operationally, the main weakness is that no set of rules can anticipate every situation or accommodate every context. In the real world, rules can often conflict, requiring prioritization and compromise beyond the capabilities of a rule based computational system. On the battlefield, much is left to the judgment of the soldier.

Some have suggested that instead, robots should learn morality the way children do, through learning and experience (Turing, 1950). In this bottom-up approach, machine learning or artificial evolution would be used to inculcate moral reasoning in a UAS. As in any instructional situation, appropriate behavior would be rewarded while undesirable behavior would be punished (Allen et al., 2005). The bottom-up approach allows flexibility in complex situations. However, one of the main concerns with this approach is that performance can never be guaranteed. As any parent knows, even perfect training is not a guarantee of perfect behavior. Furthermore, if the machine makes a bad moral choice, there is no way to trace and correct the underlying cause (Gros et al., 2012). Another objection to this approach to machine morality is that it does not account for the effect of natural law. In the Western philosophical tradition, natural law refers to the presumption that human beings are born with some intrinsic morality. Clearly a bottom-up approach would need to account for the absence of natural law in machines.

Hybrid approaches combine aspects of the top-down rule based approach with the bottom-up learning approach. Of the many hybrid approaches, the most promising is based on virtue ethics (Lin et al., 2008). Instead of morality built on actions, virtue ethics considers morality built on character. Actions are determined by their compatibility with a set of virtues, for example courage, compassion, or honesty. The advantage is that virtues themselves constitute top-down guiding principles, while learning algorithms allow bottom-up approaches to learning specific actions that are compatible with the virtues (Lin et al., 2008). A useful metaphor for the hybrid approach to developing moral beliefs and behavior can be found in grammar acquisition. Children acquire a working knowledge of grammar from experience. However, rules do exist that govern grammar usage. Similarly, robots could learn moral beliefs and behavior from experience, but those same beliefs and behaviors would be learned by conforming to guiding principles. However, combining these two opposing strategies involves harmonizing not only technical approaches, but underlying philosophies.

Perspectives in the Research Community

Divisions exist within the research community not only about how to engineer moral reasoning in robots, but also about the ethics of using autonomous lethal robots at all. The arguments against autonomous lethal robots are presented largely in the context of Just War theory. Prominent critics (Sharkey, 2008; Singer, 2009) argue that risk-free war encourages going to war, and increases the likelihood of violating of the precepts of *jus ad bellum*, the rules of going to war. Another concern is that the *jus in bello* principles of proportionality and discrimination are impossible to operationalize. No clear definition of "disproportionate suffering" or "civilian" can be created that would be airtight in combat situations.

Advocates of autonomous lethal robots assert that machines do not need to be morally perfect in their actions, only better than their human counterparts. Human beings are subject to fatigue, anger, fear, vengefulness, and other qualities that have been implicated in wartime atrocities (Arkin, 2010). Robotic systems, on the other hand, would be able to behave morally without the weaknesses unavoidable in human soldiers (Schulzke, 2011).

Opponents also argue that a precondition of *jus in bello* is that moral responsibility can be assigned for all actions (Sparrow, 2007). The very nature of autonomy means that the robot operates independently, meaning it would be unfair to hold the programmer or the operator responsible for the robot's actions. Supporters counter that responsibility can be assigned within the framework of product liability laws (Lucas, 2012). If a faulty toaster burns your house down, the manufacturer is at fault: if a faulty robot kills an innocent, the manufacturer is held morally responsible. These issues remain unresolved in the literature.

Levels of Autonomy

Complicating the issue of ethical use of autonomous lethal robots is the fact that autonomy itself has gradations. The most accepted taxonomy of levels of machine autonomy is Sheridan's range from 1, computer offers no assistance and human does it all, to 10, computer decides everything and acts autonomously (Parasuraman et al., 2000).

Another way of approaching autonomy is to describe it according to human-machine interaction, in which the lowest level would be the direct control of teleoperation, and the highest level would be the dynamic autonomy of peer-to-peer collaboration (Goodrich & Schultz, 2007).

Each military branch has its own slightly different taxonomy of automation, with some conflating autonomy and intelligence. However, most researchers encourage separation of these two constructs, since complete autonomy is possible without intelligence, for instance in a jellyfish, and intelligence is possible without autonomy, for instance in a child (Clough, 2002).

Recently, taxonomies of autonomy have focused on a three-axis model consisting of the mission complexity the robot can handle, the environmental difficulty the robot can handle, and the independence from human interaction that the robot is capable of (Huang, Pavek, Novak, Albus, & Messin, 2005). This three-axis model accounts for the obvious idea that a robot able to handle a complex mission in a challenging environment should be considered to have greater autonomy than a robot capable only of a simple mission under controlled circumstances, even if both robots have identical human interaction requirements. For simplicity, this model will ultimately categorize robot autonomy on a scale of 1-10, in which the highest level consists of absolutely no required human interaction.

NEW CONTRIBUTION

Implications for Human Factors

Although the highest levels of autonomy have not yet been achieved in aviation, some amount of perception, decision, and action autonomy have been incorporated in various subsystems, for example Traffic Collision Alerting Systems (TCAS). In some cases, automation is linked with autonomy. For example, the autopilot is physically controlling the aircraft while the flight management system determines the heading required for navigation.

Parasuraman and Wickens (2008) have advocated adjusting levels of automation, and indirectly, autonomy, according to the stage of human information processing being supported. They have cautioned against autonomy in decision making, particularly as it relates to lethality or human safety, until reliability can be assured. As noted above, however, supporters of autonomous lethal UAS argue that reliability need not be perfect, only better than human reliability in the same situation (Arkin, 2010).

Automation in aviation has created issues with system observability, mode confusion, and automation surprise (Ferris et al., 2010); reduced situation awareness (SA), trust, reliability, overreliance, and complacency (Galster et al., 2007); and skill degradation (Parasuraman et al., 2000).

Observability, mode confusion, automation surprise, and situation awareness.

Low observability, coupled with high complexity and decision making authority, creates the potential for mode confusion and automation surprise, which are specific failures of situation awareness. The bottom-up approach to engineering UAS morality would create low system observability, as the reasoning behind the UAS' choices would not be known to the operators (Gros et al., 2012). Combined with the high complexity and decision making authority inherent in an autonomous UAS, this low system observability has the potential to degrade operator SA. Top-down approaches, however, might alleviate this problem, since operators could become familiar with the fundamental moral architecture of the system.

On the flight deck, systems that provide pilots with greater feedback beyond simply system behavior have been shown to improve mode awareness (Ferris, Sarter, & Wickens, 2010). Thus, the design of human-UAS interface should include clear information about why the UAS is behaving in a certain way. For example, a UAS should alert its operator if it is confronted with a moral choice, and inform the operator of its reasoning process.

Trust, reliability, overreliance, and complacency.

Operator trust in automation effects system performance (Ferris et al., 2010). Understanding the rules that govern system behavior has been shown to increase operator trust (Galster et al., 2007). At the same time, automation that functions properly but does not conform to the operator's expectations has been shown to reduce trust (Lee & See, 2004). Trust between humans develops in a different manner from trust between humans and machines (Lee & See, 2004). The underpinnings of trust between humans and machines that have nearly human autonomy will be an interesting area of research.

In addition to transparency, one of the variables affecting operator trust is reliability, both actual and perceived (Lee & See, 2004). When malfunctions do occur, the UAS failure mode must be apparent. Clearly, fault alert systems must go beyond warning lights and horns toward rich, contextual communication to help the operator understand the failure mode of the UAS.

Excessive trust in automation, however, can lead to overreliance and complacency (Parasuraman & Riley, 1997). High workload appears to potentiate the effects of excessive trust, so that operators fail to monitor the automation as they should (Galster et al., 2007). This may particularly become an issue as the ratio of operators to UAS decreases, and operator workload increases.

As a result of overreliance, operators may fail to monitor inputs to the automation, reducing SA and making it difficult to take over should the automation fail (Parasuraman & Riley, 1997). The lack of SA created by low system observability would exacerbate this problem.

Skill degradation.

Automation of manual tasks has been shown to lead to skill degradation (Parasuraman et al., 2000). The same effect has been shown for decision making tasks (Galster et al., 2007). One potential danger of UAS autonomy is that human operator decision making skills may be lost. Reverting to lower levels of automation can successfully ameliorate skill loss, but this strategy may not be practical in fully autonomous UAS because of reduced SA due to low system observability, overreliance, and the distance inherent in the relationship between an operator and an autonomous agent. In addition, reverting to a lower level of automation is only effective if the manual skill, in this case decision making, has been learned in the first place. How can human supervisors acquire the skill of ethical decision making in battlefield situations if they are never meant to use them operationally?

Training.

Galster et al. (2007) suggest that the highest levels of machine autonomy may require types of operator training not required for the lower levels. They suggest that we already have a model for the skills required for supervising fully autonomous UAS – that is, supervising human beings. Some of the skills required include delegation and communication. Both of these skills have been taught to military and commercial aircrews for decades in the form of Crew Resource Management (CRM). Perhaps CRM training can be tailored to the unique demands of a system that includes humans, machines, and machines that behave as humans.

Implications for human centered design

Human centered design is based on the premise that if humans have final responsibility for a system, the "human operator should be at the heart of a system with full authority over all its functioning" (Parasuraman & Riley, 1997, p. 248). But if the final responsibility for system safety does not rest with the human operator, is user-centered design still relevant? What is the place of human factors engineering if the human being is no longer the heart of the system?

Further, to some extent, automation shifts the locus of error from the operator to the designer (Parasuraman & Riley, 1997). One task for those designing training and procedures will be to account for and mitigate these new error modes.

DISCUSSION

Research and development in unmanned military systems is driving toward full autonomy, even for lethal systems. This raises ethical questions about the design and implementation of moral machines. In particular, the advent of autonomous UAS will create issues in human-machine interface that go beyond what has been seen in flight deck and UAS automation to date. Human factors practitioners must be involved in autonomous UAS design from the beginning. Looking back at the issues raised by flight deck automation, and ahead to the issues unique to the relationship between operators and autonomous UAS will help prepare human factors practitioners to address the challenges raised by this developing technology.

REFERENCES

Allen, C., Smit, I., & Wallach, W. (2005). Artificial morality: Top-down, bottom-up, and hybrid approaches. *Ethics and Information Technology, 7*(3), 149-155.

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

Clapper, J., Young, J., Cartwright, J., & Grimes, J. (2007). Unmanned systems roadmap 2007-2032. *Office of the Secretary of Defense*.

Clough, B. T. (2002). *Metrics, schmetrics! How the heck do you determine a UAV's autonomy anyway*. Dayton, OH: Air Force Research Lab Wright-Patterson Air Force Base.

Department of Defense, (2011). *DoD Law of War Program.* (DoD Directive No. 2311.01E). Washington, DC: U.S. Government Printing Office.

Ethics, (2013). In *Oxford Dictionaries*. Retrieved from http://www.oxforddictionaries.com/us/definition/american_english/ethics?q=ethics

Ferris, T., Sarter, N., & Wickens, C. (2010). Cockpit automation: Still struggling to catch up. In E. Salas and D. Maurino, (Eds). *Human Factors in Aviation.* (pp. 479-503). Burlington, MA: Elsevier.

Galster, S., Barnes, M., Cosenzo, K., Hollnagel, E., Miller, C., Parasuraman, R., Reising, J., Taylor, R., and van Breda, L. (2007). Human automation integration. In *Uninhabited Military Vehicles: Human Factors Issues in Augmenting the Force.* NATO Report RTO-TR-HFM-078.

Goodrich, M. A., & Schultz, A. C. (2007). Human-robot interaction: A survey. *Foundations and Trends in Human-Computer Interaction, 1*(3), 203-275.

Gros, F., Tessier, C., & Pichevin, T. (2012) Ethics and authority sharing for autonomous armed robots. *Autonomous Agents (RDA2) 2012,* 7.

Huang, H. M., Pavek, K., Novak, B., Albus, J., & Messin, E. (2005). A framework for autonomy levels for unmanned systems (ALFUS). *Proceedings of the AUVSI's Unmanned Systems North America,* 849-863.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46(1),* 50-80.

Lin, P., Bekey, G., & Abney, K. (2008). *Autonomous Military Robotics: Risk, Ethics, and Design.* San Luis Obispo, CA: California Polytechnic State University.

Lucas Jr, G. R. (2011). Industrial challenges of military robotics. *Journal of Military Ethics, 10*(4), 274-295.

McMahan, J., & Strawser, B. J. (2013). *Killing by remote control: The ethics of an unmanned military.* Oxford, England: Oxford University Press.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 30*(3), 286-297.

Parasuraman, R., & Wickens, C.D. (2008). Humans: Still vital after all these years of automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 2008 50: 511.

Schulzke, M. (2011). Robots as weapons in just wars. *Philosophy & Technology, 24*(3), 293-306.

Sharkey, N. (2008). Cassandra or false prophet of doom: AI robots and war. *Intelligent Systems, IEEE, 23*(4), 14-17.

Singer, P. W. (2009). *Wired for war: The robotics revolution and conflict in the twenty-first century*. New York, NY: Penguin.

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

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433-460.