Lethality and Autonomous Robots: An Ethical Stance

Ronald C. Arkin and Lilia Moshkina
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332
{arkin,lilia}@cc.gatech.edu

Abstract

This paper addresses a difficult issue confronting the designers of intelligent robotic systems: their potential use of lethality in warfare. As part of an ARO-funded study, we are currently investigating the points of view of various demographic groups, including researchers, regarding this issue, as well as developing methods to engineer ethical safeguards into their use in the battlefield.

1. Introduction

Battlefield ethics has for millennia been a serious question and constraint for the conduct of military operations by commanders, soldiers, and politicians, as evidenced, for example, by the creation of the Geneva conventions, the production of field manuals to guide appropriate activity for the warfighter in the battlefield, and the development and application of specific rules of engagement for a given military context.

Breeches in military ethical conduct often have extremely serious consequences, both politically and pragmatically, as evidenced recently by the Abu Ghraib and Haditha incidents in Iraq, which can actually be viewed as increasing the risk to U.S. troops there, as well as the concomitant damage to the United State's public image worldwide.

If the military keeps moving forward at its current rapid pace towards the deployment of intelligent autonomous robots, we must ensure that these systems be deployed ethically, in a manner consistent with standing protocols and other ethical constraints that draw from cultural relativism (our own society's or the world's ethical perspectives), deontology (right-based approaches), or within other related ethical frameworks.

Under the assumption that warfare, unfortunately and inevitably, will continue into the foreseeable future in different guises, the question arises as to how will the advent of autonomous systems in the battlefield affect the conduct of war. There already exist numerous conventions, laws of war, military protocols, codes of conduct, and rules of engagement, which are sometimes global in their application and at other times contextual, which are used to constrain or guide a human warfighter. Historically, mankind has been often unable to adhere to these rules/laws thus resulting in serious violations and war crimes.

Can autonomous systems do better? In this paper, we study the underlying thesis that robots can ultimately be more humane than human beings in military situations, potentially resulting in a significant reduction of ethical violations. This class of autonomous robots that maintain an ethical infrastructure to govern their behavior will be referred to as humane-oids.

2. Understanding the Ethical Aspects of Lethal Robots

As the Army's Future Combat System (FCS) moves closer to deployment, including weapons-bearing successors to DARPA's Unmanned Ground Combat Vehicle program, serious questions arise as to just how and when these robotic systems should be deployed. There are essentially two different cases:

1. **The robot as an extension of the warfighter.** In this relatively straightforward application, the human operator/commander retains all of the decisions regarding the application of lethality, and the robot is in essence a tool or weaponized extension of the warfighter. In this case, it appears clear that conventional ethical decision-making regarding the use of weaponry applies. A human remains in control of the weapons system at all times.

2. **The robot acting as an autonomous agent.** Here, the robot reserves the right to make its own local decisions regarding the application of lethal force directly in the field, without requiring human consent at that moment, while acting either in direct support of the conduct of an ongoing military mission or for the robot's own self-preservation. The robot may be tasked to conduct a mission that possibly includes the deliberate destruction of life. The ethical aspects regarding the use of this sort of autonomous robot are

unclear at this time, and they serve as the focal point of this article.

In order to fully understand the consequences of the deployment of autonomous machines capable of taking human life under military doctrine and tactics [1,2], a systematic ethical evaluation needs to be conducted to guide users (e.g., warfighters), system designers, policy makers, and commanders regarding the intended future use of this technology. This study needs to be conducted *prior* to the deployment of these systems, not as an afterthought.

Toward that end, a three-year research effort on this topic is being conducted in our laboratory for the Army Research Office, of which we are currently in the first year. Two topics are being investigated
:

(1) What is acceptable? *Can we understand, define, and shape expectations regarding battlefield robotics?* A survey is being conducted to establish opinion on the use of lethality by autonomous systems spanning the public, robotics researchers, policymakers, and military personnel to ascertain the current point-of-view maintained by various demographic groups on this subject.

(2) What can be done? *Artificial Conscience and Reflection*. We are designing a computational implementation of an ethical code within an existing autonomous robotic system, i.e., an "artificial conscience", that will be able to govern an autonomous system's behavior in a manner consistent with the rules and laws of war.

This paper focuses on the survey procedural aspects of this work, as the design and the software implementation of an ethical code will be conducted in years 2 and 3 of this project. It is too early to report the survey results as well, as it is still open and we want to ensure that experimental bias is as far removed as possible from the results. When the survey is closed, the results will be reported in a future article.

3. Survey Design

A web-based public opinion survey is currently being conducted to establish what is acceptable to the public and other groups regarding the use of lethal autonomous systems. The overall objective of the survey is three-fold:

1) To determine people's acceptance level of the use of lethal robots in warfare in the context of the following communities: military, robotics researchers, policy makers, and general public;

2) To identify how, and if, these opinions vary depending on whether the entity employed is a human soldier, a robot as an extension of a human soldier, or a fully autonomous robot;

3) To identify any variation in acceptance based on a variety of demographic factors.

In order to promote a better understanding of the rest of the section, definitions that are used in the survey are given as:

- Robot: as defined for this survey, an automated machine or vehicle, capable of independent perception, reasoning and action.

- Robot acting as an extension of a human soldier: a robot under the direct authority of a human, including authority over the use of lethal force.

- Autonomous robot: a robot that does not require direct human involvement, except for high-level mission tasking; such a robot can make its own decisions consistent with its mission without requiring direct human authorization, including decisions regarding the use of lethal force.

3.1 Survey Structure

Based on the stated objective, the independent variables used for this survey are as follows: a) community type, b) level of authority, and c) a number of demographic variables, such as age, gender, level of education, etc., including the extent of participants' knowledge of robots and their capabilities. This survey can be described as descriptive-explanatory, where we are interested not only in how the independent variables are distributed, but also in how they are related [3]. In addition to finding out what the terms of acceptance are for using lethal robots in warfare, we would also like to see if, and how, the level of acceptance varies between the different community types, according to certain demographics factors, and for the three levels of autonomy.

The survey is divided into three sections: prior knowledge and attitudes, questions regarding the terms of acceptance and ethical issues, and demographics. The questions in the first section are presented at the very beginning of the survey, before any definitions were given. In this section, the participants are asked a number of questions to assess their prior knowledge about robots in general and in the military, as well as the participants' overall attitude towards robots and human soldiers capable of taking human life in warfare.

The questions in the second section are presented after the definitions, and, where appropriate, they are asked separately for each level of autonomy: human soldier, robot as an extension of human soldier, and autonomous robot. They are of the following categories:

1)In what situations and roles are such robots acceptable, given the robots follow the same laws

of war and code of conduct as for a human soldier?

2) What does it mean to behave ethically in warfare?

3) Should robots be able to refuse an order from a human, and what ethical standards should they be held to?

3) Who, and to what extent, is responsible for any lethal errors made?

4) What are the benefits and concerns for the use of such robots?

5) Would an emotional component be beneficial to a military robot?

Finally, the questions in the last section, those assessing demographic factors, fall into the following categories:

1) Age, gender, region of the world where the participant was raised;

2) Educational background;

3) Current occupation and military experience, if any;

4) Attitude towards technology, robots, and wars;

5) Level of spirituality.

3.2 Survey Administration

To reach the widest possible audience, the survey is being conducted online, hosted by a commercial survey company, *SurveyMonkey.com*. All of the elements of the survey: each question, survey structure and layout were designed in accordance with survey design guidelines presented in [4], and then adopted for internet use, following the recommendations in [4] and [5]. To avoid order bias, response choices were randomized where appropriate. In addition, we varied the order in which the questions involving human soldier, robot as an extension of human soldier, and autonomous robot were presented. This was accomplished by creating two different versions of the survey, where the order was reversed in the second version; the participants will be randomly assigned to each of the survey versions.

In order to improve the survey quality, we have already completed an IRB-approved pilot survey. Twenty people participated in the pilot, including representatives from each of the four aforementioned community types; some of the respondents were also briefly interviewed after the survey completion. The pilot revealed a number of minor issues that have since been addressed in the subsequent survey revision, thus improving the overall quality; it also allowed us to better estimate completion times.

For the actual survey administration we adopted the four-prong approach recommended in [4] and [5] for internet surveys, which consists of sending pre-notification, invitation to participate, a thank you/reminder, and a more detailed reminder. One slight exception is in the use of pre-notification: most of our participants are expected to be recruited through postings to mailing lists, newsgroups, and other advertising methods, and we use this recruitment stage in lieu of pre-notification messages which would otherwise be sent to individuals.

The survey has been approved by the IRB, and is currently being deployed among the robotics researchers community. Recruitment methods among other community types are being explored, and will be chosen shortly.

4. Summary and Future Work

This article has deliberately avoided taking a formal position on the issue of the appropriateness of the use of lethal force and autonomous systems, other than to state that consideration of this issue is inevitable, while data collection is underway. Information from a range of demographic populations is being gathered with the intent of ultimately providing a framework for the conduct of these operations in a manner that will be more consistent with the laws of war than perhaps even humans could aspire to. The results of the survey and the design of the supporting robotic architecture will be reported when they become available.

5. Acknowledgements

This research is funded under Contract #W911NF-06-1-0252 from the U.S. Army Research Office.

6. References

[1] Collins, T.R., Arkin, R.C., Cramer, M.J., and Endo, Y., "Field Results for Tactical Mobile Robot Missions", *Unmanned Systems 2000,* Orlando, FL, July 2000.
[2] Hsieh, M., et al., "Adaptive Teams of Autonomous Aerial and Ground Robots or Situational Awareness", in submission, 2006.
[3] Punch, K.F., *Survey Research: The Basics*, Sage Publications, 2003.
[4] Dillman, D.A., *Mail and Internet Surveys: The Tailored Design Method*, John Wiley & Sons, Inc., Hoboken, NJ, 2007.
[5] Best, S.J., Krueger, B.S., *Internet Data Collection*, Sage Publications, 2004.