Moral Values Related to Autonomous Weapon Systems

An Empirical Survey that Reveals Common Ground for the Ethical Debate

Verdiesen, Ilse; Santoni De Sio, Filippo; Dignum, Virginia

DOI
10.1109/MTS.2019.2948439

Publication date
2019

Document Version
Final published version

Published in
IEEE Technology and Society Magazine

U.S. ARMY

Moral Values Related to Autonomous Weapon Systems

Ilse Verdiesen, Filippo Santoni de Sio, and Virginia Dignum

Digital Object Identifier 10.1109/MTS.2019.2948439
Date of current version: 2 December 2019

An Empirical Survey that Reveals Common Ground for the Ethical Debate

In the political debate on Autonomous Weapon Systems strong views and opinions are voiced, but empirical research to support these opinions is lacking. Insight into which moral values are related to the deployment of Autonomous Weapon Systems is missing. We describe the empirical results of two studies on moral values regarding Autonomous Weapon Systems that aim to understand the perception of people pertaining to the introduction of Autonomous Weapon Systems. One study consists of a sample of military personnel of the Dutch Ministry of Defense and the second study contains a sample of civilians. The results indicate both groups are more anxious about the deployment of Autonomous Weapon Systems than about the deployment of Human Operated drones, and that they perceive Autonomous Weapon Systems to have less respect for the dignity of human life. The concerns for Autonomous Weapon Systems creating new kinds of psychological and moral harm is very present in the public debate, and this is in our opinion one element that deserves to be carefully considered in future debates on the ethics of the design and deployment of Autonomous Weapon Systems. The results of these studies reveal a common ground regarding the moral values of *human dignity* and *anxiety* pertaining the introduction of Autonomous Weapon Systems which could further the ethical debate.

Autonomous Weapon Systems are weapon systems equipped with Artificial Intelligence (AI). AI can be described as a system that perceives its environment and selects actions to realize its predefined goals. Russell and Norvig [1] provide an overview of many definitions combining views on *systems that think and act like humans* and *systems that think and act rationally*, but they do not present a clear definition of their own. Bryson [2] states that a machine (or system) shows intelligent behavior if it can select an action based on an observation in its environment. According to Floridi and Sanders [3], AI is characterized by the concepts of adaptability, interactivity and autonomy. Adaptability means that the system can change based on its interactions and can learn from its experience. Machine learning techniques are an example of this. Interactivity occurs when the system and its environment act upon each other and Autonomy indicates that the system itself can change its state and goals [3]. Our definition of Autonomous Weapon Systems is influenced by the description provided by Floridi and Sanders [3] and can be characterized by the concepts of adaptability, interactivity and autonomy.

Autonomous Weapon Systems are increasingly deployed on the battlefield [4]. Autonomous systems can have many benefits in the military domain, for example when the autopilot of the F-16 prevents a crash [5], or when the use of robots by the Explosive Ordnance Disposal allows dismantling bombs with less risks for military personnel [6]. Yet the use of Autonomous Weapon Systems may also cause anxiety and concerns: it is feared that these systems may end up behaving in unpredictable and unwanted ways; that they may create "accountability gaps" [7], and more generally, that their use may be in conflict with respect for human dignity. All these concerns have been voiced in the societal debate raised among others by the "Stop Killer Robots Campaign" of 82 international, regional, and national nongovernmental organizations (NGOs) in 35 countries directed by Human Rights Watch [8], but also by the United Nations. They have stated that "Autonomous weapons systems that require no meaningful human control should be prohibited, and remotely controlled force should only ever be used with the greatest caution" [9]. In fact, as stated by Kaag and Kaufman [10] the deployment of Autonomous Weapon Systems on the battlefield without direct human oversight is not only a military revolution, but can also be considered a moral one. As large-scale deployment of AI on the battlefield seems unavoidable [11], research on the ethics of the design and use of these systems is imperative.

To be sure, theoretical reflections on the ethical risks posed by the use of Autonomous Weapon Systems are not lacking in the academic literature. One common argument is that given the current status of technological development, robot systems may not be capable of sophisticated practical and moral distinctions required by the laws of armed conflict [12]–[17], and this may raise the number of wrongs and crimes in military operations [18]. Others have gone as far as claiming that letting a machine be in control of the life and death of a

The knowledge gap is that insight is lacking on which moral values the military and general public consider important when Autonomous Weapon Systems are deployed.

human being is wrong as a matter of principle, not just because of the negative consequences this may bring [19], [20]. Another set of concerns has to do with the idea that the deployment of Autonomous Weapon Systems may make attribution of moral and legal responsibility more difficult, if not impossible altogether [21]–[26]. On a related note, it has been argued that the tendency for human beings to depend increasingly on computer systems for their decision-making can lead to a reduced sense of responsibility for the consequences of those decisions [27], [28]; indeed, in relation to human-operated drone operations it has also been argued that these operations not only create a physical distance, but also a moral distance as the face of the opponent becomes less visible, which eliminates a moral-psychological barrier for killing [29].

The past five years, a few public opinion surveys on Autonomous Weapon Systems have been conducted. Carpenter [30] surveyed how people in the United States feel about the idea of outsourcing targeting decisions to machines. The Open RoboEthics initiative surveyed public opinion on Autonomous Weapon Systems in a poll in 2015 [31] and issued a report. A worldwide survey on support for Autonomous Weapon Systems was conducted in 2017 [32]. However, to the best of our knowledge, the results were not published in an academic journal.

To recap, in the debate on Autonomous Weapon Systems, strong views and opinions are voiced. The Campaign to Stop Killer Robots [33] states for example on their website that: "Allowing life or death decisions to be made by machines crosses a fundamental moral line." With reference to the so-called Martens clause in International Law, Peter Asaro has suggested that the use of Autonomous Weapon Systems may be against "the dictates of public conscience." Moreover, many scholars have raised concerns that the use of (semi-) Autonomous Weapon Systems may negatively affect military personnel's well-being, moral integrity, and sense of responsibility; and that it may also create unjust anxiety and distress in the civilians potentially affected by them.

However, we found no literature or empirical studies on moral values that are related to Autonomous Weapon Systems or on what people consider to be the "fundamental moral line." Ethical concerns are studied in the related field of Human Operated drone operations [29], [34], but this research is not yet extended to the deployment of Autonomous Weapon Systems. Therefore, the knowledge gap is that insight is lacking on which moral values the military and general public consider important when Autonomous Weapon Systems are deployed in the near future.

The knowledge gap can be filled by studying known value theories to see which values people deem important in the deployment of Autonomous Weapon Systems. Well-established value theories are those of Schwartz [35], Friedman and Kahn, Jr. [36], and Beauchamp and Walters [37], but insight in how these relate to Autonomous Weapon Systems is lacking. Deriving the values that are most relevant in the context of the deployment of Autonomous Weapon Systems and comparing these values to those related to the current technology, that of Human Operated drones, will lead to insight into the underlying motives in the debate on Autonomous Weapon Systems and to greater understanding of the views that are expressed.

Definitions of Autonomous Weapon Systems

Autonomous Weapon Systems are an emerging technology and there is still no internationally agreed upon definition [38]. Even consensus on whether Autonomous Weapon Systems should be defined at all is lacking. Although some scholars provide definitions in their writings (Table 1), others caution against such a specification. NATO states that: "Attempting to create definitions for "autonomous systems" should be avoided, because by definition, machines cannot be autonomous in a literal sense" [39]. The United Nations Institute for Disarmament Research [40] is also cautious about providing a definition of Autonomous Weapon Systems, because they argue that the level of autonomy depends on the "critical functions of concern and the interactions of different variables" [41]. They state that one of the reasons for the differentiation of terms regarding Autonomous Weapon Systems is that sometimes things (drones or robots) are defined, but in other times a characteristic (autonomy), variables of concern (lethality or degree of human control), or usage (targeting or defensive measures) are drawn into the discussion and become part of the definition.

The various definitions of Autonomous Weapon Systems are listed in Table 1. Some authors use the term military robots which have a certain level of autonomy. As military robots can be viewed as a subclass of Autonomous Weapon Systems according to the classification of Royakkers and Orbons [42], we included them in the list of definitions. In our opinion the definition in the

report of the Advisory Council On International Affairs (38) captures the description of Autonomous Weapon Systems best from an engineering and military standpoint, because it takes predefined criteria into account and is linked to the military targeting process as the weapon will only be deployed after a human decision. Therefore, we will follow this definition and define Autonomous Weapon Systems as:

> "A weapon that, without human intervention, selects and engages targets matching certain predefined criteria, following a human decision to deploy the weapon on the understanding that an attack, once launched, cannot be stopped by human intervention" ((38)).

Value Theories

In contrast to the topic of Autonomous Weapon Systems, the concept of values has been studied extensively in the fields of moral philosophy and psychology. Moral philosophy has a long and rich history in examining values and in this field theoretical questions are asked to investigate the nature of value and goodness (46). Often a distinction is made between instrumental values, which means there is reason to favor it for its effect that can lead to good things (47), and intrinsic values, which "…is a kind of value such that when it is possessed by something, it is possessed by it solely in virtue of its intrinsic properties" (48). Although traditional moral philosophy is mainly concerned with theories of what "ought to be" and is in a strict sense unaffected by empirical results, scholars have in the past decades increasingly called for a more systematic study of the relationship between the abstract ethical theories and moral practice, especially in disciplines such as experimental philosophy and applied ethics (49). The focus of this study is to investigate, from an empirical perspective, which moral values relate to Autonomous Weapon Systems and how. Therefore, we chose not to start from theoretical theories of values in traditional moral philosophy, but rather turned to (moral) psychology and applied ethics, and in particular medical and military ethics, to start our research on the values in Autonomous Weapon Systems.

In psychology, values are used by people to justify their behaviors and define which type of behaviors are socially acceptable.

In psychology, values are differentiated from attitudes, needs, norms, and behavior in that they are a belief, that they lead to behavior that guides people, and that they are ordered in a hierarchy that shows the importance of the value over other values (35). Values are used by people to justify their behaviors and define which type of behaviors are socially acceptable (50). They are distinct from facts in

TABLE 1. Overview Definitions of Autonomous Weapon Systems.

Author (s)	Definition
AIV and CAVV [38]	*"A weapon that, without human intervention, selects and engages targets matching certain predefined criteria, following a human decision to deploy the weapon on the understanding that an attack, once launched, cannot be stopped by human intervention."*
Altmann, et al. [43]	Autonomous Weapon Systems are: "…*robot weapons that once launched will select and engage targets without further human intervention*."
Galliott [44]	Military robots are: "*a group of powered electro-mechanical systems, all of which have in common that they*: 1) *Do not have an onboard human operator;* 2) *Are designed to be recoverable (even though they may not be used in a way that renders them such); and,* 3) *In a military context, are able to exert their power in order to deliver a lethal or nonlethal payload or otherwise perform a function in support of a military force's objectives."*
Horowitz [45]	"*A weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator*."
Royakkers and Orbons [42]	Military Robots are "… *reusable unmanned systems for military purposes with any level of autonomy*."
Kuptel and Williams [39]	*"Machines are only "autonomous" with respect to certain functions such as navigation, sensor optimization, or fuel management."*
UNDIR [41]	The level of Autonomy depends on the "*critical functions of concern and the interactions of different variables*"

The values of blame, trust, harm, human dignity, confidence, expectations, support, fairness, and anxiety were deemed the most important in relation to Autonomous Weapon Systems.

that values do not only describe an empirical statement of the external world, but also adhere to the interests of humans in a cultural context [51]. Values can be used to motivate and explain individual decision-making and for investigation of human and social dynamics [52].

Many definitions of values exist. The existing definitions have been summarized by Cheng and Fleischmann [52] in their meta-inventory of values and they state that: "…values serve as guiding principles of what people consider important in life." Although a quite simple description, we think it captures the description of a value best, and therefore we will adhere to the definition of Cheng and Fleischmann [52] in our study.

Universal Values

Research suggests that people across cultures identify with basic values that can be considered as universal human values [50], [51], [53]. Although people can differ in which values they find more important, there seems to be a surprisingly high consensus across cultures on the hierarchical order of the values [50]. As part of their research some researchers created so-called value inventories, which are lists of items that can be used to categorize the analysis of human values and are often accompanied by a descriptive tool for discussions of these values [52]. The most common and well-studied value inventories are those of Schwartz [35], Friedman, Kahn, Borning, and Huldtgren [51], Beauchamp and Walters [37], and Graham *et al.* [53].

Values are not only described in theory from a psychological perspective as outlined in the previous paragraph but have also been practically implemented and used by means of applied ethics to professional domains. For example in the medical field, bioethics are used to describe values that are important as guiding principles for biomedical professionals, such as physicians, nurses, and health workers. Beauchamp and Walters [37] describe four values as basis for the framework of bioethics: 1) autonomy: acting intentionally without controlling influences that would mitigate against a voluntary act, 2) beneficence: providing benefits for society as a whole, 3) justice: being fair and reasonable, and 4) nonmaleficence: not intentionally imposing risk or harm upon another.

Values Related to Autonomous Weapon Systems

Values as described in the value theories above are not often explicitly mentioned in the literature on Autonomous Weapon Systems, but most studies discuss different values or related ethical issues. Two public reports of Human Rights Watch mention the lack of human emotion, accountability, responsibility, lack of human dignity, and harm as values related to Autonomous Weapon Systems [54], [55]. Sharkey and Suchman [56] state that the values of accountability and responsibility are important to consider in the design of robotic systems for military operations.

In the field of military ethics, Johnson and Axinn [57] list responsibility, reduction of human harm, human dignity, honor, and human sacrifice as values in their discussion on whether the decision to take a human life should be handed over to a machine or not. Cummings [58] in her case study of the Tactical Tomahawk missile, looks at the universal values proposed by Friedman and Kahn, Jr. [36] and states that next to accountability and informed consent, the value of human welfare is fundamental core value for engineers when developing weapons as it relates to the health, safety, and welfare of the public. She also mentions that the legal principles of proportionality and discrimination are important to consider in the context of weapon design. Proportionality refers to the fact that an attack is only justified when the damage is not considered to be excessive. Discrimination means that a distinction between combatants and noncombatants is possible [59]. Asaro [60] also refers to the principles of proportionality and discrimination and states that Autonomous Weapon Systems open up a moral space in which new norms are needed. Although he does not explicitly mention values in his argument, he does refer to the value of human life and the need for humans to be involved in the decision of taking a human life. Other studies primarily describe ethical issues, such as preventing harm, upholding human dignity, security, the value of human life, and accountability [40], [45], [61],] [62].

Based on this literature review, on a short exploratory online survey, and on expert interviews described in [63], we selected the values blame, trust, harm, human dignity, confidence, expectations, support, fairness, and anxiety to be incorporated in our study, because these values are mentioned most often in literature and because our respondents indicated that they deem these values most important in relation to Autonomous Weapon Systems.

Research Method

To evaluate the role of the values blame, trust, harm, human dignity, confidence, expectations, support, fairness, and anxiety on Autonomous Weapon Systems, we conducted two studies, the first on a military sample and the second on a sample consisting of civilians. We will report on these studies separately given that they concern different, nonrepresentative, samples. We will first describe the method of controlled experiments we used and the scenarios. Next, we will show the operationalization of the values, and we will conclude this section with a description of the sample.

Randomized Controlled Experiments

The method we used to conduct the two studies is called a randomized controlled experiment. Oehlert (64) mentions four reasons to create experiments: 1) they allow for direct comparisons between treatments of interest, 2) they can be designed to minimize any bias in the comparisons, 3) they can be designed to keep the error in the comparison small, and 4) we are in control of the experiments, which allows us to make stronger inferences about the nature of differences we observe and especially allows us to make inferences about causation. This last point distinguishes an experiment from an observational study. A treatment in this sense is used for the different procedures we would aim to compare. We use randomization in the studies to vary the order of the scenarios and the order of the questions posed to the respondents by means of a probabilistic scheme.

Scenario

Scenarios are used in the field of cognitive science as means to study moral judgement in randomized controlled experiments (65), (66). We created a scenario that describes a military operation in which a convoy is delivering supplies in a conflict area (see Appendix B). The convoy is being approached by a vehicle at high speed. This is a situation that is likely to happen during these types of operations (https://news.un.org/en/story/2019/01/1031342, https://www.bbc.com/news/world-middle-east-49394759, https://english.defensie.nl/latest/news/2019/08/28/a-look-at-the-defence-news-19-%E2%80%93-25-august), and military personnel needs to estimate the level of threat in order to decide to attack or not. We chose to focus on drones, as this is technology that is currently used by human operators and drones are already developed with autonomy by several companies, such as BAE systems (https://en.wikipedia.org/wiki/BAE_Systems_Taranis), Dassault Aviation (https://en.wikipedia.org/wiki/Dassault_nEURON), and Boeing (https://en.wikipedia.org/wiki/Boeing_Phantom_Ray).

The actions of Human Operated drones are perceived as having more respect for human dignity than Autonomous Weapon Systems.

Although these Autonomous drones not yet deployed in military operations, we think that is likely to happen within the next five years.

Operationalization Values Construct

To measure the moral values related to Autonomous Weapon Systems, the values are operationalized in nine constructs: blame, trust, harm, human dignity, confidence, expectations, support, fairness, and anxiety. Each of these variables was measured on a self-reported scale of: 0 (strongly disagree) – 100 (strongly agree). The corresponding questions can be found in Appendix A for exact wording. The analysis of the dependent variables is of an exploratory nature and the results are depicted

Appendix A

We used the following questions to measure moral values:

1) **Blame:** The drone is to blame for the action.
2) **Trust:** The drone can be trusted to take the correct actions in the future.
3) **Harm:** The actions of the drone caused harm.
4) **Human dignity:** The actions of the drone respect human dignity.
5) **Confidence:** I am confident that the drone will take the correct actions in the future.
6) **Expectations:** The actions of the drone are according to my expectations.
7) **Support:** I support the use of these type of drones by the military.
8) **Fairness:** The actions of the drone are fair.
9) **Anxiety:** The actions of the drone worry me.

Each of these variables was measured on a self-reported scale of: 0 (strongly disagree) – 100 (strongly agree).

Demographic Variables

We also added questions to collect demographic information on the respondents and added variables on: age, gender, education level, occupation, and nationality. In addition to these general demographic questions we asked if respondents had experience with Artificial Intelligence, if they worked with drones, and if they have been in a conflict zone.

by graphs (see Figures). Each graph shows the mean of the dependent variable on the y axis and the different scenarios on the x axis to point out the differences between the scenarios.

Sample

To determine the number of scenarios for the two studies, we performed power calculations to estimate the total number of participants that we would need based on the results of the pilot studies. Based on the power calculations (effect size 0.4, a desired statistical power of 0.8, and a probability level of 0.05) we aimed for a total of 200 responses and determined that we could run 3 scenarios.

Appendix B

The default scenario that we used in the study reads as follows:

> *A military convoy is on its way to deliver supplies to one of their units at a camp near Mosul in Iraq. The commander has ordered an autonomous drone to support the convoy in the air. The autonomous drone scans the surroundings for enemy threats and carries weapons for the defence of the convoy. When the convoy is at a three-mile distance from the camp, the autonomous drone detects a vehicle behind a mountain range that is approaching the convoy at high speed. The autonomous drone detects four people in the car with large weapon-shaped objects and identifies the driver of the vehicle as a known member of an insurgency group. The autonomous drone attacks the approaching vehicle which results in the death of all four passengers, but also causes collateral damage by killing five children that were playing nearby the road.*

The above scenario represents the neutral agency condition in which we do not provide any extra information on the weapon. In the human operated scenarios, we replace words autonomous drone with the words Human Operated drone and provided no extra information on the weapon (note that the words are highlighted in blue to show the distinction in this paper and the respondents in the survey were shown scenarios in black wording). In the high agency condition, we added the following phrase to describe the agency characteristics: "*The autonomous drone independently deliberates between a series of options, weighs the pros and cons, and decides to attack the approaching vehicle,….*" We purposely kept the changes to the scenarios to a minimum so that we can attribute different results to those changes and measure their effect.

Study 1 was distributed via the snowball method by e-mail with an anonymous link to approximately 40 military personnel who further distributed the survey. This method was used because we were not allowed to collect any personal information, such as e-mail or IP addresses. Study 1 resulted in 327 responses of which 239 were complete valid responses and usable after the data preprocessing. The 239 responses (227 male) of study 1 consist of Dutch military (149 respondents) and civilian (90 respondents) personnel working at the Dutch Ministry of Defense (MOD). The number of respondents per scenario ranged between 64 and 96. Study 2 was distributed via the crowdsourcing platform Amazon Mechanical Turk and 294 valid responses were collected. The 294 responses consist of civilian respondents (168 male) and the number of respondents per scenario ranged between 110 and 91.

Results

We will report the results of both studies by describing the results of the values blame, trust, harm, human dignity, confidence, expectations, support, fairness, and anxiety for both studies in a descriptive manner. The results of the values of human dignity, trust, anxiety, and blame can be viewed in Figs. 1–8.

Study 1

The actions of Human Operated drones are perceived as having more respect for human dignity than Autonomous Weapon Systems by military personnel and civilians working at the Dutch Ministry of Defense (Figure 1). The respondents in Study 1 are more anxious about the actions of Autonomous Weapon Systems than the actions of human operated drones (Figure 2). We also found that military personnel and civilians working at the Dutch Ministry of Defense have more trust (Figure 3)

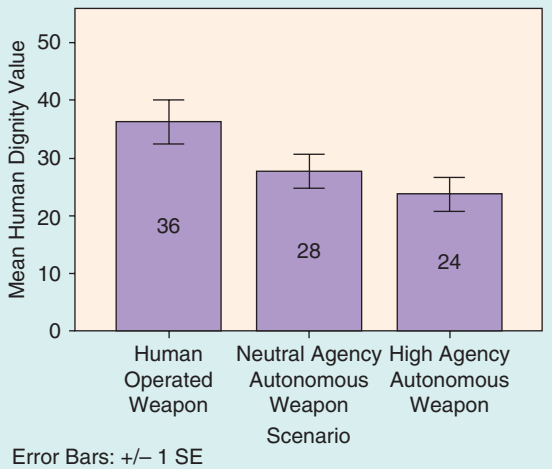

FIGURE 1. Human Dignity value study 1.

and confidence in, and support for human operated drones compared to Autonomous Weapon Systems. However, surprisingly they also assign more blame (Figure 4) to the actions of human operated drones than to the actions of Autonomous Weapon Systems. The perception of harm, fairness, and expectations of actions of human operated drones and Autonomous Weapon Systems are equal.

Study 2

The civilian respondents in Study 2 perceive the human dignity of the actions of human operated drones and Autonomous Weapon Systems equally (Figure 5). They are more anxious about the actions of Autonomous Weapon Systems in the future than of those taken by human operated drones (Figure 6). The respondents in Study 2 have an equal level of trust (Figure 7), expectations and confidence that human operated drones and Autonomous Weapon Systems will take the correct actions in the future. The actions of the human operated drone and Autonomous Weapon Systems are considered to cause equally much harm and are seen as equally fair. The civilian respondents in Study 2 assign more blame to the actions of Autonomous Weapon Systems than those of human operated drones (Figure 8). They have more support for human operated drones than for Autonomous Weapon Systems, especially when human operated drones are compared to the high agency scenario of Autonomous Weapon Systems.

Conclusion and Discussion

Our study provides an overview of the various definitions of Autonomous Weapon Systems that are currently

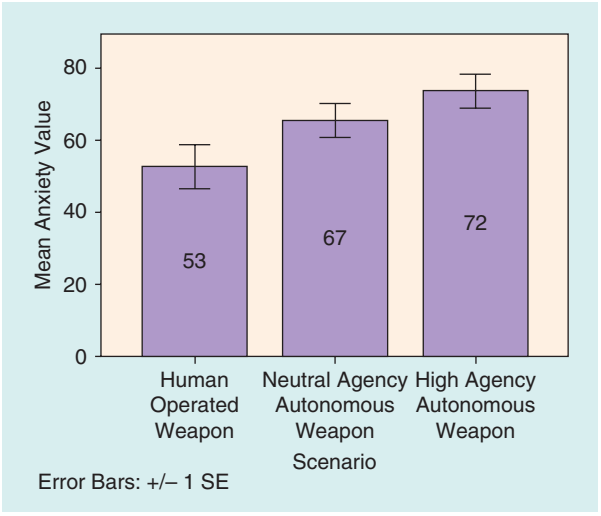

FIGURE 2. Anxiety value study 1.

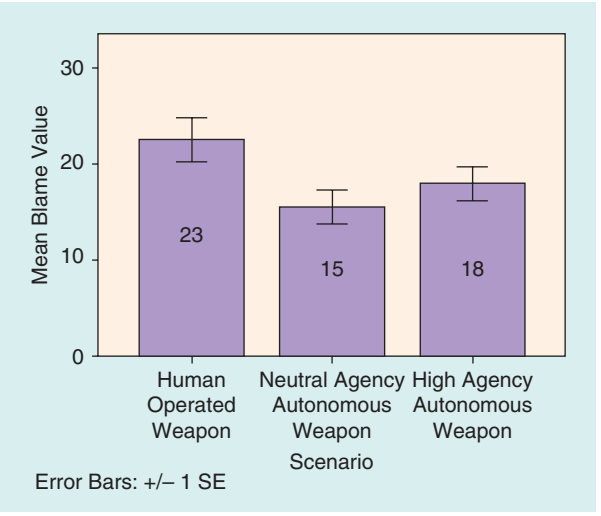

FIGURE 4. Blame value study 1.

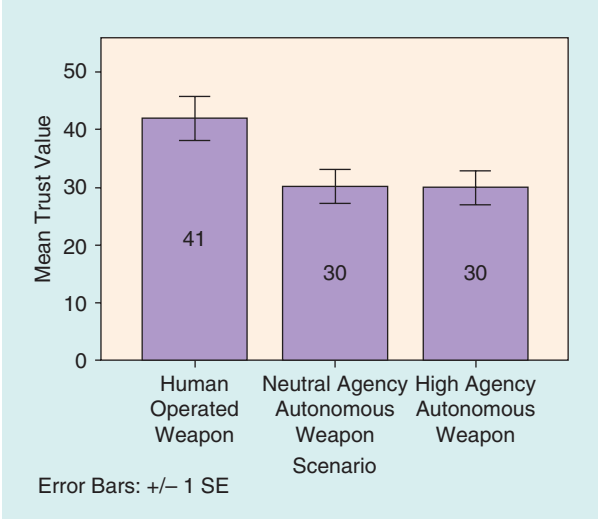

FIGURE 3. Trust value study 1.

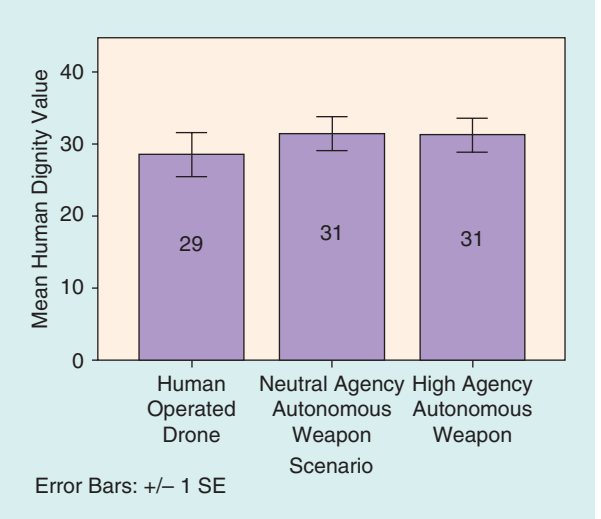

FIGURE 5. Human dignity value study 2.

used in literature and shows that there is no agreement on one single definition yet. We have identified several values that people associate with Autonomous Weapon Systems. The overview is derived from both validated value theories and from experts who are involved in the debate on Autonomous Weapon Systems or work in the military domain. We selected the values blame, trust, harm, human dignity, confidence, expectations, support, fairness, and anxiety based on our literature review, an exploratory survey, and expert interviews. The results provide insight into how a small sample of military personnel and civilians perceive these values for both the human operated drone, as current technology, and for Autonomous Weapon Systems, as future technology. Although we cannot compare both studies directly, we believe that our empirical results substantiate some of the views and opinions on Autonomous Weapon Systems affecting human responsibility in the current discourse. These could be used to move forward in the ethical debate on Autonomous Weapon Systems.

Further common ground can be found on the values of human dignity and anxiety. Our results show that military personnel and civilians are more anxious about the deployment Autonomous Weapon Systems than the deployment of human operated drones. Military personnel and civilians working at the Dutch MOD also perceive Autonomous Weapon Systems to have less respect for the dignity of human life than human operated drones. This effect was less apparent in Study 2, which consisted primarily of civilians. Human dignity and anxiety are two values that are mentioned often in the public discourse, so in our opinion it would be essential to address these values when debating the ethics of the deployment of Autonomous Weapon Systems.

Limitations

As this study is to our knowledge one of the first to gather empirical data of moral values related to Autonomous Weapon Systems, we had to derive these related values ourselves, and we can identify the following limitations for our research. First, the operationalization of the values was derived from a categorization of literature describing values. The questions that we used were based on heuristics and we did not test if these questions were correct. This selection method would be hard to replicate by others and affects the reproducibility and internal validity of the study.

Secondly, the study was conducted using samples based in northwest Europe and the U.S. Although we tried to incorporate universal and well-studies values, such as the value inventory of Friedman *et al.*, (51) and BioEthics values (37), this study takes a Western view on the moral values related to the deployment of Autonomous Weapon Systems. Also, anxiety is incorporated as a value in this

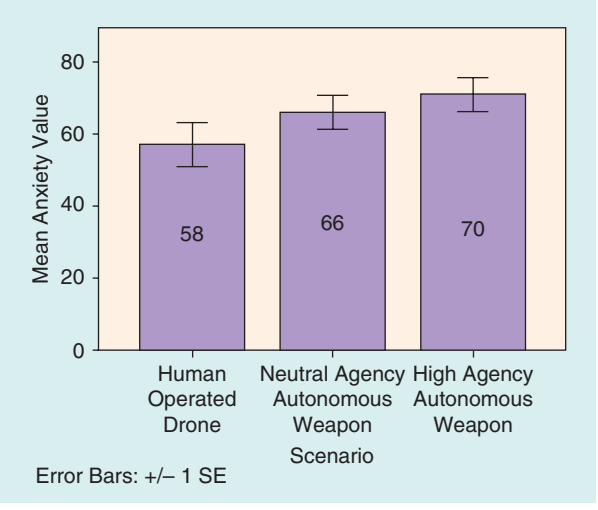

FIGURE 6. Anxiety value study 2.

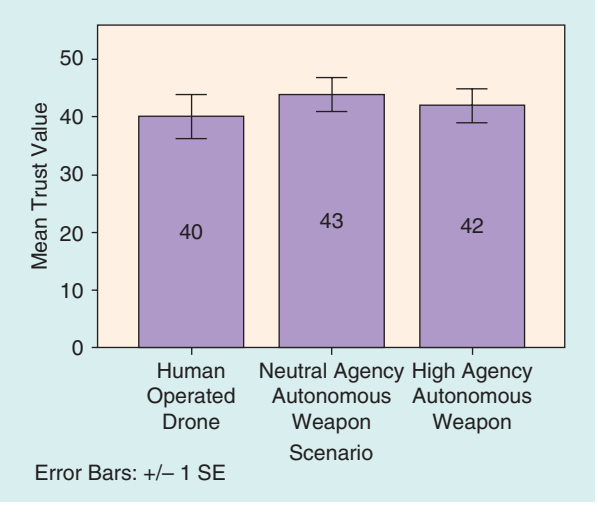

FIGURE 7. Trust value study 2.

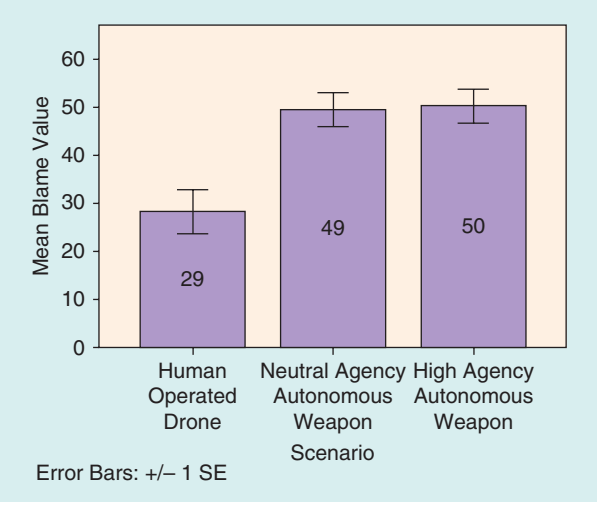

FIGURE 8. Blame value study 2.

study, because we were curious to see how Autonomous Weapon Systems as future technology would compare on this aspect to the current technology of human operated drones, but in retrospect anxiety is an emotion instead of a moral value. Therefore, this study should be regarded as exploratory and requires follow-up studies in different cultures to see if the results will hold.

Recommendations for Further Research

Given the exploratory nature of this study and the limitations mentioned above, we suggest several recommendations for further research. The first is to run follow-up studies to investigate if the results on moral values related to Autonomous Weapon Systems hold in different samples and in different cultures that have a different attitude toward drones, robots, and AI technology than we as the authors in the Western world have.

Another approach to consider which values are relevant in the deployment of Autonomous Weapon Systems is the specification of values into design requirements to see how these values would be perceived in practice. This can be made visible by means of a value hierarchy (67), which is a hierarchical structure of values, norms, and design requirements. A value hierarchy will to make the value judgements that are required for the translation explicit, transparent, and debatable. To do so, the values that are described in the natural language will need to be translated to "formal values in a formal language" (68). One way of formalizing values into norms would be to use a convention of rules represented as: "*X* counts as *Y*" or "*X* counts as *Y* in context *C*" (69). The explicitness of values in formal rules and visibility in a value hierarchy would allow for critical reflection in debates and pinpoint the value judgements that are disagreed on.

In this study we provided empirical results to fill the knowledge gap on moral values related to Autonomous Weapon Systems that aim to understand the perception of people regarding the introduction of Autonomous Weapon Systems. The recommendations for further research, being a follow-up study for the moral values in related AI fields, and the creation of a value hierarchy, will be conducted at our research group at Delft University of Technology by our graduate students. Continuing our research effort in this field would help better define the "fundamental moral line" that should not be crossed (33). The results of these studies reveal a common ground regarding the moral values of human dignity and anxiety, pertaining to the introduction of Autonomous Weapon Systems. The concerns for Autonomous Weapon Systems mentioned in the public debate deserve to be carefully considered in future debates on the ethics of the design and deployment of Autonomous Weapon Systems. We believe that our empirical results substantiate some of the views and opinions on Autonomous Weapon Systems affecting human responsibility in the current discourse; these results could be used to move forward in the ethical debate on Autonomous Weapon Systems.

Acknowledgment

The authors would like to thank Sydney Levine, Iyad Rahwan of the Scalable Cooperation Group at the Media Lab of M.I.T., and Jeroen van der Hoven of Delft University of Technology for their insightful comments, sharp discussions, and supervision of the research.

Author Information

Ilse Verdiesen and ***Filippo Santoni de Sio*** are with Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands. ***Virginia Dignum*** is with Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands, and with Umeå University, 901 87 Umeå, Sweden. Email: e.p.verdiesen@tudelft.nl.

References

(1) S.J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Malaysia: Pearson, 2016.
(2) J.J. Bryson, "Patiency is not a virtue: AI and the design of ethical systems," in *2016 AAAI Spring Symposium Series*, 2016.
(3) L. Floridi and J.W. Sanders, "On the morality of artificial agents," *Minds and machines*, vol. 14, no. 3, pp. 349-379, 2004.
(4) H.M. Roff, "Weapons autonomy is rocketing," *Foreign Policy*, Sept. 28, 2016; http://foreignpolicy.com/2016/09/28/weapons-autonomy-is-rocketing/.
(5) U.S. Air Force, "Unconscious US F-16 pilot saved by Auto-pilot," *YouTube*, Sept. 15, 2016; https://www.youtube.com/watch?v=RQBFJkMnBcA, accessed Feb. 19, 2017.
(6) J. Carpenter, *Culture and Human-Robot Interaction in Militarized Spaces: A War Story*. Taylor & Francis, 2016.
(7) C. Heyns, "Report of the Special Rapporteur on Extrajudicial, Summary, and Arbitrary Execution, United Nations Human Rights Council," *United Nations 23rd Session*, vol. 9, Apr. 2013.
(8) Campaign to Stop Killer Robots, "Campaign to Stop Killer Robots," 2018; https://www.stopkillerrobots.org/, accessed July 15, 2017.
(9) Office of the High Commissioner, "A/HRC/31/66, Joint report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association and the Special Rapporteur on extrajudicial, summary or arbitrary executions on the proper management of assemblies," *United Nations*, 2016.
(10) J. Kaag and W. Kaufman, "Military frameworks: Technological know-how and the legitimization of warfare," *Cambridge Rev. Int. Affairs*, vol. 22, no. 4, pp. 585-606, 2009.
(11) M. Rosenberg and J. Markoff, "The Pentagon's 'Terminator Conundrum': Robots that could kill on their own," *New York Times*, 2016.
(12) B. Burridge, "UAVs and the dawn of post-modern warfare: A perspective on recent operations," *RUSI J.*, vol. 148, no. 5, pp. 18-23, 2003.
(13) P. Asaro, "How just could a robot war be," *Current Issues In Computing and Philosophy*, pp. 50-64, 2008.
(14) A. Krishnan, "Killer robots," *Legality and Ethicality of Autonomous Weapons*. Farnham, U.K.: Ashgate, 2009.
(15) M. Guarini and P. Bello, "Robotic warfare: Some challenges in moving from noncivilian to civilian theaters," *Robot Ethics: The Ethical and Social Implications of Robotics*, vol. 129, p. 136, 2012.
(16) N. Sharkey, "Automated killers and the computing profession," *Computer*, vol. 40, no. 11, 2007.
(17) N. Sharkey, "Killing made easy: From joysticks to politics," *Robot Ethics: The Ethical and Social Implications of Robotics*, pp. 111-128, 2012.

[18] N. Sharkey, "The automation and proliferation of military drones and the protection of civilians," *Law, Innovation and Technology,* vol. 3, no. 2, pp. 229-240, 2011.
[19] M. Wagner, "The dehumanization of international humanitarian law: Legal, ethical, and political implications of autonomous weapon systems," *Vand. J. Transnat'l L.,* vol. 47, p. 1371, 2014.
[20] W. Wallach, "Terminating the terminator: What to do about autonomous weapons," *Science Progress,* vol. 29, 2013.
[21] A. Matthias, "The responsibility gap: Ascribing responsibility for the actions of learning automata," *Ethics and Information Technology,* vol. 6, no. 3, pp. 175-183, 2004.
[22] R. Sparrow, "Killer robots," *J. Applied Philosophy,* vol. 24, no. 1, pp. 62-77, 2007.
[23] M. Santoro, D. Marino, and G. Tamburrini, "Learning robots interacting with humans: From epistemic risk to responsibility," *AI & Society,* vol. 22, no. 3, pp. 301-314, 2008.
[24] Human Rights Watch, "Mind the gap: The lack of accountability for killer robots," 2015.
[25] F. Santoni de Sio and E. Di Nucci, "Drones and responsibility: Mapping the field," in *Drones and Responsibility*. Routledge, 2016, pp. 11-24.
[26] F. Santoni de Sio and J. Van den Hoven, "Meaningful human control over Autonomous Systems: A philosophical account," *Frontiers in Robotics and AI,* vol. 5, p. 15, 2018.
[27] M.L. Cummings, "Automation and accountability in decision support system interface design," 2006.
[28] D. Saxon, "Autonomous drones and individual criminal responsibility," in *Drones and Responsibility*. Routledge, 2016, pp. 27-56.
[29] M. Coeckelbergh, "Drones, information technology, and distance: Mapping the moral epistemology of remote fighting," *Ethics and Information Technology,* vol. 15, no. 2, pp. 87-98, 2013.
[30] C. Carpenter. "How scared are people of "killer robots" and why does it matter?" *Open Democracy*, Jul. 4, 2014; https://www.opendemocracy.net/charli-carpenter/how-scared-are-people-of-%E2%80%9Ckiller-robots%E2%80%9D-and-why-does-it-matter, accessed Oct. 3, 2018.
[31] Open Roboethics Initiative, "The Ethics and Governance of Lethal Autonomous Weapons Systems: An International Public Opinion Poll," 2015; http://www.openroboethics.org/laws_survey_released/, accessed Jul. 15, 2017.
[32] "Three in ten Americans support using Autonomous Weapons," *Ipsos*, Feb. 7, 2017; https://www.ipsos.com/en-us/news-polls/three-ten-americans-support-using-autonomous-weapons, accessed Oct. 03, 2018.
[33] "The Problem," *Campaign to Stop Killer Robots*, http://www.stopkillerrobots.org/the-problem/, accessed Jul. 15, 2017.
[34] B.J. Strawser, "Moral predators: The duty to employ uninhabited aerial vehicles," in *Handbook of Unmanned Aerial Vehicles*. Springer, 2010, pp. 2943-2964.
[35] S.H. Schwartz, "Are there universal aspects in the structure and contents of human values?," *J. Social Issues,* vol. 50, no. 4, pp. 19-45, 1994.
[36] B. Friedman and P.H. Kahn, Jr, "Human values, ethics, and design," *The Human-Computer Interaction Handbook,* pp. 1177-1201, 2003.
[37] T.L. Beauchamp and L.R. Walters, *Contemporary Issues in Bioethics*. Wadsworth, 1999.
[38] "Autonomous weapon systems:The need for meaningful human control," *Advisory Council on International Affairs,* no. 97, Oct. 2015 [Online] Available: http://aiv-advice.nl/8gr.
[39] A. Kuptel and A. Williams, "Policy guidance: Autonomy in defence systems," 2014.
[40] United Nations, *The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values,* 2015 [Online] Available: http://www.unidir.org/files/publications/pdfs/considering-ethics-and-social-values-en-624.pdf.
[41] *Framing Discussions on the Weaponization of Increasingly Autonomous Technologies,* 2014 [Online] Available: http://www.unidir.org/files/publications/pdfs/framing-discussions-on-the-weaponization-of-increasingly-autonomous-technologies-en-606.pdf.
[42] L. Royakkers and S. Orbons, "Design for values in the armed forces: Nonlethal weapons weapons and military military robots robot," *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains,* pp. 613-638, 2015.
[43] J. Altmann, P. Asaro, N. Sharkey, and R. Sparrow, "Armed military robots: Editorial," *Ethics and Information Technology,* vol. 15, no. 2, p. 73, 2013.
[44] J. Galliott, *Military Robots: Mapping the Moral Landscape*. Ashgate, 2015.
[45] M.C. Horowitz, "The ethics and morality of robotic warfare: Assessing the debate over Autonomous Weapons," *Daedalus,* vol. 145, no. 4, pp. 25-36, 2016.
[46] M. Schroeder. "Value theory," *Stanford Encyclopedia of Philosophy*, Feb. 5, 2008, rev. Jul. 28, 2016; https://plato.stanford.edu/entries/value-theory/, accessed Aug. 7, 2017.
[47] T. Rønnow-Rasmussen, "Instrumental values–Strong and weak," *Ethical Theory and Moral Practice,* vol. 5, no. 1, pp. 23-43, 2002.
[48] B. Bradley, "Two concepts of intrinsic value," *Ethical Theory and Moral Practice,* vol. 9, no. 2, pp. 111-130, 2006.
[49] M. Alfano and D. Loeb, "Experimental moral philosophy," *Stanford Encyclopedia of Philosophy*, Mar. 19, 2014; https://plato.stanford.edu/entries/experimental-moral/, Aug. 7, 2017.
[50] S.H. Schwartz, "An overview of the Schwartz theory of basic values," *Online Readings in Psychology and Culture,* vol. 2, no. 1, p. 11, 2012.
[51] B. Friedman, P.H. Kahn, Jr., A. Borning, and A. Huldtgren, "Value sensitive design and information systems," in *Early Engagement and New Technologies: Opening Up the Laboratory,* Springer, 2013, pp. 55-95.
[52] A.S. Cheng and K.R. Fleischmann, "Developing a meta-inventory of human values," in *Proc. American Society for Information Science and Technology,* vol. 47, no. 1, pp. 1-10, 2010.
[53] J. Graham *et al.*, "Moral foundations theory: The pragmatic validity of moral pluralism," 2012.
[54] B. Docherty, *Losing Humanity: The Case Against Killer Robots*, 2012.
[55] B. Docherty, *Mind the Gap: The Lack of Accountability for Killer Robots*, Human Rights Watch, 2015.
[56] N. Sharkey and L. Suchman, "Wishful mnemonics and autonomous killing machines," in *Proc. AISB*, 2013, vol. 136, pp. 14-22.
[57] A.M. Johnson and S. Axinn, "The morality of autonomous robots," *J. Military Ethics,* vol. 12, no. 2, pp. 129-141, 2013.
[58] M.L. Cummings, "Integrating ethics in design through the value-sensitive design approach," *Science and Engineering Ethics,* vol. 12, no. 4, pp. 701-715, 2006.
[59] T. Hurka, "Proportionality in the morality of war," *Philosophy & Public Affairs,* vol. 33, no. 1, pp. 34-66, 2005.
[60] P. Asaro, "On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making," *Int. Rev. Red Cross,* vol. 94, no. 886, pp. 687-709, 2012.
[61] J.I. Walsh and M. Schulzke, "The Ethics of Drone Strikes: Does Reducing the Cost of Conflict Encourage War?," *DTIC Document*, 2015.
[62] A.P. Williams, P.D. Scharre, and C. Mayer, "Developing Autonomous Systems in an ethical manner," in *Autonomous Systems: Issues for Defence Policymakers*: NATO Allied Command Transformation (Capability Engineering and Innovation), 2015.
[63] I. Verdiesen, "Agency perception and moral values related to Autonomous Weapons: An empirical study using the Value-Sensitive Design approach," *Semantic Scholar*, 2017.
[64] G.W. Oehlert, *A First Course in Design and Analysis of Experiments*, 2010.
[65] J.F. Kominsky, J. Phillips, T. Gerstenberg, D. Lagnado, and J. Knobe, "Causal superseding," *Cognition,* vol. 137, pp. 196-209, 2015.
[66] F. Cushman and L. Young, "Patterns of moral judgment derive from nonmoral psychological representations," *Cognitive Science,* vol. 35, no. 6, pp. 1052-1075, 2011.
[67] I. Van de Poel, "Translating values into design requirements," in *Philosophy and Engineering: Reflections on Practice, Principles and Process*. Springer, 2013, pp. 253-266.
[68] H. Aldewereld, V. Dignum, and Y.-h. Tan, "Design for values information and communication technologies in software development," *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains,* pp. 831-845, 2015.
[69] J.R. Searle, *The Construction of Social Reality*. Free Press, 1997.