Design Requirements for a Moral Machine for Autonomous Weapons

Ilse Verdiesen[1(✉)], Virginia Dignum[1], and Iyad Rahwan[2]

[1] Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands
e.p.verdiesen@tudelft.nl
[2] MIT Medialab, 75 Amherst Street, Cambridge, MA 02139, USA

Abstract. Autonomous Weapon Systems (AWS) are said to become the third revolution in warfare. These systems raise many questions and concerns that demand in-depth research on ethical and moral responsibility. Ethical decision-making is studied in related fields like Autonomous Vehicles and Human Operated drones, but not yet fully extended to the deployment of AWS and research on moral judgement is lacking. In this paper, we propose design requirements for a Moral Machine (Similar to http://moralmachine.mit.edu/) for Autonomous Weapons to conduct a large-scale study of the moral judgement of people regarding the deployment of this type of weapon. We ran an online survey to get a first impression on the importance of six variables that will be implemented in a proof-of-concept of a Moral Machine for Autonomous Weapons and describe a scenario containing these six variables. The platform will enable large-scale randomized controlled experiments and generate knowledge about people's feelings concerning this type of weapons. The next steps of our study include development and testing of the design before the prototype is upscaled to a Massive Online Experiment.

Keywords: Autonomous weapons · Ethical decision-making
Moral acceptability · Moral machine

1 Introduction

As the reach and capabilities of Artificial Intelligence (AI) systems increases, there is also an increasing awareness of the ethical, legal and societal impact of the potential actions and decisions of these systems. Many are calling for guidelines and regulations that can ensure the responsible design, development, implementation, and policy of AI in general, and Autonomous Weapon Systems (AWS) specifically [1]. Others are calling for a ban on these weapons and there is an increasing awareness that international regulations are long due [2]. Many experts agree that AWS that kill without human intervention are "morally wrong" and that their use should be controlled under the 1983 Convention on Certain Conventional Weapons (CCW)[1] in order to become part of International Humanitarian Law [3]. However, banning this autonomous technology for

[1] cf. https://ihl-databases.icrc.org/ihl/INTRO/500?OpenDocument.

B. Gallina et al. (Eds.): SAFECOMP 2018 Workshops, LNCS 11094, pp. 494–506, 2018.
https://doi.org/10.1007/978-3-319-99229-7_44

military use may not be practical given that similar technologies could well be dual-used as they are also available in the commercial sector. The CCW met late 2017 in Geneva and started a substantive review of relevant developments related to AWS. Of particular relevance to this discussion are issues regarding the level of autonomy, embodiment (including "killer robots", human enhancement, and virtual machines), validation and verification, and safety. Complementary to this discussion, is equally important to identify how the (civilian) public perceives AWS and their use, both in defense and attack situations.

The defense industry has always been one of the main areas of application of AI, and is also the area in which responsibility and ethical impact are the most salient. AWS are weapon systems equipped with Artificial Intelligence (AI) reasoning mechanisms. Such weapons have the capability to select and engage targets without human intervention. In the complex discussion of how the development of AWS should be controlled, there is a need to give full consideration, not only to national and commercial interests, but most importantly to have a profound understanding of the feelings people associate with this type of weapon and their perception of the moral acceptability of its use.

These considerations should lead setting regulation that can ensure that the development and use of AWS is controlled and that there are sufficient powerful monitoring mechanisms and institutions that can control this type of technology in all its application domains (military and commercial). Towards this aim, it is necessary to understand what autonomy in the context of AWS really means, how it relates to meaningful human control, and how people respond to the potential use of these weapons in different situations, for attack or defense, or with or without collateral damage.

A growing body of researchers is focusing on responsible design of AI, which incorporates social and ethical values, to prevent undesirable societal outcomes of this technology. Several proposals for principles to describe Responsible AI exist, including the Asilomar principles[2], ART - *Accountability*, *Responsibility* and *Transparency* [4], and the IEEE Ethically Aligned Design recommendations[3]. *Accountability* refers to the justification of the actions taken by the AI, *Responsibility* ensures appropriate data governance by establishing a chain of responsibility amongst stakeholders. *Transparency* is concerned with describing and reproducing the decisions the AI makes and adepts to its environment [5]. These works however, refer to AI in general and are as such to not provide specific guidance to the issue of AWS.

Autonomous systems are increasingly deployed on the battlefield [6]. Autonomous systems can have many benefits in the military domain, for example when the autopilot of the F-16 prevents a crash [7]. Other advantages are that Autonomous systems and AWS can act as force multipliers, meaning that less military personnel is needed for a mission which in turn is more efficient and could lead to long-term savings [8]. Another advantage would be that decisions of AWS would not be impaired by human emotions and could process much more incoming sensory information at a higher speed than humans are able to [9].

2 https://futureoflife.org/ai-principles.
3 https://ethicsinaction.ieee.org/.

Many arguments have been raised opposing AWS. The 'Campaign Stop Killer Robots' [2] states for example on their website that: '*Allowing life or death decisions to be made by machines crosses a fundamental moral line. Autonomous robots would lack human judgment and the ability to understand context.*'. The United Nations are also voicing their concerns and state 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*' [10]. The deployment of AWS on the battlefield without direct human oversight is not only a military revolution according to Kaag and Kaufman [11], but can also be considered a moral one. As large-scale deployment of AI on the battlefield seems unavoidable [12] the research on ethical and moral responsibility is imperative.

Ethical decision-making in AI and robots is an emerging field and scholars are studying moral judgment related to these technologies. For example, Malle [13] proposes a framework combining the (until recently) separate fields of *robot ethics*, in which ethical questions about the design, deployment and treatment of robots by humans are addressed, and *machine morality*, which is concerned with questions about the moral capacities of a robot and how these should be computationally implemented. Cointe, Bonnet and Bossier [14] propose a model in which an agent can judge the ethical aspects of its own behaviour and that of other agents in a multi-agent system. The model describes an Ethical Judgment Process (EJP) which allows agents to evaluate the behaviour of other agents. Bonnefon, Shariff and Rahwan [15] have studied the ethical decision an Autonomous Vehicle has to make, being self-protection or utilitarian, when confronted with pedestrians on the road. In this research, the Moral Machine at MIT is used to gain insight in how people judge on scenarios with an Autonomous Vehicle to see how their moral judgment compares to those of other people.

Ethical concerns have also been studied in the related field of Human Operated drone operations. Coekelbergh [16] argues that drone operations not only create a physical distance, but also a moral distance as the face of the opponent becomes less visible which eliminates the moral-psychological barrier for killing. Another ethical concern according to Strawser [17] is that due to the remote distance to the battlefield human operators can experience cognitive dissonance in which the war feels more like a video game than reality. These ethical concerns are refuted by Strawser [17] as he states that due to the increased distance the human operators have more time to evaluate a target, because their own safety is not at risk, but he also argues that more empirical research needs to be done to assess the psychological effects of the large distance to the battlefield.

We found that studies on ethical decision-making and moral judgement relating to AWS are currently lacking. Several scholars like Malle [13], Cointe et al. [14] and Bonnefon et al. [15] are studying ethical decision-making in AI and robots or Human Operated drones [16, 17], but their research has not yet been extended to the deployment of AWS. Therefore, we propose a design for a Moral Machine for Autonomous Weapons to conduct a large-scale study of the moral judgment of people on the deployment of this type of weapons. We build on the concept of the Moral Machine, that was developed by the Scalable Cooperation group of the Media Lab at MIT [18].

We will focus on the deployment of AWS in the near future, which we define as: *within the next 5 years* as these type of weapons are not yet deployed on the battlefield.

This entails that we will not study weapons equipped with Artificial General Intelligence or futuristic technology that is not possible to construct yet, but we will focus on technology that is currently being developed, specifically drones with autonomous targeting capabilities.

The paper is structured as follows; Sect. 2 first defines Autonomy and AWS and describes the Moral Machine for Autonomous Vehicles. In Sect. 3 we delineate our proposal for the Moral Machine for Autonomous Weapons. We present the results of an online survey that we ran to get a first impression of the importance of the variables and specify the scenarios and its variables. We conclude in Sect. 4 with a discussion on the limitations of our work and discuss recommendations for further research in Sect. 5.

2 Related Work

This section first discusses fundamental work on Autonomy prior to defining AWS. We further provide a description of the existing Moral Machine for Autonomous Vehicles, designed by the Scalable Cooperation group of the Media Lab at MIT [18], which we use as basis for our work.

2.1 Autonomy

Autonomy lays at the core of Artificial Intelligence. However, it is a most undefined and misunderstood concept. Often, autonomy is taken to be an absolute property of the system, i.e. an agent is either autonomous- capable of performing tasks in the world by itself, without explicit human control - or not. However, autonomy should be seen as a relational notion involving three classes of entities [19]: (i) the main subject, x whose autonomy is being considered/evaluated; (ii) a function/action/goal μ that must be realized or maintained by the main subject and on which the autonomy is evaluated; (iii) a secondary subject y (which may be a plurality of subjects) with respect to whom the main subject should be considered autonomous given the specified function/action/goal. That is, stating "x is autonomous" is meaningless as it does not indicate the object and recipient of that autonomy, whereas the statement "x is autonomous about μ with respect to y" provides enough characterization of the focus and context of autonomy. E.g. in the case x is a self-driving car (x), autonomy means that x can autonomously decide on the travel route (μ) given a destination set by its user (y).

On the basis of these relationships and the nature of the entities, several degrees and types of autonomy can be identified, including executive autonomy (i.e. the agent is autonomous in its means rather than its ends or goals; as in the self-driving car example above), goal autonomy (i.e. the agent is endowed with goals, or ends, of its own), or social autonomy (i.e. the agent is self-sufficient meaning that it can execute its goals by itself without other agents or resources) [19].

In the AWS arena, an important concept is that of *meaningful human control*, which signifies control over the selection and engagement of targets, that is, the "critical functions" of a weapon, including where and how weapons are used; what or whom they are used against; and the effects of their use [20]. In terms of the different types of autonomy above, this indicates that, at most, AWS can have executive autonomy but never the autonomy to decide on its own goals nor the moment of action. Human oversight is required to set goals and to monitor execution by the machine.

2.2 Autonomous Weapon Systems

Although the debate on AWS has drawn a lot of attention in the recent years and a few public opinion surveys have been conducted [21, 22], we found that the topic in not well delineated in the academic literature. AWS are an emerging technology and there is still no internationally agreed upon definition [23]. Even consensus whether AWS should be defined at all is lacking. Although some scholars provide definitions, 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*.' [24:10]. The United Nations Institute for Disarmament Research (UNDIR) is also cautious about providing a definition of AWS, because they argue that the level of autonomy depends on the '*critical functions of concern and the interactions of different variables*' [25:5]. They state that one of the reasons for the differentiation of terms regarding AWS 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. In our opinion, the definition in the report of the Advisory Council on International Affairs (AIV & CAVV) captures the description of AWS 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 AWS 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.'* [23:11].

2.3 Moral Machine for Autonomous Vehicles

Ethical decision making in Autonomous Vehicles is studied by Bonnefon et al. [15] by means of the Moral Machine. The original Moral Machine is a '*…platform for gathering data on human perception of the moral acceptability of decisions made by autonomous vehicles faced with choosing which humans to harm and which to save.'* [26:42–43]. The website has three modes for users: (1) a *Judge* mode in which users can decide the outcome for 13 series of scenarios, (2) a *Design* mode in which users can design their own scenarios, and (3) a *Browse* mode in which users can view the scenarios of others. The main feature is the Judge mode in which users choose between two scenarios that contain different variables of: *Characters* {gender, social value, age, species, fitness, utilitarianism}, *Interventionism* {omission, commission}, *Relationship to vehicle* {driver, pedestrian} and *Concern for law* {legal action, illegal action}. The Moral Machine is set-up as a Massive Online Experiment (MOE) [27] which is aimed at recruiting a large sample pool with a diverse background in a short amount of time at a low cost. These features are clear advantages of a MOE, but the downside is that the conditions are hard to control, as users can take surveys multiple times and are self-selected, which means that they can join the experiment and drop out whenever they want [26]. By May 2017, approximately 3 million users over 160 countries assessed over 30 million scenarios making it one of the biggest large-scale moral judgment tools that exist. The Moral Machine for Autonomous Vehicles is a novel tool to conduct a MOE which generates a vast amount of data in a short time. However, being novel, no report of substantial impact of the methodology could be found. Nevertheless, the Moral Machine concept could provide much insight in the moral judgement regarding the deployment of AWS. As empirical data on this insight is currently lacking, we propose to apply the Moral Machine concept to the domain of AWS and drafted design requirements for a Moral Machine for Autonomous Weapons.

3 Moral Machine for Autonomous Weapons

AWS are a sensitive topic and it invokes a primary response of anxiety and unease with people. In our opinion, it would not be prudent to develop an open platform to gather data on a large-scale at first, because an open platform could attract negative sentiment and unwanted actions which would be counterproductive for our research. For example, people creating scenarios in which certain groups of people, such as women or Muslims, are specifically targeted. Therefore, it is advisable to take a more step-wise approach in scaling up to a Massive Online Experiment.

3.1 Method

To run a large-scale follow-up study of the moral judgments of people regarding AWS a randomized controlled experiment with a limited set of conditions and sample should be set-up. Four reasons to use experiments as research technique are mentioned by Oehlert [28]: (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 we can make inferences about causation. A treatment in this sense are the different scenarios we would aim to compare. It is important that the effects of a treatment can only attributed to one cause that can be measured and cannot be attributed to multiple causes which is also referred to as confounding. Randomisation helps ensure that participants are assigned to a scenario by chance and not based on pre-existing features, such as time when the survey is taken or location.

3.2 Survey

To get a first impression on the importance of variables to be incorporated in the Moral Machine, we ran an online survey with 209 participants (142 males) on Amazon

Mechanical Turk. These six variables are inspired by the Moral Machine for Autonomous Vehicles and based on the researchers' intuition. The respondents were paid a small fee ($0.50) and were asked for each of the six variables: '*Please indicate to what extent the {variable X} is relevant to the moral decision-making on Autonomous Weapons:*'. Each of the items was measured on a self-reported scale ranging from '0 = not at all' to '100 = very important'. The results (Fig. 1. Results survey) indicate that the *Outcome* ($M = 70.25$, $SD = 28.05$) and *Type of Mission* ($M = 68.09$, $SD = 26.82$) are considered most important, but the other variables *Type of Weapon* ($M = 66.44$, $SD = 28.06$), *Type of Character* ($M = 62.83$, $SD = 30.24$), *Number of Characters* ($M = 62.52$, $SD = 28.70$) and *Location* ($M = 63.08$, $SD = 29.21$) are also relevant.

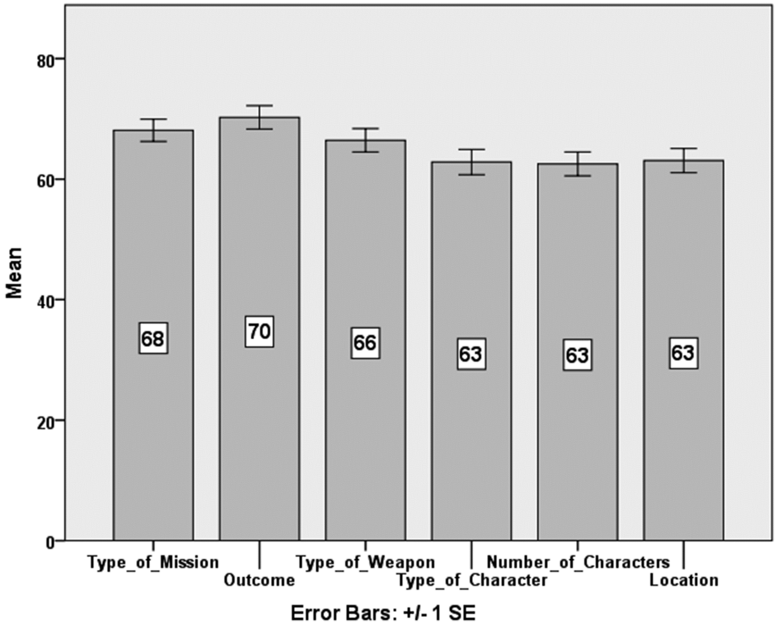

Fig. 1. Results survey

In addition to the questions on the variables we asked the respondents to: '*Please list one other aspect that you consider important for the moral decision-making on Autonomous Weapons:*'. Based on this open question, the following items are relevant to include into the design as variables: (1) Type of collateral damage (hospital, school, temple, water, psychological damage); (2) Activity of target (hostile or passive); (3) Threat level; (4) Chance of success; (5) Type of ammunition; (6) Type of decision by machine (automatically select target and acts or chooses best action based on human selection of target); (7) Duration of mission; (8) Means of control (automatic or human deactivation).

3.3 Scenarios

In this section, the variables for the Moral Machine for Autonomous Weapons, which are based on the online survey, are defined and depicted in two example scenarios. The variables are based on the scenario in which a military convoy is supported by a drone in the air. The 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 drone detects a vehicle behind a mountain range that is approaching the convoy at high speed. The 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 drone needs to decide if it attacks the approaching vehicle which could result in the death of all four passengers and might cause collateral damage by killing people that are nearby the road.

3.4 Variables for Scenarios

For the prototype of the Moral Machine for Autonomous Weapons six variables are defined: *Type of Weapon* (W), *Location* (L), *Character* (C), *Number of Characters* (N), *Outcome* (O) and *Mission* (M). Similar to the Moral Machine of Autonomous Vehicles [26], the variables are depicted in figures and pictograms. These are suggestions and will need to be tested to verify how people understand these variables and if their understanding concurs with the intended meaning by the researchers. The variables also will need to be professionally designed when implemented in a Moral Machine for Autonomous Weapons. For now, the *Location* variable is based on a cartoon and game like depiction of a desert and village and the collateral damage of the *Outcome* variable is depicted with a smiley. These depictions have been chosen to avoid resemblance with real-life locations, but the cartoon-like pictures and smiley's need to be tested if these are a clear representation or if a photograph is more suitable when used in a study.

Location. This dimension shows the *Location* (L) as setting for the scenario which can either be in the desert or in the village (Figs. 2 and 3). In this dimension, we test if which location is more morally acceptable for people to deploy the AWS. If *Location* is the discriminative variable, then both different images are shown on each of the scenario. Otherwise the same location is depicted in both scenarios as in the example (Fig. 9).

Fig. 2. Location variable L = {desert}

Fig. 3. Location variable L = {village}

Type of Weapon. This dimension shows the *Type of Weapon* (W) that is deployed to support the convoy. This dimension tests if people judge the current technology, the Human Operated drone, to be more morally acceptable than the AWS, which is future technology. This can be used to get insight in the support for the technologies. The *Type of Weapon* is either a Human Operated drone (left hand side) or an AWS (right hand side) (Fig. 4). If *Type of Weapon* is the discriminative variable, then the different pictograms are shown on the scenario for people to choose between them as is shown in the example (Fig. 9).

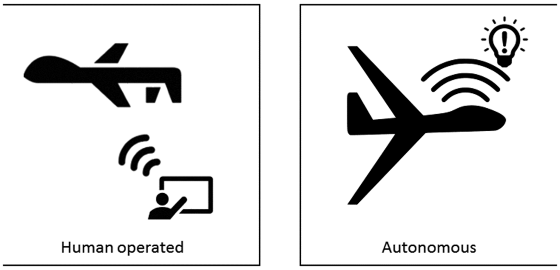

Fig. 4. Type of weapon variable

Character. This dimension shows the type of *Character* (C) that is involved as bystanders in the scenario which can either be a man (on the left), a woman (in the middle) or a child (on the right) (Fig. 5). By varying the characters, we gain insight which bystanders people find morally acceptable when an Autonomous or Human Operated Weapon is used. If *Character* is the discriminative variable, then different characters are shown on each of the scenario. Otherwise the same characters are depicted in both scenarios.

Fig. 5. Character variable

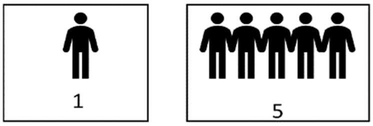

Fig. 6. Number of characters

Number of Characters. This dimension shows the *Number of Characters* (N) that are involved in the scenario which can range from one character (on the left) up to five characters (on the right) (Fig. 6). This dimension allows us to gain insight into how many bystanders people find morally acceptable when an Autonomous or Human Operated Weapon is used. If the *Number of Characters* is the discriminative variable, then different numbers of characters are shown on each of the scenario. Otherwise the same number of characters are depicted in both scenarios.

Outcome. This dimension shows the *Outcome* (O) of the scenario which can either be with no collateral damage (on the left) or an outcome with collateral damage of the number of characters involved in the scenario (on the right) (Fig. 7). This allows us to gain insight into how the outcome influences the moral acceptability of when an AWS is deployed. If *Outcome* is the discriminative variable, then different pictograms are shown on each of the scenario. Otherwise the same outcome variable is depicted in both scenarios.

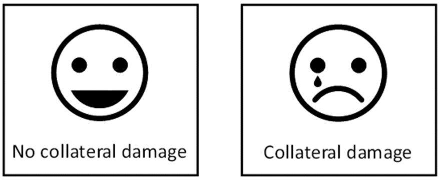

Fig. 7. Outcome variable

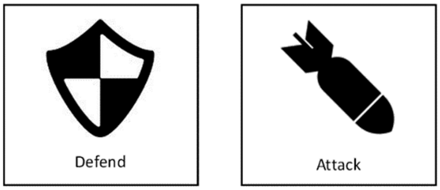

Fig. 8. Mission variable

Mission. This dimension shows the *Mission* (M) of the weapon in the scenario which can either be to defend the convoy only when a direct threat is perceived (on the left) or to attack vehicles that are on the target list even as they do not pose a direct threat for the convoy (on the right) (Fig. 8). In this dimension, we test which type of mission is morally acceptable to people. If the *Mission* is the discriminative variable, then different pictograms are shown on each of the scenario. Otherwise the same mission variable is depicted in both scenarios.

3.5 Example Scenario

The variables described above can be used to create scenarios in which each scenario differs on only one variable. The question presented to the user in the scenario is the same question as that is being asked when judging the scenarios in the original Moral Machine. In this section, we depict a scenario as an example to show the concept, but for the sake of brevity chose not to show all variables in an endless list of examples. The example in Fig. 9 shows a convoy in a desert location. The difference between the scenarios is that on the left the convoy is defended by a Human Operated drone and in

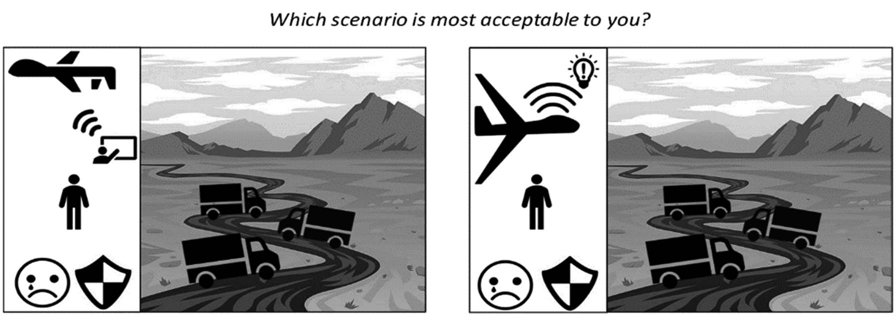

Fig. 9. Example scenario

the scenario on the right by an AWS. In both cases the mission of the drone is to defend the convoy and there is one person near the road which is killed by collateral damage.

Implementation. To test the scenarios, we will design the website that allows people to take the survey after obtaining a password via a web-interface to a secure server. We will gather initial data and user feedback in pilot studies to check the research set-up and clarity of the survey. The next step will be to scale up to a large-scale open platform, like the Moral Machine, where people can judge the scenarios to collect a large amount of data in several countries. However, due to the sensitivity of the topic we believe it would not be advisable to allow people to create their own scenarios or share their results on social media as the *Design* feature of the original Moral Machine for Autonomous Vehicles offers. This will be a process that will take several iterations until the final study can be tested. The original Moral Machine collected data from June 23, 2016 until May 2017 [26] and we propose to run the study on AWS for the same duration to get a large enough sample to truly call it a Massive Online Experiment.

4 Discussion

This section discusses the limitations of this design proposal and the recommendations for further research. Given the preliminary nature of the work this proposal has several limitations. The first is that the work is conceptual. Not only do we need to test the scenarios and layout of the variables to verify if there are easily understood, but also we need to check whether the understanding of the concepts concur with the intended meaning of the researchers before they can be implemented. Secondly, at this stage it is not clear if the scenarios will lead to usable and statistically robust research data on the moral judgement of people on AWS and the scenarios might have to be tweaked in order to obtain this data. Thirdly, the study uses drones in the scenarios, but this is only one type of AWS. Many other weapons can be surveyed, for example the anti-aircraft missiles or autonomous torpedoes of submarines. The final limitation is that the scenarios present a simplified view of military operations and much more variables than the ones tested in the scenarios, such as the changes in the circumstances and overall threat level due to unexpected events or strategy of the opponent, are at play. It is exactly because of the complexity of the theme, that is of the utmost importance to proceed with great caution and to carefully design and measure each choice. This experiment is a first step towards this aim.

5 Conclusions

In the discussion of how the development of AWS should be controlled, there is a need to give full consideration not only to national and commercial interests but most importantly to have a profound understanding of the feelings people associate with this type of weapon. This paper proposes a MOE in order to get a grounded view on the perception of the moral acceptability of AWS of the wider public on this question.

The limitations discussed above call for recommendations for further research. The first is that the variables should be validated with a panel consisting of different demographical composition and cultures to verify the understandability and clarity. Secondly, to validate that the scenarios generate accurate and usable research data they should be tested in several pilot studies to verify the usability of the proposed research set-up of a randomized controlled experiment before launching the website. Thirdly, after the initial scenarios are tested a follow-up study is recommended to expand the *Type of Weapon* variable in which other types of AWS are tested. For example, anti-aircraft missiles or autonomous torpedoes. Other expansions of the variables would include a different *Location*, for example villages in the West compared to villages in the Middle-East or Asia. It is also possible to include different types of *Characters*, such as animals or people with different social values, e.g. criminals, doctors and pregnant women. Lastly, the research could be extended to incorporate more scenarios that include realistic military variables in order for the MOE to be representative for military operations. Included could be for example different threat levels, influences from the decision-making process or circumstances that change rapidly during the execution of the mission in order to get a thorough understanding of the moral judgement of people regarding Autonomous Weapon Systems.

References

1. IEEE Global Initiative, The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (2017)
2. Campaign to Stop Killer Robots. Campaign to Stop Killer Robots (2015). https://www.stopkillerrobots.org/. Accessed 15 July 2017
3. ICRC, Ethics and autonomous weapon systems: An ethical basis for human control? International Committee of the Red Cross (ICRC), Geneva, p. 22 (2018)
4. Dignum, V.: Responsible Autonomy (2017). arXiv preprint: arXiv:1706.02513
5. Dignum, V.: Introduction to AI (2016). https://rai2016.tbm.tudelft.nl/contents/
6. Roff, H.M.: Weapons autonomy is rocketing (2016). http://foreignpolicy.com/2016/09/28/weapons-autonomy-is-rocketing/
7. US Air Force, Unconscious US F-16 pilot saved by Auto-pilot 2016, Catch News: Youtube
8. Etzioni, A., Etzioni, O.: Pros and Cons of Autonomous Weapons Systems. Military Review, May–June 2017
9. Arkin, R.C.: The case for ethical autonomy in unmanned systems. J. Mil. Ethics **9**(4), 332–341 (2010)
10. General Assembly United Nations, 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, p. 23 (2016)
11. Kaag, J., Kaufman, W.: Military frameworks: technological know-how and the legitimization of warfare. Camb. Rev. Int. Aff. **22**(4), 585–606 (2009)
12. Rosenberg, M., Markoff, J.: The Pentagon's 'Terminator Conundrum': Robots That Could Kill on Their Own, in The New York Times (2016)
13. Malle, B.F.: Integrating robot ethics and machine morality: the study and design of moral competence in robots. Ethics Inf. Technol. **18**, 243–256 (2015)

14. Cointe, N., Bonnet, G., Boissier, O.: Ethical judgment of agents' behaviors in multi-agent systems. In: Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems. International Foundation for Autonomous Agents and Multiagent Systems (2016)
15. Bonnefon, J.-F., Shariff, A., Rahwan, I.: The social dilemma of autonomous vehicles. Science **352**(6293), 1573–1576 (2016)
16. Coeckelbergh, M.: Drones, information technology, and distance: mapping the moral epistemology of remote fighting. Ethics Inf. Technol. **15**(2), 87–98 (2013)
17. Strawser, B.J.: Moral predators: the duty to employ uninhabited aerial vehicles. In: Valavanis, K.P., Vachtsevanos, G.J. (eds.) Handbook of Unmanned Aerial Vehicles, pp. 2943–2964. Springer, Dordrecht (2010). https://doi.org/10.1007/978-90-481-9707-1_99
18. Scalable Cooperation Group. Moral Machine (2016). http://moralmachine.mit.edu/. Accessed 27 Sept 2016
19. Castelfranchi, C., Falcone, R.: From automaticity to autonomy: the frontier of artificial agents. In: Hexmoor, H., Castelfranchi, C., Falcone, R. (eds.) Agent Autonomy, pp. 103–136. Springer, Boston (2010). https://doi.org/10.1007/978-1-4419-9198-0_6
20. Article 36. Killing by machine: Key issues for understanding meaningful human control (2015). http://www.article36.org/autonomous-weapons/killing-by-machine-key-issues-for-understanding-meaningful-human-control/. Accessed 9 May 2019
21. Open Roboethics initiative. The Ethics and Governance of Lethal Autonomous Weapons Systems: An International Public Opinion Poll, 5 Nov 2015. http://www.openroboethics.org/laws_survey_released/. Accessed 15 July 2017
22. Jackson, C.: Three in Ten Americans Support Using Autonomous Weapons (2017). https://www.ipsos.com/en-us/news-polls/three-ten-americans-support-using-autonomous-weapons. Accessed 17 June 2018
23. AIV and CAVV, Autonomous weapon systems: the need for meaningful human control, A.C.O.I.O.P.I.L. Advisory Council on International Affairs, Editor, pp. 1–64 (2016)
24. Kuptel, A., Williams, A.: Policy guidance: autonomy in defence systems (2014)
25. UNDIR, Framing Discussions on the Weaponization of Increasingly Autonomous Technologies, pp. 1–14 (2014)
26. Awad, E.: Moral Machine: Perception of Moral Judgment Made by Machines. Massachusetts Institute of Technology, Boston (2017)
27. Reips, U.-D.: Standards for internet-based experimenting. Exp. Psychol. **49**(4), 243 (2002)
28. Oehlert, G.W.: A First Course in Design and Analysis of Experiments. W.H. Freeman, New York (2010)