Autonomous Weapon Systems, Public Opinion, and the Moral Equality of Combatants

Daniel Lim

daniel.lim672@dukekunshan.edu.cn
Arts & Humanities Division
Duke Kunshan University
Kunshan, China

Philip Santoso

lie.santoso@dukekunshan.edu.cn
Social Sciences Division
Duke Kunshan University
Kunshan, China

Abstract—A novel deontological objection to Autonomous Weapon Systems (AWS) based on the moral equality of combatants has recently been developed. While this is an interesting objection, there is reason to believe that it: (1) fails to maintain a moral distinction between AWS and long-distance human-guided weaponry and (2) fails to show that AWS are truly independent in terms of targeting and engaging enemies. Moreover, based on a random sampling of U.S. citizens, public opinion in the U.S. suggests that popular assessment of AWS is more sensitive to relative effectiveness rather than deontological reason. Consequently, this deontological objection to AWS fails and deontological objections to AWS, more generally, can be overridden by consequentialist considerations.

Keywords-autonomous weapon systems; moral equality of combatants; deontology.

I. INTRODUCTION

In current debates over the development and use of Autonomous Weapon Systems (AWS), deontological arguments have been brought to bear against AWS [1][2][3][4]. A deontological argument against AWS can be characterized as an argument against AWS even if AWS were able to provide the best possible outcomes in terms of legitimate military aims (e.g., minimizing casualties, both military and civilian). In this paper, we aim to show that a specific kind of deontological argument fails and that deontological arguments, more generally, can be overridden by consequentialist considerations.

This paper is organized as follows. In Section II we will briefly summarize several prominent deontological arguments that have been raised against AWS. In Section III we will examine a recent, novel deontological argument based on the moral equality of combatants developed by Skerker et al. [5] that purports to fill gaps in existing arguments. In Section IV we will argue that Skerker, et al.'s argument fails because the concept of 'independent targeting' cannot be fleshed out in a way that (i) draws a clear distinction between AWS and conventional weapons and (ii) can be practically implemented. In Section V we will share results from a study that suggests deontological arguments based on the distinction between independent and dependent targeting can be overridden by considerations of the effectiveness of the weapons involved. We will end with concluding remarks in Section VI.

II. DEONTOLOGICAL ARGUMENTS

Here, we will briefly detail some of the prominent deontological arguments that have been raised against AWS. Some have argued that AWS create responsibility gaps [1]. If AWS were to violate the international laws that govern war, there would be no one to hold morally responsible for such crimes so it would be wrong to deploy AWS. Others have argued that AWS disrespect human combatants because they fail to establish 'interpersonal relationships' with the enemy or fail to acknowledge the humanity of the enemy [2][3]. Consequently, it would be wrong to deploy AWS. Finally, some have argued that combatants must act for the proper reasons in order to justly engage in war [4], but because AWS cannot act for reasons *simpliciter* (let alone the proper reasons) it would be wrong to deploy AWS.

These deontological arguments all have shortcomings. The argument from responsibility gaps would fail if AWS could flawlessly replace human combatants. If AWS could ensure adherence to the international laws that govern war, then the need to locate morally responsible parties for breaches of international laws seems to disappear. The argument from disrespect based on the failure to establish interpersonal relationships fails because it is unable to distinguish AWS from commonly used long-distance, human-guided weaponry (e.g., drones). Consequently, if this argument succeeds against AWS it would also succeed against many weapons that are currently in use. The argument from disrespect based on a failure to acknowledge the humanity of the enemy fails because AWS could, in a sense, be more respectful to the enemy than human combatants. If AWS were to become superior to human combatants in terms of discriminating between legitimate and illegitimate targets and providing assurance that illegitimate targets are not unjustly treated, then AWS would be more respectful. Finally, the argument from acting based on proper reasons fails because human combatants do not always act based on proper reasons. Consequently, this would invalidate much that currently goes on in war. Moreover, if most human combatants were to act from improper reasons, then we would have a legitimate reason to replace human combatants with AWS.

III. THE MORAL EQUALITY OF COMBATANTS

In a recent paper, Skerker et al. have developed a novel deontological argument against AWS that they believe will overcome the shortcomings of the alleged deontological reasons introduced above [5]. They argue that all military personnel at war enter into a martial contract with enemy combatants. Through this contract military personnel:

"… cede a right not to be directly targeted with lethal violence to… agents able to reason about moral considerations (including whether to exercise their rights at others' expense). Such agents would also need to have an understanding of the authority of moral reasons, reasons grounded in the value of human life." [5]

In other words, military personnel cede a right not to be directly targeted with lethal violence to agents who are *moral equivalents*. Moral equivalents of human combatants have, among other things, "the capacity for moral responsibility" and "moral maturity."

Since AWS do not have the capacity for moral responsibility and are not morally mature, they cannot enter into the martial contract that all military personnel are required to enter. So, human combatants are unable to cede a right not to be harmed to AWS. It would therefore be wrong to deploy AWS against human targets because those targets are unable to cede the right not to be harmed to AWS. Let us call this the Moral Equality of Combatants (MEC) argument against AWS.

One might be tempted to think that this objection to AWS has an obvious flaw: it is too strong since the same objection can easily be raised against conventional weapons. When a human combatant uses a gun to shoot a human enemy, one might argue that it was the bullet that harmed the enemy. However, because the enemy could not cede the right not to be harmed to the bullet (because bullets are not moral equivalents), it would be wrong to use guns against human targets.

This response, however, fails to take seriously the critical distinction between AWS and conventional weapons. Conventional weapons do not *target* human enemies on their own. It is the human combatant who targets and ultimately harms the enemy. While it might be natural to say, for example, that "the bullet killed him," technically speaking, the bullet is nothing more than the means by which the human combatant kills the enemy. So the enemy, by entering the martial contract with opposing human combatants, is not violated because the right not to be harmed was *properly ceded* to the opposing human combatant, not the bullet.

Contrary to conventional weapons, AWS target enemies *on their own*. Because of this, it is not possible for human combatants "to count as targeting combatants targeted by the AWS they deploy." [5]. As a result, the enemy cannot cede the right not to be harmed to the human combatant who deploys the AWS. The enemy harmed by an AWS is disrespected because such an enemy never successfully ceded the right not to be harmed to anyone in this situation.

Not only does the MEC argument enjoy prima facie plausibility, it is strengthened by the fact that it addresses the shortcomings of the deontological arguments introduced above. The MEC argument, unlike the argument from responsibility gaps, is not contingent on AWS that fail to flawlessly replace human combatants. Even if AWS could flawlessly replace human combatants, deploying AWS would still be wrong because humans targeted by AWS would not be able to cede the right not to be harmed.

The MEC argument, unlike the arguments from disrespect, does not depend on the inability of AWS to establish interpersonal relationships nor does it depend on the inability of AWS to respect the humanity of human enemies. Moreover, the MEC argument, unlike the argument from acting based on proper reasons is not contingent on the inability of human combatants to wage war based on proper reasons. The MEC argument requires only that AWS are not moral equivalents for the argument to succeed in showing that deploying AWS is morally wrong.

IV. INDEPENDENT TARGETING

The success of the MEC argument critically depends on the distinction between existing weapons and AWS. This distinction, according to Skerker, Purves, and Jenkins, lies squarely on the claim that AWS target their enemies 'on their own'. We might call this the ability to *independently* target enemies. No other weapons have this ability.

It is unclear, however, whether the concept of independent targeting can be clarified with sufficient detail in a way that vindicates the MEC argument. If the concept cannot be applied to AWS or if the concept can be applied to existing weapons, the MEC argument would be undermined. Let us begin with an extremely strong definition of independent targeting.

INDEPENDENCE$_1$: An entity S independently targets an enemy only if S does not receive any initial guidance or constraints regarding who or what to target.

According to this definition, it is clear that AWS fail to independently$_1$ target enemies because they are programmed (and possibly trained on datasets) in ways that provide guidance and constraints regarding who or what they target. Since AWS fail to independently$_1$ target it can be argued that it is the human combatant who programmed the AWS that did the targeting. In this sense, AWS are merely tools and moral responsibility ultimately falls on its human programmer. So deploying AWS still makes it possible for AWS victims to cede their right not to be harmed.

INDEPENDENCE$_2$: An entity S independently targets an enemy only if S makes the *final* targeting decision.

According to this definition, it is clear that AWS independently$_2$ target enemies. Since independence$_2$ does not require the total absence of guidance or constraints, once AWS are deployed, like human combatants, they are continually making decisions and negotiating who or what to target up until harm is inflicted. So, we can say that AWS,

though given initial guidance and constraints, make the final targeting decisions.

It is equally clear that bullets do not independently$_2$ target enemies. All the targeting is done by human combatants wielding guns. Once the gun is fired and the bullet is propelled, the bullet makes no decisions about who or what to target. So there is a principled distinction between human combatants and AWS on the one hand and conventional weapons like guns on the other.

The problem for this definition, however, is that it fails to distinguish AWS from existing weapons like Raytheon's AIM-120 AMRAAM (advanced medium range air-to-air missile). The AIM-120 AMRAAM is considered a 'fire-and-forget' weapon because, though its target is pre-selected, once it reaches a certain proximity to its target, its on board radar-guidance system allows it to autonomously track the target. It continues to negotiate the location of the target up to the point that harm is inflicted. Because it makes the final targeting decision, we should also say that the AIM-120 AMRAAM independently$_2$ targets enemies. The deployment of this weapon along with many other existing 'fire-and-forget' weapons (developed in at least a dozen nations) would fall prey to the MEC argument.

INDEPENDENCE$_3$: An entity S independently targets an enemy only if S makes the *final* targeting decision and it is false that S receives initial guidance or constraints that targets a *single* individual.

Perhaps the difference between existing 'fire-and-forget' weapons is that these weapons, unlike AWS, are always targeted at specific, individually identifiable targets. The AIM-120 AMRAAM, though it makes the final targeting decision, is not independent$_3$ because this weapon receives initial guidance or constraints that targets a single individual. Though this may show that the AIM-120 AMRAAM does not independently$_3$ target enemies, it is unclear that the same can be said for the AGM-114 Hellfire Longbow Variant (or simply Hellfire). This is also a 'fire-and-forget' weapon but, unlike the AIM-120 AMRAAM, the Hellfire can lock onto its target *after* being launched. It seems that it does *not* receive initial guidance or constraints that targets a single individual. It targets *any* member of a set of individuals that match the specifications programmed into its radar guidance system. So it is arguable that the Hellfire, an existing weapon, may be classified as a weapon that independently$_3$ targets enemies and therefore also falls prey to the MEC argument.

Perhaps, there is a deeper issue lurking. Though the distinction between targeting a single individual and targeting a set of individuals may seem theoretically clear, the distinction may prove untenable in practice. To see this we might consider the task of translating *singular* propositions (i.e., propositions that make claims about specific individuals) into *categorical* propositions (i.e., propositions that relate two categories). Categorical propositions in standard form come in four possible formats: (1) All S are P, (2) No S are P, (3) Some S are P, and (4) Some S are not P – where S and P are placeholders for categories. Here is a categorical proposition in the first format:

All weapons are dangerous things.

Two categories are being related: weapons (S) and dangerous things (P). This proposition is asserting that everything in the category of weapons belongs in the category of dangerous things. Without getting into details, the value of translating propositions into categorical propositions in standard form is that arguments based on such propositions can easily be assessed through formal logic.

Consider singular propositions like the following:

Philip is British.

Is it possible to translate this into a categorical proposition? One way is to treat Philip as a criterion for membership into a category.

All people identical to Philip are people who are British.

What this might show is that the distinction between singular propositions and categorical propositions is not rigid. Singular propositions can be treated as categorical propositions.

How does the non-rigid distinction between singular and categorical propositions relate to the distinction between targeting a single individual and targeting a set of individuals? Perhaps we want to use a weapon, like the AIM-120 AMRAAM, to kill a specific individual X. It may be tempting to say that this weapon fails to independently$_3$ target its enemy because it received guidance that targets a single individual. As just shown, we can easily construe the targeting of a single individual in terms of a category: all targets identical to X. Put in this way, we might argue that it is false that the AIM-120 AMRAAM receives initial guidance or constraints that targets a single individual. Rather it was targeted at a set of individuals.

This would seem to be a pyrrhic victory since the meaning behind the locution 'all targets identical to X' is, for all intents and purposes, a way of referring to a single individual. It may, however, only be pyrrhic in a purely theoretical sense. Consider how one might operationalize 'being identical to X'. How could this provide concrete guidance in identifying X? There seems to be something of a tautology at work here. If I were to tell you to target all people identical to, say, Max Eisenhardt (i.e., Magneto), then, assuming you did not know who Max Eisenhardt is, to tell you to target all and only those individuals identical to Max Eisenhardt would do you no good.

Presumably, one would have to rely on a variety of characteristics that are putatively unique to Max Eisenhardt - characteristics that would be readily available for a human combatant engaged in war to asses. What might such characteristics be? Most likely they would be physical attributes (e.g., height, skin color, facial features). Then what would be meant by 'identical to Max Eisenhardt' would essentially be a list of descriptions. Perhaps something like the following:

All people 190 cm tall with light brown skin color and facial features $f_1, f_2, \ldots$ are targets to be engaged.

Practically speaking, given the number of combatants in a given military, it will be possible that more than one individual will satisfy these descriptions (with some threshold accuracy).

While there might be a neat theoretical distinction between targeting a single individual and a set of individuals, it can reasonably be argued that the practical distinction between targeting a single individual and a set of individuals is rather blurry. As such one might reasonably argue that AWS, when given actual guidance and constraints on who or what to target, do not engage in independent$_3$ targeting.

V. PUBLIC OPINION

The debate over the ethics of AWS will not, however, be decided solely on theoretical grounds. A critical factor in how this debate evolves is public opinion. And this is not merely a descriptive fact about the nature of societies. Apparently, public opinion is encoded into the very laws that govern international warfare. The Martens Clause states:

> "Until a more complete code of the laws of war is issued, the High Contracting Parties think it right to declare that in cases not included in the Regulations adopted by them, populations and belligerents remain under the protection and empire of the principles of international law, as they result from the usages established between civilized nations, from the laws of humanity and the requirements of the public conscience." [6]

Here, we highlight the term 'public conscience' which emphasizes the importance of aligning international law with public opinion.

To probe 'public conscience' we ran a survey of the U.S. public to see, among other things, what their views are regarding AWS [7]. In particular, we wanted to see whether their support for or against AWS would be affected by the distinction between weapon systems that independently target enemies and weapon systems that rely on human controllers to determine who or what to target. Moreover, we wanted to see whether their support for or against AWS would be affected by their 'effectiveness'. That is, whether the weapon systems would be better than human combatants using conventional weapons in terms of reducing the number of overall casualties.

1,600 respondents took our survey. 52% were female and age was evenly distributed. Half the respondents were given a scenario describing weapon systems that rely on human controllers to determine who or what to target (dependent condition) and the other half were given a scenario describing weapon systems that independently target enemies (independent condition). Moreover, with each of these conditions we varied the way the weapon systems were described in terms of relative effectiveness. The weapon systems either performed worse, the same as, or better than human soldiers using conventional weapons. Moreover, we

added a control condition that did not comment on their effectiveness. Respondents were then asked whether they "agree that weapons like these should be developed" based on a 5-point Likert Scale ('strongly disagree', 'somewhat disagree', 'neither disagree nor agree', 'somewhat agree', and 'strongly agree').
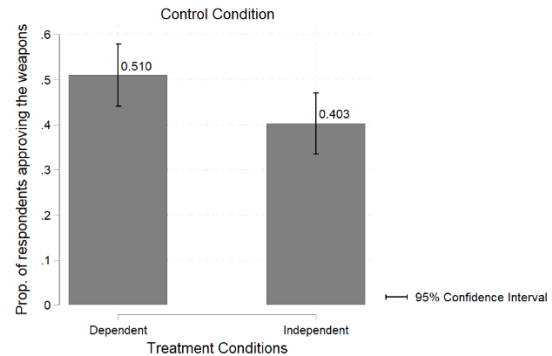

Figure 1.   51% of respondents in the dependent condition and 40% in the independent condition supported the development of the relevant weapons.

We see in the control condition, as shown in Figure 1, weapons that independently target enemies enjoy less approval than their dependent counterparts (where approval stands for a 'somewhat agree' or 'strongly agree' response). Not only is the public less familiar with AWS with the capacity for independent targeting, there is legitimate concern over whether or not such weapons can perform with the same level of effectiveness as human soldiers with conventional weapons.
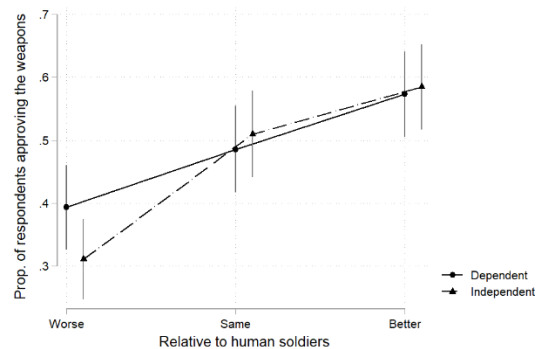

Figure 2.   The proportion of respondents steadily increased as the weapons improved in their effectiveness across both dependent and indepent conditions.

That, being said, we also found that the effectiveness of the relevant weapon system affected approval, as shown in Figure 2. It is interesting to note that though weapons which independently target enemies start with a lower proportion of approval (in the worse condition), they eventually end up slightly eclipsing the weapons requiring human controllers for targeting (in the better condition). What this suggests is that whatever deontological worries the public might harbor regarding AWS, these worries are not static. Were AWS able

to provide the best possible outcomes in terms of legitimate military aims (e.g., minimizing casualties, both military and civilian), it seems the public conscience would not be so deeply affected with the deployment of AWS.

VI. CONCLUSION AND FUTURE WORK

Though there may be other kinds of deontological arguments raised against AWS in the future, the widely discussed arguments discussed today hinge on the distinction between independent and dependent targeting. This distinction, while theoretically interesting, either cannot be appropriately fleshed out or cannot serve a useful ethical purpose. Moreover, our survey results suggest that whatever deontological reasons people might have for rejecting AWS, the weight of these reasons can be overridden by consequentialist considerations.

To make progress on our results we would like to expand our study to probe military personnel. It may be that experience in military contexts (especially live combat) may affect the way people understand the ethics of AWS. Moreover, it would be interesting to see if the trends discovered with a U.S. population will remain consistent in other cultures. It would be interesting to extend this work to probe the 'public conscience' in other key nations developing AWS like China and Russia.

REFERENCES

[1] R. Sparrow, "Killer Robots," Journal of Applied Philosophy, vol. 24, pp. 62-77, 2007.

[2] R. Sparrow, "Robots and Respect: Assessing the Case against Autonomous Weapons Systems," Ethics and International Affairs, vol. 30, pp. 93-116, 2016.

[3] T. Nagel, "War and Massacre," Philosophy and Public Affairs, vol. 1, pp. 123-144, 1972.

[4] D. Purves, R. Jenkins, and B. J. Strawser, "Autonomous Machines, Moral Judgment, and Acting for the Right Reasons," Ethical Theory and Moral Practice, vol. 18, pp. 851-872, 2015.

[5] M. Skerker, D. Purves, and R. Jenkins, "Autonomous Weapons Systems and the Moral Equality of Combatants," Ethics and Information Technology, vol. 22, pp. 197-209. 2020.

[6] International Committee of the Red Cross. "The Martens Clause and the Laws of Armed Conflict" [Online]. Available from:
https://www.icrc.org/en/doc/resources/documents/article/other/57jnhy.htm 2021.06.26

[7] M. Horowitz, "Public Opinion and the Politics of the Killer Robots Debate," Research and Politics, vol. 3, pp. 1-8. 2016.