Philosophy & Technology (2023) 36:41
https://doi.org/10.1007/s13347-023-00647-2

RESEARCH ARTICLE

A Moral Bind? — Autonomous Weapons, Moral Responsibility, and Institutional Reality

Bartlomiej Chomanski[1]

Received: 15 December 2022 / Accepted: 31 May 2023 / Published online: 9 June 2023

Abstract

In "Accepting Moral Responsibility for the Actions of Autonomous Weapons Systems—a Moral Gambit" (2022), Mariarosaria Taddeo and Alexander Blanchard answer one of the most vexing issues in current ethics of technology: how to close the so-called "responsibility gap"? Their solution is to require that autonomous weapons systems (AWSs) may only be used if there is some human being who accepts the ex ante responsibility for those actions of the AWS that could not have been predicted or intended (in such cases, the human being takes what the authors call the "moral gambit"). The authors then propose several institutional safeguards to implement in order to ensure that the moral gambit is taken in a fair and just way. This paper explores this suggestion in the context of the institutional settings within which AWSs are most likely to be deployed. It raises some concerns as to the feasibility of Taddeo and Blanchard's proposal, in light of the recent empirical work on the incentive structures likely to exist within militaries. It then presents a potential problem that may arise in case the accountability mechanisms are successfully implemented.

Keywords Autonomous weapons systems · Responsibility gap · Moral responsibility · Accountability

1 Introduction

In "Accepting Moral Responsibility for the Actions of Autonomous Weapons Systems—a Moral Gambit" (2022), Mariarosaria Taddeo and Alexander Blanchard answer one of the most vexing issues in current ethics of technology: how to close the so-called "responsibility gap"? The responsibility gap arises when assigning responsibility for the "actions" of autonomous artificial agents — mostly, but not exclusively, discussed in the context of autonomous weapons systems (AWSs). As

✉ Bartlomiej Chomanski
b.chomanski@gmail.com

[1] Department of Philosophy, Adam Mickiewicz University, Poznan, Poland

Taddeo and Blanchard document, it is almost[1] universally acknowledged that it makes little sense to hold machines themselves responsible; rather, responsibility for what they do must always lie with a human being. However, as it turns out, determining *which* human being(s) to assign responsibility to is far from easy (for a classic articulation of this problem in the context of AWS, see Sparrow, 2007; see also Matthias, 2004; Nyholm, 2018; Gordon & Nyholm, 2021, Sect. 2b). The difficulty arises because the machines' autonomous actions may well not be intended or predicted by anyone.

Taddeo and Blanchard argue, further, that moral responsibility so assigned must be *meaningful*, in the sense that it "bears in a justified and fair way on those who have played a key role in the realisation of the effects of using AWS" (2022, 4). It must also attach to human beings *qua* individuals, not as representatives of some organization.

Responsibility assignments are important, as Taddeo and Blanchard explain, because meaningful moral responsibility "enables backward-looking responsibility, as it *fosters accountability*. It may also enable forward-looking responsibility, insofar as the prospect of the blame and praise linked to a given decision/action would *facilitate morally sound choices and careful conduct* [emphasis added]" (2022, 15). People morally responsible can be held accountable; and they will be motivated by potential praise and blame to act in a just and fair manner.

2 The Moral Gambit

So, how do Taddeo and Blanchard aim to close the responsibility gap?

Following a similar approach advocated by Marc Champagne and Ryan Tonkens (2015), Taddeo and Blanchard propose that responsibility for the unpredictable actions taken on the battlefield by AWS be assigned to those who willingly and knowingly accept it ex ante — through what Taddeo and Blanchard call a "moral gambit." This, in their view, is the only way to fairly resolve the responsibility question.

As Taddeo and Blanchard put it,

> All one may ask is for designers, developers and deployers to take meaningful moral responsibility for the intended actions, while being aware of the risk that unpredicted outcomes may occur and *accepting* moral responsibility also for the unpredictable effects that may follow the decision to deploy AWS. Let us specify this aspect. In accepting this responsibility, the human agents make a moral gambit: they design/develop/use an AWS, being fully aware of the risks that it may perform some unpredicted actions… The human agents remain aware that independently of all these efforts, it will not be feasible to predict all possible actions of an AWS and their effects on the context of deploy-

[1] For a view that seeks to make sense of the idea of assigning responsibility to machines themselves, albeit in a legal, rather than moral, context, see Chomanski (2021).

> ment. Nonetheless, if they decide to proceed with the design/development/use of these systems, then they make a moral gambit and accept to be morally responsible for the unforeseen AWS outcomes and their effects [emphasis in original] (2022, 16).

Taddeo and Blanchard develop a number of suggestions for the creation of institutions to facilitate making the moral gambit in as full knowledge and understanding of the stakes involved and other relevant information as possible.

The institutional design proposed by Taddeo and Blanchard would try to ensure that AWSs are interpretable and that their predictability can be reasonably assessed; that the decision-makers have as much knowledge and understanding of these systems as possible; and that there are accountability mechanisms: specifically, "The decision as to use or not use non-lethal AWS should always follow a risk/benefit analysis and be justified according to the principle of necessity" (2022, 19). Moreover, "A process to identify mistakes and unwanted outcomes, to assess their impact and costs and to define redressing remedy measures should be established" (2022, 20). Finally,

> Ethics-based auditing of both the non-lethal AWS and of the processes for their acquisition and deployment should be established … with the aim of facilitating accountability as well as to identify possible points of failure and address them promptly, so to improve the decision-making and the redressing processes. (ibid.)

Overall, for the moral gambit to close the responsibility gap in a meaningful way, the basis for what AWSs do must be capable of being understood; the decision-makers must actually possess the knowledge to understand it; they must also be incentivized to only deploy AWS in justified circumstances; and, in case of mistakes, redress, remedies, and audits must be available.

In addition to the moral gambit proposal, there is much more to recommend in Taddeo and Blanchard's paper, including a rich discussion of the unpredictability of autonomous artificial agents, an engaging appraisal of alternative methods for assigning responsibility for what AWSs do, and an inventive argument why the moral gambit should be limited only to non-lethal uses of AWS (the conclusion I accept for the sake of argument). Nevertheless, in this piece, I want to focus on some pointers for further discussion and research that flow from Taddeo and Blanchard's institutional proposals. (I also think, though I do not develop this suggestion much further in this article, that Champagne & Tonkens' approach raises similar reservations.)

3 The Moral Bind?

In this section, I make the following argument: As far as institutions are concerned, one may have a reasonable worry, based on past performance and present incentive structures, that Taddeo and Blanchard's proposal for implementing the moral gambit is too idealistic, and hence unlikely to achieve meaningful responsibility assignments in the real world. Rather, for all the institutions they propose, responsibility may remain hopelessly "nominal" more often than not.

Assume the non-lethal AWS, deployed in accordance with all the rules Taddeo and Blanchard envision, turns out to have unpredictably "cause[d] significant damage, including bodily harm, disproportionate destruction to property, infringements of liberty and breaches of the principle of distinction [hereafter simply significant damage]" (2022, 18). Jay, a high-ranking official within the military, took — as a "deployer" of the AWS — a moral gambit on using the AWS. He lost. Now, in accordance with Taddeo and Blanchard's principles, Jay is morally responsible for these unpredictable harms. What should happen to Jay?

Taddeo and Blanchard do not say much about situations like these, as far as legal remedies are concerned. They prefer to focus on *moral* responsibility alone, leaving questions of legal responsibility for a further discussion.[2] Still, it seems to me that for moral responsibility to be meaningful, some accountability (legal or otherwise) should be a requirement (otherwise, both the backward-looking and forward-looking aspects of responsibility that Taddeo and Blanchard mention appear absent). One plausible principle regarding Jay in these circumstances could thus be as follows:

> PRESUMPTIVE MORAL-LEGAL RESPONSIBILITY LINK: if Jay is morally responsible for significant damage caused by the AWS, then, *pro tanto*, Jay should be held legally responsible for that damage.

Such a principle would suggest that, where agents are morally responsible for some significant harms, then, absent excellent reasons to the contrary, they should also be held legally responsible.

An even weaker principle could also be formulated, as follows:

> PRESUMPTIVE MORAL RESPONSIBILITY-ACCOUNTABILITY LINK: if Jay is morally responsible for significant damage caused by the AWS, then *pro tanto*, Jay should be held accountable for the damage.

This principle suggests that, where agents are morally responsible for some significant harms, then, absent excellent reasons to the contrary, they should have to face some adverse consequences (not necessarily legal punishment, however).

It seems to me that, were Jay *not* to be held accountable in some ways (legal, professional, reputational, etc.), then an important element of both backward-looking and forward-looking aspects of moral responsibility would be missing. With great responsibility comes accountability.

Though one important aspect of moral responsibility goes amiss in the absence of accountability, this is not to say that the moral gambit approach is entirely impotent without it. Imagine a case where Jay is not held accountable by anyone, but thinks *of himself* as being morally responsible for what happened. Personal consequences of accepting moral responsibility are no trivial matter. Moreover, Jay may

[2] To be clear, I'm not criticizing Taddeo and Blanchard's choice to focus on moral responsibility only. Their strategy to do so is, to my mind, well-justified.

also be blamed for the outcome by agents outside his institutional framework (or, at any rate, those who do not have the institutional capacity to hold him accountable). This may, in turn, affect both the backward-looking and forward-looking aspects of moral responsibility (e.g., identifying and blaming morally responsible people who escape accountability can serve to demonstrate failures in institutional accountability mechanisms).

However, in this paper, I explore the consequences of responsibility/accountability assignments within institutional settings, rather than as a matter of personal ethics. It seems to me that the institutional proposal for formalizing the moral gambit forms a significant enough part of Taddeo and Blanchard's article to justify focusing on it. Moreover, without properly functioning institutional accountability mechanisms, the moral gambit approach would nevertheless remain seriously incomplete. In proposing such mechanisms, if only in sketch form,[3] Taddeo and Blanchard, in my view, implicitly recognize this.

Within the institutional setting that Jay may inhabit (if he were, say, a member of the US military), neither legal responsibility nor accountability seem likely, regardless of the degree of moral responsibility Jay has. For starters, as a matter of empirical fact, government officials are held legally accountable not nearly as often as everyday people, and if they are, they are punished with much less severity. As Michael Huemer (2021) notes (and provides receipts):

> The legal system directs its efforts almost entirely toward private criminals. The wrongs done by government officials are rarely taken seriously, rarely investigated, and barely punished when uncovered … Government officials regularly regard themselves and each other as above the law. Even those who do not abuse their own power seldom make any serious effort to hold accountable those officials who do. In place of the shockingly punitive approach taken by the state toward private criminals …, we see a fantastic leniency, a readiness to see everything from the point of view of the perpetrator and to find any reason to avoid harsh punishments, when it comes to crimes committed by those in government. (173)

While Huemer talks about all government officials, the pattern he describes appears especially vivid when it comes to the military.[4] One need not look too far back to find cases where, intuitively, members of the military should be held accountable for doing certain things — and yet they were not.

On August 29, 2021, a "botched" US military drone strike in Kabul killed Zemari Ahmadi, an aid worker based in Afghanistan, along with seven children and two other innocent adults (Savage et al., 2022). The strike became a matter of

[3] Given this point, it may be objected that my criticism of the moral gambit approach is inapt. Taddeo and Blanchard offer a schematic and abstract proposal for the institutional implementation of the moral gambit, with many details to be filled out. It could be unfair to criticize them for failing to consider feasibility constraints at this stage of the argument. However, it seems to me that the feasibility concerns I am raising are general enough to be applicable even to very schematic solutions. Hence, they are worth considering if only as issues to be addressed by further research.

[4] See also Coyne (2022).

controversy, with wide media coverage and pressure mounting on the Pentagon to at least conduct an inquiry.

It is worth quoting at length a *New York Times* story on the actions taken by the US military to deal with those responsible. As the newspaper reported on December 13, 2021:

> None of the military personnel involved in a botched drone strike in Kabul, Afghanistan, that killed 10 civilians will face any kind of punishment, the Pentagon said on Monday [Dec 13, 2021].
>
> The Pentagon acknowledged in September [of 2021] that the last U.S. drone strike before American troops withdrew from Afghanistan the previous month was a tragic mistake that killed the civilians, including seven children, after initially saying it had been necessary to prevent an Islamic State attack on troops. A subsequent high-level investigation into the episode found no violations of law but stopped short of fully exonerating those involved, saying such decisions should be left up to commanders.
>
> Defense Secretary Lloyd J. Austin III, who had left the final word on any administrative action, such as reprimands or demotions, to two senior commanders, approved their recommendation not to punish anyone. The two officers, Gen. Kenneth F. McKenzie Jr., the head of the military's Central Command, and Gen. Richard D. Clarke, the head of the Special Operations Command, found no grounds for penalizing any of the military personnel involved in the strike (Schmitt, 2021, np.)

Tellingly, the very fact that the US military launched an investigation and admitted committing a "tragic mistake" (without holding any single person accountable) has been *praised* by commentators as a step in the right direction for its departure from the long-standing practice of not admitting any wrongdoing for civilian deaths due to drone strikes (Wargaski, 2022).

Furthermore, the incident described above seems to fit a long-recognized pattern. Military officials have shown very limited willingness to pursue accountability for people under their command. The problem appears institutional. As Amnesty International (2014) points out, in a report on the accountability for civilian casualties in Afghanistan, there are.

> important structural flaws in the US military justice system that hinder the investigation and prosecution of crimes against civilians. Most importantly, the military justice system is "commander-driven" and, to a large extent, relies on soldiers' own accounts of their actions in assessing the legality of a given operation. As a 2013 report of the Defense Legal Policy Board concluded, the functioning of the system depends very much on initial, ground-level reporting from troops at the point of contact. It is, in significant ways, a system of self-policing. Yet troops have scant incentive to report possible violations up the chain of command, and many reasons not to. Commanders, too, have little reason to push investigations forward, particularly in cases in which the commander's own conduct or judgment might be called into question. Because the military justice system lacks independent prosecutorial authorities, it is

> the commander who decides whether a case will be referred to trial, resolved administratively, or dropped altogether. Any prosecution, no matter how clearly in the interests of justice it is, can be vetoed by a defendant's commanding officer. Given these obstacles, it is no wonder that few cases make it to court. It is only in the rarest of circumstances—where fellow soldiers are so appalled by another soldier's behaviour that they insist on reporting it up the chain of command, where commanders support a prosecution, and, sometimes, where the media draws unwanted attention to flagrant abuses—that criminal cases involving civilian casualties go forward. (5)

Thus, even with flagrant cases of wrongdoing that can be traced to their source without excessive effort, where what is fair and just seems pretty clear, the military has little incentive to pursue justice in any meaningful way, *for structural reasons*.

It is unclear how we can trust institutions that make accountability decisions of this sort — in cases that are relatively highly publicized and involve loss of innocent life, rather than, say, "mere" significant non-lethal damage that Taddeo and Blanchard talk about — to implement any kind of recommendations, such as "ethics-based auditing," in a way that promotes (let alone ensures) meaningful responsibility and accountability.[5] In any case, Taddeo and Blanchard do not give any reasons to think militaries over the world would actually abide by their recommendations.

Moreover, as Richard Hanania (2022) explains, there is little hope of meaningful civilian-democratic oversight of whatever measures the military decides to enact, since the public, elected officials, and bureaucrats tend to have very little knowledge of foreign policy.

Hanania says:

> [in the arena of foreign policy, w]e can find clear examples of concentrated interests [e.g. the military; the weapons manufacturers] that should have an incentive to influence policy. Moreover, given how remote international affairs are from the experiences of most Americans, it is small wonder that regular citizens and even many government officials remain ignorant toward foreign policy issues, and the existence of classified information further hinders public understanding. Finally, the nature of geopolitics and the difficulties inherent in establishing causal relationships and making forecasts regarding major issues further cement the power of concentrated interests to shape public discourse. (129)

This prompts Hanania to conclude that "[l]eaders, like the public, can thus be easily led by those perceived to have foreign policy expertise" (ibid.). In many cases, those will be the high-ranking military officials.

Hanania adds that there is likewise little hope for vigorous journalistic oversight of these policies:

[5] Of course, any system of responsibility assignments can be inscribed in official documents. Whether people in charge of enforcing such rules and recommendations have an incentive to actually follow them is a different matter. Historical anecdote and the incentive structure of military justice suggest that there are serious feasibility constraints on how effective such a system can be in the real world.

> The process of the national security establishment managing public opinion has to a large extent been bureaucratized and operates under the radar. While the Pentagon employs thousands of military and civilian personnel to manage public opinion, as of 2010 only ten reporters covered the Pentagon full-time …, meaning that the press trying to hold government officials accountable is simply overmatched. (2022, 156)

The above quotation assumes that members of the press are incentivized to "ask hard questions" of military officials but are faced with severe resource constraints. However, elsewhere in the book, Hanania offers a less rosy picture of the journalistic ethos:

> Journalists and pundits who report on and shape public opinion tend to have extremely close relationships with other members of what is called 'the foreign policy community.' … In some cases, the lines between journalism, scholarship, and advocacy become blurred. … Even reporters for the most prominent news organizations rely on government officials for access and may be selected into their careers based on their belief in an interventionist American policy abroad… [Furthermore,] foreign policy reporting is shaped by powerful interests. In reporting on foreign affairs, the phrase 'government officials say' is virtually ubiquitous, and what follows is normally passed along as if it is a neutral piece of information rather than calculated to shape public perceptions of an issue. (58-9)

This assessment is borne out by other research. As Michel Haigh (2013) points out, "research shows war coverage by the media is largely uncritical and often patriotic …. When the press acts as a cheerleader, confidence in the press does not decline …. During the first Gulf War, 'media generally reacted with predictable boosterism' (Mueller, 1994, p. 74) instead of asking the difficult questions" (2469, other references omitted).

In this institutional setting of uninformed voters, uninformed elected officials, less-than-critical press coverage, and reliance on the military as unequivocal experts, cases of non-lethal significant damage caused by AWS are likely to be dealt with in much the same way as the botched drone strike described above, with no clear accountability enforced on anyone (especially given that they are likely to generate much less public and journalistic scrutiny, at least once the novelty of using "robots" in war wears off).

Such institutional responses to apparent mistakes and even wrongdoing, on the part of members of the military, are unlikely to "[foster] accountability …[or] enable forward-looking responsibility, insofar as the [unlikely] prospect of the blame … linked to a given decision/action would [not] facilitate morally sound choices and careful conduct" (Taddeo & Blanchard, 2022, 4). This is because those taking the gambit would likely not be blamed, or held accountable, except perhaps for the most egregious mistakes.

Hence, the accountability mechanisms, at least as enforced by the US military officials, seem to promote extreme leniency, in line with Huemer's comments. This is likely to lead to the overuse of AWS even in situations where benefits are hard to discern — and it may also lead to much of the significant damage that Taddeo and Blanchard rightly worry about.

In light of the above, we can tentatively conclude that it is altogether unlikely that such institutions as the Pentagon will take steps to implement Taddeo and Blanchard's policy recommendations, and, even if they were implemented, there is likely to be very little interest from the public, the elected officials, and the press, to see to it that they are working as designed (and, as the Amnesty International report makes clear, decision-makers within the military itself would have a strong incentive to avoid imposing accountability, if they can do it without provoking civilian outrage). For reasons specified by Hanania, it also seems unlikely that such mechanisms will be imposed upon the military by civilian leaders themselves (who will probably defer to expert judgment in such matters).

4 Overabundance of Caution

One objection to the critique I just offered is that it misses the target: Taddeo and Blanchard present an *idealized* account — something our imperfect institutions and leaders should try to aspire to. It is beside the point to say that the institutions are *in fact* imperfect — everyone agrees with that.

However, it seems to me that Taddeo and Blanchard's paper strays into the non-ideal when the authors outline how to build responsibility-enhancing institutions. This, I take it, is intended to be put to work in the real world, not its idealized version. (If it were otherwise, if all were behaving in an idealized manner, then, for example, ethics auditing would be superfluous.) So it is apt to look at real-world evidence to assess its feasibility. As I argued in the preceding section, the feasibility of such institutions achieving their goals is undermined by incentive structures of more or less all the actors involved.

Still, being hostage to empirical fortune, my argument could turn out to be mistaken. It could turn out that evidence and theorizing of a Hanania or a Coyne have been superseded by more plausible models, much friendlier to the feasibility of Taddeo and Blanchard's institutions. However, even supposing that the accountability for the use of AWS could be established and enforced along the lines envisaged by Taddeo and Blanchard, a different — in a way, the exact opposite — problem emerges.

The problem is as follows: A well-known critique of the Food and Drug Administration (FDA) alleges that the agency displays excessive cautiousness when it comes to their drug approval process (Higgs, 1994; Kazman, 1990; Miller, 2000; Peltzman, 1973). The FDA, so the criticism goes, is too conservative, not because its members are particularly cautious compared to the general population, but rather due to the incentive structures in place.

The argument goes like this: the FDA will get blamed for the actual harms, such as illnesses and deaths caused by a dangerous drug that the agency hastily allowed to go to market; but it will *not* get blamed for all the untreated illnesses and deaths caused by taking too long to approve a lifesaving drug. This is so even if the latter death toll may well be higher. Consequently, the FDA will tend to spend an excessive amount of time pondering whether to approve lifesaving drugs — which, ultimately, kills people.

There is an analogous worry that meaningful moral responsibility assignments proposed by Taddeo and Blanchard — if faithfully implemented and followed — would incentivize just such a potentially deadly conservatism. People deciding whether to take moral gambits will be aware that they'll be less likely to get blamed for delaying the deployment, or altogether refusing to deploy AWS, even if the AWS could have saved lives or prevented significant damage in the circumstances. In contrast, decision-makers *are* likely to be blamed for whatever disasters in fact occur due to the moral gambit going awry and the AWS causing unjustified damage. This applies not just to situations where the risks of using AWS are in fact too high. It likewise applies in cases where there are good reasons (perhaps even lifesaving reasons) to use AWS rather than, say, human combatants.

Consider a stylized example: as an ethics-based audit of the AWS responsibility assignments draws to a close, an opportunity arises to deploy the AWS in a certain location, where a risk–benefit analysis clearly shows it to be justified. Suppose that internal regulations and years-long practice discourage the use of systems currently under ethics audits. Kay, the decision-maker, aware of both the risk–benefit analysis and internal regulations and practice, decides to wait until the completion of the audit. Her reasoning: to launch the AWS, Kay will have to take the moral gambit and accept moral responsibility ex ante; if anything goes wrong, she'll receive harsh blame for not following established practice and acting hastily; if she does nothing, she will receive no blame. She chooses the latter course, even though deploying the AWS may have saved innocent lives.

Does that sound far-fetched? Consider the following example of the FDA's decision process, as recounted by Henry Miller:

> regulators… fear… being perceived as too eager to approve new products. In the early 1980s, when I headed the team at the FDA that was reviewing the NDA [New Drug Application] for recombinant human insulin, the first drug made with gene-splicing techniques, we were ready to recommend approval a mere four months after the application was submitted (at a time when the average time for NDA review was more than two and a half years). With quintessential bureaucratic reasoning, my supervisor refused to sign off on the approval -even though he agreed that the data provided compelling evidence of the drug's safety and effectiveness. "If anything goes wrong," he argued "think how bad it will look that we approved the drug so quickly." (When the supervisor went on vacation, I convinced his boss to sign off on the approval.) The supervisor was more concerned with not looking bad in case of an unforeseen mishap than with getting an important new product to patients who needed it. (2000, 41-2)

As with the FDA, the result of the bureaucratic incentive structure rewarding excessive caution may be that lifesaving AWSs are waiting idly in the wings rather than saving actual lives, because the moral gambit approach promotes an ultraconservative attitude towards risk. (Not taking) moral gambits may cost innocent lives.

There are at least two objections one could raise to the analogy between the FDA and the (properly working) moral gambit approach to responsibility assignments. First, without any empirical support, it is difficult to decide whether the differences between military institutions and regulatory bodies such as the FDA may serve to undermine the analogy. For instance, the FDA's mission is the protection of public health, and so prevention of death and disease. On the other hand, the purpose of the military is, at least in part, the infliction of harm (and death) on the enemy. Moreover, military personnel in decision-making capacity may be willing to tolerate more risks than the bureaucrats at the FDA. This may suggest that the problem of excessive caution is less of a concern with military organizations.

While these differences are real,[6] there is nevertheless reason to believe that, at least in some cases, excessive caution, driven by unwillingness to face blame for action rather than inaction, on the part of military organizations does lead to tragedy. It is a separate question whether such caution is driven by fear of blame, but common sense suggests it would be at least a part of an explanation.

The tragic Srebrenica massacre and, especially, its aftermath may serve as a case in point. The massacre was orchestrated by Bosnian Serb militia forces who killed around 8000 Bosnian Muslim men and boys near the town of Srebrenica in Bosnia and Herzegovina (Ryngaert & Schrijver, 2015). The massacre occurred despite the presence of United Nations (UN) peacekeeping troops in a nearby base. The troops arguably did not intervene because the peacekeeping mandate only allowed the use of lethal force in self-defense (Koster, 2020), and they were not the target of the attack.

The massacre, and the accompanying passivity of UN forces, prompted changes in the UN peacekeeping doctrine. As international lawyers Cedric Ryngaert and Nico Schrijver put it, "the peacekeeping missions stopped being just passive forces interposed between parties to a conflict; instead, almost all of them were endowed with a primary mandate to protect civilians" (2015, 222), including the permission to use lethal force to do so.

Now, I am not claiming that the only, or main, reason for UN troops' inactivity in the face of the Srebrenica massacre was the fear of being held responsible for violating the mandate. As with most such events, a variety of reasons was surely at play. However, instructively, even after the changes to the UN's peacekeeping doctrine, the use of lethal force is extremely rare. In the words of Ryngaert and Schrijver again, "[a]lthough most UN peacekeeping missions can currently use—lethal—force

[6] Though one may think that the military's official mission statements would emphasize national security and defense, just as the FDA would emphasize its own protective role in its mission, this isn't the case. While the FDA's mission statement claims the organization "is responsible for protecting the public health by ensuring the safety, efficacy, and security of human and veterinary drugs, biological products, and medical devices" (FDA, 2018, np), the U.S. Army, for instance, considers its purpose to be "To deploy, fight and win our nation's wars by providing ready, prompt and sustained land dominance by Army forces across the full spectrum of conflict as part of the joint force. The Army mission is vital to the Nation because we are the service capable of defeating enemy ground forces and indefinitely seizing and controlling those things an adversary prizes most – its land, its resources and its population." (The United States Army, nd, np).

to protect civilians—so not only in self-defence, in practice it turns out that such force is hardly used" (ibid., 223). As the authors bemoan,[7]

> [i]t is most unfortunate, 20 years after Srebrenica, that troop-contributing nations still consider that the risks of using force and even of having boots on the ground are too high, that peacekeepers are not always following orders of the UN but rather of their own capitals, that non-compliance with UN orders is not reported, that hierarchical decision-making causes delays in reaction, that missions are weak and spread too thinly, that proper information-gathering fails, and that *peacekeepers fear penalties for action rather than for inaction* (ibid., 223-4, emphasis added).

At least in the view of these scholars (as the italicized fragment suggests), military personnel's excessive caution due to fear of being held responsible is *a contributing factor to the decision not to deploy lethal force and, potentially, incur loss of civilian life*. Though other motivations may surely counteract it in any given case, it remains at least an element of the motivational economy of military decision-makers. Consequently, it is reasonable to expect, all the differences notwithstanding, that the concern articulated by the analogy with the FDA needs to be taken seriously. There is a risk that the moral gambit approach will, in a sense, work "too well" and the relevant decision-makers will be incentivized to deploy AWSs less frequently than the situation demands.

A different objection could be that the story of Kay that I told earlier is too naive to be comparable to real-world military decision-making, with responsibility likely distributed among a number of agents, rather than focused in a single person, as my analogy describes. However, the evidence presented above suggests that despite the (presumably[8]) distributed responsibility, the problem of excessive caution can also plague real-life militaries. Hence focusing on a single decision-maker for ease of exposition needn't undermine the analogy.[9]

5 Conclusion

I take Taddeo and Blanchard to be making the following argument: meaningful moral responsibility for using AWS is supposed to be both forward-looking and backward looking. It is supposed to incentivize careful scrutiny of reasons to deploy AWS, and to promote accountability for deployment decisions. To facilitate meaningful moral responsibility assignments, we have a set of institutional recommendations to be implemented prior to AWS use.

[7] The authors base this assessment at least in part on a report written by the UN Office of Internal Oversight Services.

[8] If responsibility was *not* distributed in the real-life cases I cited, then it is also not naive to analogize Kay to real-life cases.

[9] How do we reconcile the evidence cited here with that presented in the preceding section? There seem to be a number of differences between the two cases. I would speculate that chief among them is the interaction and potential conflict between orders from the UN and those from national governments of the troop-supplying nations. But I have no space to pursue this suggestion further.

I sought to identify some problems with this proposal: first, experience suggests that the institutions most likely to deploy AWS have a history of not assigning responsibility and executing accountability for apparently egregious lethal errors of their members; second, there is a theoretical explanation for that: the incentive structure of military and civilian leaders, as well as the public opinion, makes such outcomes all the more likely — and suggests that, whatever official policies are implemented, including ones that Taddeo and Blanchard propose, relative lack of scrutiny from civilian officials, public opinion, and the press, may render them toothless.

On the other hand, if the accountability mechanisms actually were dentate, this might incentivize people to avoid taking moral gambits even if the situation requires it, merely for fear of being blamed should anything go wrong.

I am not arguing, though, that this should lead us to abandon the moral gambit approach, or indeed, that the status quo is somehow justified and that military organizations should not be held to high moral standards. Rather, it seems that further research is required to determine whether the faults — whether its implementation fails or succeeds in the real world — are weighty enough to prompt a search for a different solution. In a way, I see one contribution of my article as stressing the need to assess how each solution to the problem of the responsibility for the use of AWS alters the balance between what we might call "false negatives" (not deploying the weapons when the situation calls for it) and "false positives" (deploying them when one shouldn't). Responsibility assignments (plus the associated institutions for fostering accountability) are one way of attempting to achieve such balance, but a further (or perhaps a more fundamental) question is which types of errors we should be focused on minimizing, given real-world constraints.

Abbreviations *AWS*: Autonomous weapons systems; *FDA*: The Food and Drug Administration; *NDA*: New drug application; *UN*: United Nations

Acknowledgements The author wishes to thank two anonymous reviewers for this journal for many helpful comments and suggestions.

Author Contribution BC — 100%.

Data Availability Not applicable.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The author declares no competing interests.

References

Amnesty International. (2014). *Left in the dark: Failures of accountability for civilian casualties caused by international military operations in Afghanistan — Summary*. Amnesty International Ltd. Available at: https://www.amnesty.org/es/wp-content/uploads/2021/06/asa110082014en.pdf.

Champagne, M., & Tonkens, R. (2015). Bridging the responsibility gap in automated warfare. *Philosophy & Technology, 28*(1), 125–137.

Chomanski, B. (2021). Liability for robots: Sidestepping the gaps. *Philosophy & Technology, 34*(4), 1013–1032.

Coyne, C. (2022). *In search of monsters to destroy: The folly of American Empire and the paths to peace*. Independent Institute.

Food and Drug Administration. (2018). The FDA mission. Available at: https://www.fda.gov/about-fda/what-we-do. Accessed 18 Apr 2023

Gordon, J. S., & Nyholm, S. (2021). Ethics of artificial intelligence. *Internet Encyclopedia of Philosophy* np. https://iep.utm.edu/ethics-of-artificial-intelligence. Accessed 10 Dec 2022

Haigh, M. M. (2013). The relationship between the media, the military, and the public: Examining the stories told and public opinion. In A. Valdivia (Ed.), *The international encyclopedia of media studies volume V: Media effects/media psychology* (pp. 2468–84). Blackwell.

Hanania, R. (2022). *Public choice theory and the illusion of grand strategy: How generals, weapons manufacturers, and foreign governments shape American Foreign Policy*. Routledge.

Higgs, R. (1994). Banning a risky product cannot improve any consumer's welfare (properly understood), with applications to FDA testing requirements. *The Review of Austrian Economics, 7*(2), 3–20.

Huemer, M. (2021). *Justice before the law*. Palgrave Macmillan.

Kazman, S. (1990). Deadly overcaution: FDA's drug approval process. *Journal of Regulation and Social Costs, 1*(1), 35–54.

Koster, M. (2020). The Netherlands looks again at its controversial role in Srebrenica. *Balkan Transitional Justice.* Available at: https://balkaninsight.com/2020/07/10/the-netherlands-looks-again-at-its-controversial-role-in-srebrenica/. Accessed 19 Apr 2023

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology, 6*(3), 175–183.

Miller, H. (2000). *To America's health: A proposal to reform the Food and Drug Administration*. Hoover Institution Press.

Mueller, J. (1994). *Policy and opinion in the Gulf War*. University of Chicago Press.

Nyholm, S. (2018). Attributing agency to automated systems: Reflections on human-robot collaborations and responsibility-loci. *Science and Engineering Ethics, 24*(4), 1201–1219.

Peltzman, S. (1973). An evaluation of consumer protection legislation: The 1962 drug amendments. *Journal of Political Economy, 81*(5), 1049–1091.

Ryngaert, C., & Schrijver, N. (2015). Lessons learned from the Srebrenica massacre: From UN peacekeeping reform to legal responsibility. *Netherlands International Law Review, 62*(2), 219–227.

Savage, C., et al. (2022). Drone strike video shows killing of civilians in Afghanistan. *The New York Times* (January 19, 2022). Available at: https://www.nytimes.com/2022/01/19/us/politics/afghanistan-drone-strike-video.html. Accessed 10 Dec 2022

Schmitt, E. (2021). No U.S. troops will be punished for deadly Kabul strike, Pentagon Chief decides. *The New York Times* (December 11, 2021). Available at: https://www.nytimes.com/2021/12/13/us/politics/afghanistan-drone-strike.html. Accessed 10 Dec 2022

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

Taddeo, M., & Blanchard, A. (2022). Accepting moral responsibility for the actions of autonomous weapons systems—A moral gambit. *Philosophy & Technology, 35*(3), 1–24.

The United States Army. (nd). The army's vision and strategy. Available at: https://www.army.mil/about/. Accessed 18 Apr 2023

Wargaski, R. (2022). U.S. drone warfare and civilian casualties. *Eagleton Political Journal*. Available at: https://eagletonpoliticaljournal.rutgers.edu/us-the-world/u-s-drone-warfare-and-civilian-casualties/. Accessed 10 Dec 2022