Robots and Respect: A Response to Robert Sparrow

Ryan Jenkins and Duncan Purves

Robert Sparrow recently argued in this journal that several initially plausible arguments in favor of the deployment of autonomous weapon systems (AWS) in warfare are in fact flawed, and that the deployment of AWS faces a serious moral objection.[1] Sparrow's argument against AWS relies on the claim that they are distinct from accepted weapons of war in that they either fail to transmit an attitude of respect for enemy combatants or, worse, they transmit an attitude of disrespect. In this reply we argue that this distinction between AWS and widely accepted weapons is illusory, and therefore cannot ground a moral difference between AWS and existing methods of waging war. We also suggest that if deploying conventional soldiers in a given situation would be permissible, but we could expect to cause fewer civilian casualties by instead deploying AWS, then it would be consistent with an intuitive understanding of respect to deploy AWS in this situation.

Sparrow's Objection

Drawing on Thomas Nagel's influential "War and Massacre,"[2] Sparrow argues that AWS fail to manifest an attitude of respect toward the adversary. Quoting Nagel, he says that during warfare, "whatever one does to another person intentionally must be aimed at him as a subject . . . [and] should manifest an attitude to *him* rather than just to the situation" (p. 106). For Sparrow, establishing this interpersonal relationship with the adversary requires at least "acknowledging the morally relevant features that render them combatants or otherwise legitimately subjected to a risk of being killed" (p. 107). Deploying AWS, he argues, prevents the establishment of this interpersonal relationship since "in some fundamental sense there is no one who decides whether the target of the attack should live or die" (p. 107).

Ethics & International Affairs, 30, no. 3 (2016), pp. 391–400.

doi:10.1017/S0892679416000277

Before turning to our specific critiques of Sparrow's argument, it is important to highlight several virtues of Sparrow's criticism of the deployment of AWS. First, some existing methods of waging war appear to be objectionable precisely because they are, in some sense, disrespectful. Consider, for example, land mines and nuclear weapons, two roundly criticized methods of waging war. One plausible explanation of the common and deep moral aversion to the use of these weapons is that they sever the interpersonal relationship between the person deploying them and his or her targets, and thus the use of such weapons fails to respect the humanity of the enemy. Second, Sparrow advances the debate about the ethics of deploying AWS by presenting a noncontingent objection to AWS. Failing the principle of "respect," the deployment of AWS would be morally problematic *even if* AWS became better than human soldiers at discriminating between combatants and noncombatants, did not ultimately lower the threshold for going to war, and did not lower the value placed on the lives of those targeted by them. Thus, Sparrow's objection, if successful, would constitute a powerful and enduring reason to oppose the deployment of AWS. As we have argued elsewhere, in anticipation of the advance of machine learning and machine vision, theorists should examine the ethical status of delegating to machines the task of killing itself, rather than resting their objections on the inferior performance (for now) of machines versus humans. We applaud Sparrow's contribution to this discussion.[3]

Cruise Missiles, AWS, and Respect

If the disrespectful nature of using AWS is to ground a morally important distinction between AWS and accepted weapons of war, then there must be some way in which those deploying accepted weapons are respectful of those they target, whereas those deploying AWS are not. If there is no relevant distinction, then Sparrow's argument has far-reaching implications: it would force us to abandon technologies that have been widely used for over a century. We argue that there is no feature of AWS that makes their use less respectful of their targets than the use of cruise missiles or long-distance artillery.

Our argument here will proceed by way of a disjunction: Either AWS choose and engage their targets deterministically or else they are "artificial agents" (to use Sparrow's phrase), and have some species of free will that allows them to act nondeterministically. In either case, there is no apparent distinction vis-à-vis

respect between the use of AWS and the use of widely accepted long-range weapons. Consider each of these possibilities in turn.

If AWS are deterministic, then their operation is no different from other weapons of war that seem morally acceptable. Recall that respect for a target requires establishing an interpersonal relationship between the attacker and the target, one whereby the attacker acknowledges the features that render the target legitimately subjected to potentially lethal violence. Consider several plausible conditions for forming this sort of interpersonal relationship. Perhaps an interpersonal relationship can only be formed between an attacker and target if the attacker stands in close proximity to the target. But if that is the case, then many commonly accepted methods of waging war are immoral because they fail to satisfy this requirement. Dropping bombs from high-altitude aircraft does not involve any close personal contact with enemy combatants, and neither do less advanced weapons such as long-range artillery.

Nor can the requirement be that the attacker knows with certainty the identities of her targets. This would rule out cruise missiles, which may choose from among several possible targets when arriving at a kill box, or long-range artillery, which may rain down on targets whose identities are unknown. This also rules out the possibility that an attacker might establish an interpersonal relationship by *deciding* whether the target of the attack should live or die, for one cannot make such a decision when one does not know the target's identity when launching the attack.

One might argue that cruise missiles and long-range artillery still allow the transmission of an attacker's intention, insofar as they allow commanders to transmit their violence in a causally direct way against the adversary. To deploy AWS, by contrast, is to funnel one's intention through an intermediary that could be an "artificial agent," and so is an originator of actions. This brings us to the other disjunction: the possibility that AWS have some species of free will that allows them to target nondeterministically. Thus, the original commander's intention can get muddled by the deployment of AWS, and there is no longer the direct interpersonal link between a commander and the adversary. Sparrow argues that AWS are relevantly different from other weapons for this very reason.

The artificial agency of AWS might muddle the transmission of intention in two apparent ways. First, it could introduce additional causal links between the commander and the target. But multiplying the number and complexity of the causal links between a commander and the adversary does not plausibly make a moral difference.[4] For example, it makes no moral difference whether I fire a gun at

someone or press a button that *then* fires a gun at them. Nor would this distinguish AWS from long-range artillery and cruise missiles, which also introduce additional causal links between the commander and target, thereby weakening the interpersonal relationship.

Second, deploying AWS could muddle the commander's intention by introducing some uncertainty as to whether the commander's intention will successfully be transmitted to the adversary. However, the introduction of epistemic uncertainty does not seem to make a moral difference here, either. Reducing the certainty that the commander's intention will be transmitted to the target is relevant only if it is morally required to transmit this intention. We argue below, however, that this is not necessary. If it is not morally required to transmit an intention to the target, then it cannot be morally relevant that using AWS makes us *less certain* that one will transmit an intention. Sparrow is right that AWS differ from long-range artillery and cruise missiles in these ways. However, neither of these distinctions is relevant to the permissibility of long-range weapons. Any difference between AWS and cruise missiles in these two regards is a difference of degree, not of kind.

Finally, consider a stronger claim that AWS, because they are indeterministic "artificial agents" who have free will, do not simply muddle but obliterate the intention of the commander. Because the choice originates fully within the AWS there is a metaphysical indeterminacy regarding whom they will select for destruction. We contend that even if the targeting choice of an AWS were metaphysically indeterminate, this would not make its deployment morally problematic.[5]

Suppose, for example, that we plan to launch a cruise missile at an enemy compound. Suppose further that the warfighters at this compound are determined by an indeterministic computer, so that every morning it randomly selects a different group of warfighters to call into work. In this case we see no reason to say that the transmission of intention is broken in a way that is morally problematic. If it would be permissible to attack this compound regardless of who occupies it on any given day, then the fact that there is no fact of the matter about who will be present on the day of attack cannot make it impermissible. The same applies if the transmission of intention is broken by the weapon itself, rather than by an enemy's computer algorithm. Imagine, for example, a cruise missile that is designed to decide between equally liable targets such that there is no fact of the matter at launch as to which target it will engage. These examples suggest that we may be permitted to "pull the trigger" if we have good reason to think that

our targets are liable to attack *whoever they are*, that is, even when their identities are metaphysically indeterminate.

Ultimately, however, we draw a different lesson from the above comparisons between AWS and existing long-range weapons; that is, it is mysterious how AWS or other widely accepted forms of weaponry establish *any* meaningful interpersonal relationship in cases like these. This view is motivated by a simple claim, which is that it is impossible to establish an interpersonal relationship with persons whose identities are either unknown or metaphysically indeterminate. And yet we routinely use long-range weapons to target people whose identities are either unknown or, as we can suppose in the example above, metaphysically indeterminate. These weapons remain widely accepted despite not establishing the interpersonal relationship that Sparrow argues is necessary for respect for one's target, thus casting doubt on the claim that this relationship is a necessary feature of waging just war.

Sparrow concedes that epistemic uncertainty and metaphysical indeterminacy regarding the identities of targets fail to establish that only AWS, and not cruise missiles or long-range artillery, sever the interpersonal relationship between an attacker and a target, but he contends that this similarity may in fact count against the permissibility of both types of weapons:

> It is important to note that this comparison is not entirely favorable to either AWS or other sorts of weapons. People often *do* feel uneasy about the ethics of anonymous long-range killing (p. 108).

But—and this point is crucial—condemning cruise missiles and long-range artillery renders Sparrow's position significantly more extreme than the one he set out to defend. Rather than merely justifying a prohibition on an emerging technology whose permissibility is widely questioned, Sparrow's respect-based objection to AWS has the result of condemning established methods of warfare whose use is widely accepted. Indeed, if sound, his arguments would seem to require a return to the bad old days of bloody, close-range, ground warfare, which ensures massive casualties on both sides of a conflict. Only then could we be sure to establish the interpersonal relationship that "respect" requires.

Disrespect and Popular Opinion

Sparrow later suggests that in many cases—and at least, we might add, for the near future—AWS will continue to allow the formation of an interpersonal relationship

between commander and adversary, but that it will be a morally problematic relationship, namely, one of *dis*respect. Sparrow's evidence for this claim is that public opinion surveys repeatedly show a strong aversion to the use of killer robots as tools of war. The most plausible interpretation of this hostility, Sparrow claims, is that "Most people already feel strongly that sending a robot to kill would express a profound disrespect of the value of an individual human life" (p. 109). We do not deny that public aversion to a technology counts against its adoption, but public opinion can be swayed by an array of factors, only some of which are indicative of the moral truth of a matter. For example, people are generally less tolerant of risks associated with novel technologies than they are of risks associated with familiar ones.[6] It is precisely because of biases against novel technologies that ethicists should not be satisfied to let public opinion carry the day, especially in the absence of a robust moral distinction that makes the use of AWS disrespectful in a way that the use of other widely accepted weaponry is not.

Minimizing Civilian Casualties and Reasonable Expectations

We have argued that Sparrow fails to identify a way in which the deployment of AWS is uniquely disrespectful toward the people they target. We will now argue that the use of AWS is, in one important way, *more* respectful of human life than the use of cruise missiles or human soldiers.

Suppose that AWS will eventually be *better* at avoiding civilian casualties than human-controlled weapons.[7] In this case, the deployment of AWS surely also expresses a kind of respect for the civilians who *would have otherwise been killed* as a consequence of the mistakes made by human warfighters. Choosing a more conventional means of warfare when we have an alternative means that minimizes the risk of death to innocent civilians surely expresses a *lack* of regard for the humanity of those civilians. By using AWS instead, we express our intention to make every possible effort *not* to kill through error. If we are concerned with respect directed toward those who are killed by AWS *and* toward those whose lives are protected by AWS, the consideration of respect does not clearly tell against the deployment of AWS. The requirement of discrimination underscores the need to respect those who are not only the direct targets of our attacks but those whose fate depends on the method of fighting we choose.

Sparrow argues that appealing to the possibility that AWS may one day kill fewer civilians than human soldiers sets the wrong standard for establishing their ethical use. When asking whether the deployment of AWS is morally permitted, he argues that one should not ask whether they *will* kill fewer civilians than human soldiers, but rather what sort of conduct it is reasonable to *expect* from them. He suggests that perfection is the appropriate moral expectation when deploying human soldiers to battle. One can reasonably expect humans—but not AWS—to target zero illegitimate targets in war, so appealing to the fact that AWS may in fact kill fewer innocents cannot justify their deployment. We think he is mistaken.

Sparrow says that the relevant expectation to set for humans is perfection because humans have both the "power and the freedom" not to deliberately target noncombatants or use disproportionate force (p. 104). It is important to distinguish here between two notions of expectation that might be conflated: one is statistical and one is normative. If we understand "expectation" statistically, where our reasonable expectations are determined by how humans and AWS have behaved in the past, then perfection is not an appropriate standard for humans. Sparrow himself admits that human beings are not *actually* perfect decision-makers in war, and he admits that machines may be *better* decision-makers in many contexts (p. 96). Thus, we can expect (that is, predict) that deploying AWS rather than human soldiers would be a way of lowering casualties. Certainly, if we are to compare the best possible humans to the best possible AWS currently available in setting our expectations, it will not be a contest. The humans would win. However, if we are to compare the most common human behavior to the near-future prospects for AWS, then we may soon find that AWS are superior. Thus, if we understand "expectation" statistically, then the appropriate standards are the exact reverse of what Sparrow argues: in the not-too-distant future it will be more appropriate to expect (predict) results that are closer to perfection from a sufficiently sophisticated machine than from a human soldier. Perhaps, however, Sparrow does not have the predictive or statistical sense of "expectation" in mind, but is instead concerned with a deeper question about what sort of conduct it is reasonable or appropriate to morally *demand* of a human agent compared with a machine. This would explain why the appropriate moral standard for human soldiers is perfection.

Would it be reasonable to expect (that is, *morally demand*) of a machine that it will never target noncombatants? The answer must be "no," since machines are

not moral agents. When it comes to the question of what it is reasonable to morally demand of a machine, Sparrow is clearly right that we can morally demand much more of human soldiers than of AWS. But that is because it is not reasonable to morally demand *anything* of machines. However, it is difficult to see how the mere fact that it is appropriate to make moral demands of human soldiers but not of machines could render the deployment of AWS morally problematic, especially given that we can *expect* (predict) AWS to be at least as good as, if not better than, human soldiers at meeting the requirements of discrimination and proportionality. If our concern is adherence to proportionality and discrimination, then AWS are still the preferred option.

Sparrow argues that accepting this argument requires adopting a controversial consequentialist framework. However, given the concern of deontologists with not violating the rights of innocents not to be harmed, deontologists should accept that we have a strong moral reason to deploy AWS in the place of human soldiers *if doing so will predictably minimize civilian casualties*. This reasoning is only problematically consequentialist if it ignores some further deontological constraints, such as those against disrespect.[8]

Finally, a few words on consequentialist reasoning. It is true that consequentialist reasoning has been relegated to the sidelines of just war theory for much of its history. Nagel's "War and Massacre," the spiritual progenitor of Sparrow's article, is perhaps the best known explication of this alleged incompatibility between just war theory and consequentialist reasoning. But surely there is a role for consequentialist reasoning to play in preferring, for example, fewer civilian deaths to more civilian deaths, other things being equal. Consider a distinction from Jeff McMahan: when considering whether a use of force is proportionate, we can distinguish between its narrow and its wide proportionality.[9] A weapon is proportionate in the narrow sense when it inflicts proportionate harm on the people who are liable to be harmed. But that same weapon may be criticized for being disproportionate in the wide sense when it subjects those *not* liable to harm to a disproportionate risk of harm—so-called collateral damage.

Sparrow, in his discussion of respect, is at risk of conflating two senses of the term. For most of his article, he is concerned with whether the use of AWS is respectful toward the people who are actually killed by them, presuming they are liable to be killed. We can call this respect in the narrow sense. However, Sparrow's discussion slides into considering whether deploying AWS is respectful toward innocent civilians who are put at risk. We can call this respect in the wide

sense. What determines whether a weapon is respectful in the wide sense must be, at least in part, the extent to which it subjects civilians to harms that could be avoided by other weapons. Thus, choosing to use conventional weapons when we could use AWS, supposing AWS are a more reliable means of minimizing harm to civilians, would be disrespectful toward civilians in the wide sense. More simply, if the use of force is justified with regard to one method of waging war, and if we have reason to believe that another method of waging war is more discriminate, then it is clearly wrong (viz. disrespectful) to choose the less discriminate method.[10] We feel Sparrow overlooks this conception of respect.

Sparrow would likely argue that this sidesteps the issue of whether AWS are *mala in se*, and he is right. But until that is compellingly shown, we feel that AWS have the potential to be more respectful to civilians who would be at greater risk from other methods of waging war. Both deontologists and consequentialists should be motivated to deploy weapons that are respectful in this sense.

Conclusion

The moral status of the use of AWS sits at the locus of some of the most vexing issues in philosophy: action theory, liability to harm, and moral responsibility. Philosophers are only just beginning to grapple with the ethical questions posed by this developing technology, and we admire the contributions of Sparrow and other critics to these debates, even if we find their arguments ultimately unconvincing. We maintain that, if truly autonomous weapon systems become feasible, the militaries of the world will find their advantages irresistible. Our preceding arguments notwithstanding, we remain concerned that deploying AWS will be tantamount to deploying psychopaths who are unable to appreciate moral reasons, and whose decisions are therefore bound to be deficient.[11] And so we sympathize with the swelling chorus of activists, philosophers, and humanitarian lawyers who are calling for a ban on the development or deployment of AWS, at least as a precaution until they can be more fully examined from a moral standpoint. Though it is conceivable that expert opinion will ultimately settle on the side of preferring AWS to human soldiers, we should not allow the progress of our technology to outstrip the progress of our wisdom.

NOTES

[1] See Robert Sparrow, "Robots and Respect: Assessing the Case against Autonomous Weapon Systems," *Ethics & International Affairs* 30, no. 1 (2016), pp. 93–116.
[2] Thomas Nagel, "War and Massacre," *Philosophy & Public Affairs* 1, no. 2 (1972).

[3] Duncan Purves, Ryan Jenkins, and Bradley J. Strawser, "Autonomous Machines, Moral Judgment, and Acting for the Right Reasons," *Ethical Theory and Moral Practice* 18, no. 4 (2015), pp. 851–72.
[4] Alastair Norcross, "Off Her Trolley? Frances Kamm and the Metaphysics of Morality," *Utilitas* 20, no. 1 (2008), p. 65.
[5] To be sure, the possibility of metaphysical indeterminacy in targeting decisions seems to be the impetus for the "responsibility gaps" objection, which Sparrow notes. We have addressed this objection elsewhere. See Purves, Jenkins, and Strawser, "Autonomous Machines."
[6] See Paul Slovic, "Perception of Risk," *Science* 236 (1987), pp. 280–85. See also Chauncey Starr, "Social Benefit Versus Technological Risk," *Science* 165 (1969), p. 1232.
[7] Indeed, Sparrow is willing to entertain this possibility. We, in fact, think the outcome is quite likely. Sparrow is worried, and justifiably so, about a machine's ability to understand and appreciate the nature of morality as a meaning-laden and contextual domain of knowledge and behavior. However, the results of recent advances in machine learning, which have been nothing short of staggering, have rendered moot these concerns about machine "understanding." AlphaGo and Watson have made it clear that machines can outperform humans in domains where we once thought we enjoyed a great privilege and indomitable superiority. And this is true whether these machines *understand* the context in which they are acting, or the meaning and significance of their choices.
[8] The fact that we cannot legitimately *demand* that AWS minimize civilian casualties seems significant only if it generates a "responsibility gap" that renders attributions of responsibility for the actions of AWS difficult or impossible. But this is not a new problem. For discussions of the problem of responsibility attributions, see Andreas Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata," *Ethics and Information Technology* 6, no. 3 (2004), pp. 175–83; Robert Sparrow, "Killer Robots," *Journal of Applied Philosophy* 24, no. 1 (2007), pp. 62–77; and Heather Roff, "Killing in War: Responsibility, Liability, and Lethal Autonomous Robots," in Fritz Allhoff, Nicholas G. Evans, and Adam Henschke, eds., *Routledge Handbook of Ethics and War: Just War Theory in the Twenty-First Century* (Milton Park, Oxon: Routledge, 2013). The inability to make moral demands of machines may ultimately count *against* deploying human soldiers and in favor of deploying AWS. Michael Robillard and Bradley Strawser ["The Moral Exploitation of Soldiers," *Public Affairs Quarterly* 30, no. 2 (2016)] have argued that soldiers are often victims of "moral exploitation" by having moral responsibility "outsourced" to them in virtue of their vulnerable position. Replacing human soldiers with AWS holds the potential to resolve this deontological worry about exploitation.
[9] Jeff McMahan, *Killing in War* (New York: Oxford University Press, 2009).
[10] Ryan Jenkins, "Cyberwarfare as Ideal War," in Adam Henschke, Fritz Allhoff, and Bradley Strawser, eds., *Binary Bullets: The Ethics of Cyberwarfare* (New York: Oxford University Press, 2016).
[11] Purves, Jenkins, and Strawser, "Autonomous Machines."