Simpson, Thomas W and Müller, Vincent C. (2016), 'Just war and robots' killings', *Philosophical Quarterly,* 66 (263), 302-322.
http://www.sophia.de
http://orcid.org/0000-0002-4144-4957

Just war and robot's killings

Thomas W. Simpson & Vincent C. Müller

Philosophical Quarterly 2016, 66 (263), 302-22

University of Oxford & Anatolia College/ACT
http://www.sophia.de

Abstract
May lethal autonomous weapons systems—'killer robots'—be used in war? The majority of writers argue against their use, and those who have argued in favour have done so on a consequentialist basis. We defend the moral permissibility of killer robots, but on the basis of the non-aggregative structure of right assumed by Just War theory. This is necessary because the most important argument against killer robots, the responsibility trilemma proposed by Rob Sparrow, makes the same assumptions. We show that the crucial moral question is not one of responsibility. Rather, it is whether the technology can satisfy the requirements of fairness in the re-distribution of risk. Not only is this possible in principle, but some killer robots will actually satisfy these requirements. An implication of our argument is that there is a public responsibility to regulate killer robots' design and manufacture.

1. Introduction

Rich countries are increasingly using semi-autonomous weapons platforms in war. The examples that have captured public imagination are the Predator and Reaper 'drones'. For those who can afford them, these systems are expendable in a way that humans are not. Development of such technology is set to continue, so that these platforms can be used to fulfil more roles than they currently do, and fulfil more effectively those roles in which they are currently employed (US Department of Defense 2013; Singer 2009). The same trajectory of development also aims at the development of autonomous machines, which—who?—can do our soldiering for us. Call these latter *killer robots*. We are not there yet, but there is good reason to think that we will one day be.

But would it be morally permissible to deploy killer robots? Those arguing for its moral impermissibility have dominated the philosophical debate. It is a minority report that has argued for its permissibility, under at least some circumstances, and it has done so on a consequentialist basis. Ronald Arkin's work exemplifies this: so long as fewer lives would be lost as a result of killer robots' warfighting than that of humans, then their deployment is justified (2009, 2010; also Sullins 2010). This paper takes the latter position but on different grounds. Just War theory assumes a structure of right in which aggregated welfare is not the decisive moral consideration. Our task is to show how, assuming this same structure, robots may kill. This is needed because the most important argument against killer robots, the responsibility trilemma posed by Robert Sparrow, adopts the same assumptions. (For a survey and rejection of other moral arguments given against killer robots; see Müller, Forthcoming.) Sparrow's argument serves as a foil to develop a positive account of responsibility attribution for technologies, and the demands of such in war. On this account, the crucial question is not whether, for every death in war, someone is morally responsible. It is whether killer robots can satisfy the requirements of fairness in the re-distribution of risk that they impose. A consequence of our view is that national governments and international governance structures have a responsibility for the regulation of killer robots.

The issue is timely and important. The United Nations' Convention on Certain Conventional Weapons is at the time of writing engaged on a consultative process regarding whether killer robots should be banned. One implication from this is that a kind of realist argument for their permissibility is undermined. The realist argues that the lust for power will result in others developing killer robots anyway, so it can be only a form of moralism—in Tony Coady's sense, as utopian moral thinking—which requires us to forgo such weapons at considerable risk to our own security (Coady 2008: 14). The argument is undermined because a similar process has already resulted in wide agreement that another technology should not be used in war, viz. anti-personnel landmines under the 1997 Ottawa Treaty. There is special merit in addressing the question now, before the technology is developed. This is on the psychological principle that it is easier to forgo a prospective gain than make a sacrifice, after the benefit has been incorporated in the baseline. In this case the gain or sacrifice is one of security. While most of the policy debate addresses whether killer robots could comply with international law, variants of Sparrow's objection generally form the central moral case. For instance, it is specifically credited as such by a prominent contribution, namely that by Human Rights Watch (2012: 42-5).

The paper proceeds as follows. We first state Sparrow's argument. He argues that there is a 'responsibility gap' for robots' killings in war, due to the kind of autonomy they could possess, and that this renders their deployment impermissible (§2). We deny that such a responsibility gap could exist for killer robots (§3). We then introduce the concept of a 'tolerance level' for engineered systems that have lower levels of au-

tonomy than that considered by Sparrow, which explains how responsibility and blame for malfunction are assigned. We outline how the same structure of responsibility attribution applies to killer robots (§§4-5). The crucial question here is whether the tolerance level for killer robots is 'strict'. Could any malfunctions in the exercise of lethal force by robots be permissible? We show that they could; our argument depends on a widely acceptable, if not wholly uncontroversial, claim about the ethics of risk imposition (§6). Replies to possible objections follow (§7). In conclusion, we identify the responsibility for regulation (§8).

2. Partial autonomy

Sparrow poses the following case.

> Imagine that an airborne AWS [Autonomous Weapon System], directed by a sophisticated artificial intelligence, deliberately bombs a column of enemy soldiers who have clearly indicated their desire to surrender. The AWS had reasons for what it did ... [but] they were not the sort to morally justify the action. Had a human being committed the act, they would immediately be charged with a war crime. Who should we try for a war crime in such a case? (2007: 66-7)

The concluding question contains his argument *in nuce*. Someone needs to be tried for the war crime; nobody is identifiably responsible; so no means of war that will likely lead to such killings is permissible.

Focus on the second premise. Sparrow contends that there is nobody who could in principle be held responsible for robots' unjust killings. He identifies three possible loci of responsibility. In arguing that none are appropriate, he poses a trilemma. The designers of the killer robot ought not be held responsible, for the robot is autonomous by design. The task for robotics engineers is to build systems that will learn from experience and adapt future behaviour in ways that are not scripted. Success in doing so renders the outcome of the system's action no longer controllable or predictable by the designer; so they are no longer responsible. Nor should the military commander who deployed the killer robot be held responsible; again, the robot is autonomous by design. This distinguishes killer robots from long-range area weapons such as artillery (70). The responsibility relation for such a commander is thus akin to that which normally holds between her and a subordinate. The former issues orders, the outcome of compliance with which she is responsible for, but she is not responsible for the subordinate's non-compliance. Finally, the machine cannot be held responsible. Responsibility is appropriate only if the subject is an appropriate target of praise and blame, punishment and reward. Consider punishment. Robots cannot be punished, at least not until they have the internal complexity sufficient for the frustration

of desires and of suffering. But at the point when they are capable of this, they will likely possess the autonomy constitutive of full moral agents. Because their lives will then have moral value, there would be no moral saving in sending them to war in place of human soldiers (73).

The trilemma turns on the existence of what Andreas Matthias has called a 'responsibility gap', in which no person is responsible for the actions of a system. For Matthias, a responsibility gap occurs when a technological system is designed to adapt its behaviour to its environment, thus rendering its operation not fully predictable (Matthias 2004). Sparrow's responsibility gap is more specific. For him, responsibility is related to autonomy—roughly following the *Grundlegung*, consisting in acting for reasons and the self-choosing of ends—at least according to the following conditional. If an agent acts autonomously, it is not possible to hold anyone else responsible for their actions. Alongside this is a more subtle point. Autonomy is a property that comes in degrees. At low levels of autonomy, the creator or 'minder' of such a quasi-agent is responsible for its interactions. At full autonomy, only that agent is. But there is a 'grey area' between the two, where there is some degree of autonomy (65-6). In this, the agent is acting for reasons, not with the competence sufficient for full autonomy, but with sufficient competence to absolve the minder of responsibility. Most teenagers are examples; so are many who suffer from mental illness or cognitive limitations. Sparrow's view is that the responsibility gap exists; the laws of war require responsibility for killings; so compliance with the laws of war requires that one not fight in a way that will result routinely in 'gappy' killings. Call this the *near-autonomy responsibility gap*. One of the things wrong with using child soldiers is that they are in this grey area of non-responsibility (73-4). Killer robots would be too.

3. Responsibility for nearly-autonomous robots' killings

We deny that the near-autonomy responsibility gap could apply to killer robots. Examining the analogy with children makes the point.

It is very plausible that the near-autonomy responsibility gap exists. Further, it applies for some killings. Take the tragic 1993 murder of the British toddler, James Bulger. During the briefest moment while his mother's attention was elsewhere while shopping, Bulger was abducted and killed by Robert Thompson and Jon Venables, both then aged 10. Presume that Thompson's and Venables' parents could not reasonably have foreseen that their boys might act in such a way. No-one is fully morally responsible for Bulger's death. It occurred in the near-responsibility gap.

Now consider Sparrow's proposed parallel, child soldiers. Children have been widely abducted and armed as soldiers by the so-called 'Lord's Resistance Army' over the last 25 years, operating variously in northern Uganda, the D. R. Congo and south Sudan. We propose that there is no ambiguity about responsibility for the kill-

ings that these children have conducted. Joseph Kony, the leader of the LRA, is responsible. So are those of his lieutenants who are of sufficient age for moral responsibility, and who have actively taken part or been complicit in the forcible manipulation and arming of those children. In putting weapons in their hands, subjecting them to extreme psychological duress, and then leading them in actions where unjust killings were the probable outcome, Kony and confederates are responsible for the resulting deaths.

The LRA poses the issue in stark terms because there is no moral justification for the rape, torture and killing its members have conducted. Not only were unjust killings the probable result of the LRA's 'military' activities, they were also its purpose. But we doubt that this feature of the case results in misleading intuitions. It is negligence to entrust to children the tools for a task which is itself permissible, when the tools require technical and moral competence for their operation. The negligent are responsible for the outcome. The fact that children may *nearly* possess sufficient technical and moral competence to perform the task does not absolve from responsibility those adults who delegate the task to them. Take car driving. People need opportunities to learn how to drive. But it is negligence to give a learner the keys to a car and then not to accompany them or to fail to guide them as appropriate when they are behind the wheel. A test then establishes that someone is of sufficient competence to themselves be responsible. The near but not full competence of a learner does not create a responsibility gap.

This applies also to war. The near but not full competence of a child-soldier does not create a responsibility gap. Weapons should not be entrusted to children because they lack the technical and moral competence to use them appropriately. (There is a plethora of other reasons why they should not, of course.) Those who do so become responsible for any misuse. In a military that employs child-soldiers, most likely a great many people are responsible for their killings: from the top of the military's chain of command down. In a well-governed state, responsibility also extends upwards to the political level. Those people who arm, train, indoctrinate, and lead or allow child-soldiers out on operations, know full well what the likely consequences are and in result are appropriately held accountable for the results.

Mutatis mutandis for nearly-autonomous killer robots. The near-autonomy of killer robots does nothing to absolve those who deploy them from responsibility for such robots' unjust killings.[1] Finding whom to pin the blame on will vary on each to

[1] Caveat: while controversial, it is nonetheless possible that there is a moral division of labour, in which soldiers low in the chain of command are justified, not merely excused, in following orders that are themselves unjust. Yitzhak Benbaji (2009) gives a contractarian argument for the justifiability of moral labour-shifting for the *ad bellum* conditions. It is no great extension of his argument to include some tactical practices which are unjust by the *in bello* conditions, but

occasion, but will usually include senior military figures and politicians. Our present point is that the near-autonomy or otherwise of an agent who performs an unjust killing is not relevant to the exculpation of higher authorities who prepare and commission that agent to kill.

4. Engineering tolerance

While some nearly-autonomous agents threaten a responsibility gap, killer robots do not. But there is a related gap that they *do* create, even in war: a 'blameworthiness' gap. This is overlooked because the schema Sparrow uses to allocate responsibility is incomplete. This section identifies that gap for engineered systems generally. The next section applies it to systems designed to kill.

Consider a bridge. If it collapses, there is a threat to life. Responsibility for minimising the threat to life is allocated in the following way. The engineer is responsible for designing the bridge in a way that is sufficiently robust, and then for building it to that standard: using only those materials with the appropriate properties, carrying out appropriate testing, using competent personnel in building it, and so on. In well-governed societies, a public official inspects the design and construction of the bridge. In implementing a regulatory regime, the official is charged with ensuring conformity to the legally required standards. Road users are then responsible for using the bridge within any publicly declared restrictions.

'Sufficient robustness' for a system is the degree of reliability it must meet. The degree of reliability required is the result of a judgment regarding how the costs of system design and construction should be balanced with the value of the system, the expected costs of its failure, and the distribution of those costs. The distribution of those costs is a factor because the risks of a system's failure do not always fall on those who benefit from the system. Adapting an engineering term, call this degree of required reliability the *tolerance level*. In practice, the tolerance level is usually given by a specification of the conditions under which a system should perform its function. The costs of designing and constructing it to that standard are then borne accordingly, with willingness to pay revealing that a system built to those standards is regarded as valuable overall. For a bridge, the tolerance level specifies at least the environmental conditions under which it should work, such as the temperature range (e.g. -20º to +50º C), for how long (e.g. 100 years), under what earthquake shocks (e.g. 5.0 on the Richter scale), and with what loads (e.g. vehicles not heavier than 5tons).

---

which are ordered by legitimate authority. Were it impermissible to deploy killer robots—contrary to our view—this caveat could justify soldiers who were appropriately ordered to launch them.

The tolerance level is a normative concept. It incorporates instrumental and moral considerations to determine how reliable a system should be. The normativity of the tolerance level may be legal or moral, according to the source of demand for a particular degree of reliability. Its requirements are realised or not by engineers in the systems and structures they build. Specifically, setting the tolerance level requires resolving a series of nice questions that are addressed by philosophers under the rubric of the 'ethics of risk'. In particular, it is an open question what justifies the imposition of risk on others, especially when they do not consent to the activity that creates the risk, and individually bear no *ex ante* benefit from it, such that its expected value is negative for them. But the impermissibility of *any* imposition of risk would prohibit a lot of mundane activities in modern life, leading to the 'problem of paralysis'. (For overviews of the debate, see Hayenhjelm & Wolff 2012; Hansson 2013. We return to the ethics of risk in §6 below.)

Now imagine the bridge collapses. Any of these actors could be morally responsible and legally liable. Consider the former only. If a 7ton truck is driven over a bridge declared publicly to be authorised for use by vehicles of 5tons or less only, then the driver is responsible for the collapse. If the engineer has failed to design and build the bridge to the tolerance level, for instance by using too much sand in the concrete, then they are responsible; in such a case, the inspecting official is likely responsible too. Equally, it may be that the engineer is exempt responsibility. It is at least plausible that if they have designed and constructed the bridge to the legally required tolerance level, and are permissibly ignorant that the regulatory regime's requirements are not as stringent as morality requires, then moral responsibility for any loss is public.

It is possible that no-one is responsible, however. Suppose there has been rainfall of a level reasonably expected only once in 300 years, resulting in a force of water pressing against the piers beyond the bridge's tolerance. No-one is responsible for the natural event. Building a bridge of sufficient robustness to withstand the water would have been a misallocation of resources that could be better used elsewhere. Any deaths that result from its collapse, though tragic, are no-one's fault. This is in the same way that, in the normal course of events, no-one is responsible when a death occurs due to a lightning strike or hungry crocodile. For deaths that result from conditions *inside* a system's required tolerance level, one at least of the user, engineer or regulator is responsible. But for deaths *outside* of a system's required tolerance, it is possible that no-one is responsible. Engineered systems create the possibility of another responsibility gap, whereby there is a risk of harm or death for which no-one is responsible, when a system functions in conditions outside of its required tolerance.

For some technologies, deciding whether it has performed within its tolerance level cannot be done on a single-case basis. Take drugs. Andy's hallucinating from taking the anti-malarial Mefloquine does not show one way or the other whether it is performing within tolerance. It is only if the aggregate side-effects of a popula-

tion's taking Mefloquine are known, as well as the degree of protection provided against malaria across that population, that a judgment can be made as to whether the drug is performing within tolerance. Suppose that no more than one in 1,000 patients ought to suffer hallucinations as a side-effect from Mefloquine. Whether Andy's hallucinations are inside or outside the required tolerance, depends on facts about how many other people have taken Mefloquine and suffered hallucinations. A need for aggregated judgment arises when the possible disutility of an engineered system cannot be inferred from failure on individual occasions of use, but can be identified in the long run probabilistically. One of the important functions that public regulation plays is of adequate testing of new technologies where users cannot consent appropriately to those risks without help, due to unavoidable ignorance, or where the benefit is gained only by imposing risk without consent; of consideration of whether the system is overall valuable; and then of licensing it for use. Licensing of these systems is of types (compounds; models; etc), whereas licensing of large-scale civil engineering projects, for instance, is of tokens. Performing this regulatory function is a moral responsibility for public authorities.

The Mefloquine case shows that the following conditional is false: if a system performs outside its tolerance level, no-one is responsible for resultant harms. Those who own and sell the drug—call them *PharmaCo*—are responsible. This is because PharmaCo own and sell a product that they know can cause serious harm. However, the case is compatible with the following conditional, which we endorse: if a system performs outside its tolerance level, no-one is blameworthy for resultant harms. (Anyone responsible for creating conditions that exceed the tolerance level is excluded from consideration.) The conditional is true because you cannot be blamed for not doing what you are not required to do, and the tolerance level just is the degree of reliability that is morally required. The revised conditional allows that PharmaCo may owe compensation for harm that results from Mefloquine. But PharmaCo may still be justified in selling the drug, despite the risk, because the product is overall valuable.

Tolerance levels are sometimes strict. That is, absolute reliability is required. This is usually because the consequences of system failure are too serious to be permitted. Recall the distinction between the sources of the reasons that determine the tolerance level. A tolerance level may be strict in virtue of the demands of either morality or law. When the tolerance level is morally strict, and absent excusing conditions, then someone is blameworthy for every failure of a system. If the tolerance level is legally strict, then someone is legally liable for every failure. This is compatible with that person's not being morally responsible, and because responsibility is a condition of blameworthiness, so there is no inference from legal liability to blame. (A legally-strict tolerance level is thus a form of strict liability, as found in Anglo-American tort law.) Strictness is one end of a spectrum of the degree of reliability required; non-strict

tolerance levels are of varying degrees of stringency. Because resources are finite and the world sometimes harsh, no system can be designed and constructed so that failure is impossible. When the tolerance level is strict, the owners of such a system carry a second-order level of risk, according to the source of the strictness—the risk may be legal, financial, moral, or combination thereof. The possible costs may be so serious that they outweigh any value in the project.

5. Assigning responsibility and blame for robots' killings

In this section we apply the foregoing schema. Robots are engineered systems. They are, and will be, expected to perform their function with a sufficient degree of reliability, that degree being defined for each task. When a robot is designed for war, one of the requirements of the tolerance level is to define what degree of reliability the robot ought to achieve in attacking appropriate targets only. Morality's predominant concern here is with the possibility of inappropriate targeting leading to that gravest of wrongs: of people being unjustly killed.

When is someone unjustly killed? Someone is morally liable to attack only if they have forfeited the right that they not be killed. (We omit the 'moral' qualifier henceforth.) On the usual account, this right is forfeited in civilian life when someone poses an unjust threat, and force is then permissibly used in self- or other-defence. If someone is liable to attack, then their rights are not transgressed when they are attacked. It is highly likely that there are situations in which people may permissibly act in a way that leads to rights being transgressed. This happens if it is permissible to turn the runaway trolley, leading to the death of an innocent worker on the side-track rather than five innocent workers further down the main line. Using terminology introduced by Judith Thomson, say that in this case Y *infringes* X's right, but does not *violate* it. He violates her right 'only if it is not merely true that Y let [a state of affairs which X has a claim to] fail to obtain but more, that Y ought not have let [the state of affairs] fail to obtain' (1990: 122). In answer to our above question, then, apply the distinction to war as follows. Someone is killed in war unjustly only when their right not to be attacked is violated. If they are liable to attack, or that right is merely infringed, then the killing satisfies the demands of justice at war.[2]

It is controversial what the criteria are for liability to attack in war. The inherited legal view is that someone is liable for attack if and only if they are a combatant, and regardless of which side they fight for. 'Revisionist' Just War theorists assign liability very differently, according to the degree of threat posed, causal contribution to that threat, and the justice of that threat. Centrally, they propose that soldiers who

[2] On Thomson's official formulation, violations are a species of infringement. We omit the 'mere' qualification from non-violating infringements: on our use, an infringement is not a violation.

fight for a cause that satisfies the conditions of *jus ad bellum* are not liable to attack.[3] We are neutral on the correct account of liability. For simplicity of exposition, we use the conventional combatant/non-combatant distinction to mark liability to attack. Appropriate changes can be readily made according to one's criteria.

Given the forgoing, responsibility for robots' killings in war would be distributed as follows. One of the functions of killer robots' tolerance level would be to specify that degree of reliability robots should achieve in killing only combatants. (There is *pro tanto* reason to set the legally required tolerance level at not less than that required by morality.) *Ex hypothesi*, combatants have no complaint against being killed. But non-combatants do. As killer robots may fail to respect their claim not to be attacked, non-combatants would be exposed to lethal risk, and non-consensually so. So public authorities would have a duty to regulate killer robots' design and manufacture. Absent excusing conditions, failure to regulate would render public authorities blameworthy for robots' killings inside the tolerance level. Engineers would be responsible for designing and building killer robots to the tolerance level, and failure to do so would render them blameworthy for unjust killings inside it. Military commanders would be responsible for using killer robots only within the specified parameters of use; these parameters in turn must be within those used to set the tolerance level. If soldiers deploy them outside the set parameters, they are blameworthy for ensuing killings.

It is conceptually possible, however, that there exists a set of killings by robots of non-combatants, for which no-one is blameworthy: those that occur *outside* the tolerance level. If the tolerance level is not strict, then it could be permissible to deploy killer robots which would attack, say, one non-combatant for every 100 combatants attacked. No-one would be blameworthy for the resulting deaths. Even though such killings would be both tragic and a transgression of the rights of those non-combatants, it is conceptually possible that they could be infringements, and not violations.

Our gap relates to Sparrow's as follows. For clarity, call ours the *permissible malfunction* blameworthiness gap. The gaps are commensurable, for Sparrow's near-autonomy responsibility gap entails an extensionally equivalent blameworthiness gap. They hold at differing levels of autonomy. Recall that full autonomy, for Sparrow, consists in the self-choosing of ends and of acting for reasons. The near-autonomy gap arises when an agent possesses *some* ability to choose his own ends and to evaluate reasons, but lacks the degree of self-ownership and of facility in evaluating reasons sufficient for him to be held morally responsible (e.g. children). The permissible mal-

---

[3] This presumes that just combatants comply with other demands of morality on the use of force, such as proportionality. McMahan (2009) is the revisionist *locus classicus*, against Walzer (2006). See Fabre (2012: 118-28) and Frowe (2014) for striking claims about who is liable, and Rodin (2014) for who is not.

function gap holds only at a lower level of autonomy (e.g. automated elevators). What marks the distinction between the relevant levels of autonomy? We conjecture that it consists in the ability of a system to decide on its own ends. At a lower level of autonomy, although a system may perform computations that functionally equate to the evaluation of reasons, the ends are non-revisable and have been set by the programmers. Our proposal seeks to identify the point at which the engineers or operators of a system lose the control or influence on its behaviour that is required to hold them responsible for its actions, and thus blameworthy, in the same way that parents at some point cease to be responsible for the actions of their children. For the sake of our argument, however, the soundness of this conjecture is inessential. What matters is that, on the spectrum of possible degrees of autonomy that a robot might possess, some could pose the permissible malfunction blameworthiness gap but not the near-autonomy gap.

5. Morality's demands on the tolerance level

What are the implications of the forgoing? We have presumed that the schema by which responsibility and blame for the proper functioning of engineered systems is distributed in war is the same as in civilian life. When a system is designed to kill, the conceptual possibility then exists that some resulting deaths of those not liable to be killed would be infringements but not violations of victims' rights. This is so just if the tolerance level is not strict. The pertinent question is then the following: does morality require that the tolerance level be strict? If strict, someone would be blameworthy for any killing of non-combatants by a robot. If not, some killings of non-combatants by robots at war would be morally permissible. No-one would be blameworthy.

We endorse the latter view, but first set out the argument for the strictness of the tolerance level before replying. It goes as follows. Killer robots would impose the risk of the gravest of harms on non-combatants, namely non-liable death. The risk would be new. They would not have consented to it. No compensation can plausibly outweigh the non-consensual, non-liable loss of one's life in war. Non-combatants derive no *ex ante* benefit from killer robots' use. And crucially, the re-distribution of risk imposed by the use of killer robots would be unfair. By sending robots to war in place of people, country *B*'s combatants are immunised from the risk of death or injury in combat. But country *B* has done so by 'offloading' risk onto country *A*'s non-combatants, and this is impermissible. (This is impermissible even if *B*'s combatants are not liable to attack, and so does not depend on the status of the doctrine the moral equality of combatants.) In such cases, tolerance levels should be strict. The earlier bridge case is misleading, for it differs on at least three of those features: in using the bridge, travellers would have consented to the (small) risk of collapse, have expected *ex ante* to benefit from its use, and done so as part of a suite of risks commonly undertak-

en in travelling. An inexact but closer analogy to killer robots would be the arrival of cars on the roads, at the point in time when the technology was uncommon. Imagine a *Wind in the Willows* world where only Toads have cars, and Rats, Badgers and Moles are always pedestrians. In this, the latter have not consented to the risk of being run over; they derive no *ex ante* benefit from Toads' driving; and the imposed risk is new. In that world, *someone* is blameworthy every time a pedestrian is run over. Usually, it is a Toad.

To make the objection more precise, consider what James Lenman calls the 'Precaution Thesis'. Suppose a population of 20 million people each faces a 1 in 500,000 baseline risk of dying, in the absence of intervention, resulting in 40 expected deaths. He then contrasts the following interventions:

> *Policy E.* Reduce the risk to each of the 20 million to 1 in 1 million.
>
> *Policy F.* Reduce the risk to 19 million of the population to 1 in 19 million while the risk to the remaining 1 million is increased to 1 in 100,000.

Policy F is preferable on simple utilitarian grounds; it results in 11 expected deaths, while E results in 20. Yet absent some unstated countervailing considerations, it is morally objectionable. The risk to some is increased so that the risk to many might be diminished. Madeleine Hayenhjelm calls this *the aggregation worry* (2012: 918). The issue here is the same as that central and more general objection to utilitarianism, that its aggregative weighting of welfare ignores 'the separateness of persons'. Others' welfare does not justify my victimisation (Rawls 1999: 26-7; Nozick 1974: 33; Scanlon 1982: 123). Motivated by the contrasting policies E and F, and adopting Scanlon's contractualist framework, Lenman proposes that it is not sufficient justification for the imposition of risk on a population that one seek to minimise that risk by taking precautions *simpliciter*. Rather, we must 'seek to minimize the risk of harm to *each* of them, aiming any precautions we take at the safety of *each* affected person' (2008: 106). The Precaution Thesis is then the conjunction of two subordinate necessity claims. It requires, first, that those who impose risk on others take all reasonable precautions to avoid their coming to harm. Further, second, it requires that they do so in a way that is justifiable according to the reasons that *each individual* has. The Thesis is not endorsable only by contractualists. Non-aggregative moral theory more generally—the broad assumption of this paper—should do so.

The objection to a non-strict tolerance level for killer robots can now be stated. Non-combatants cannot reasonably be expected to consent, actually or hypothetically, to a risk of death which is new and from which they derive no expected benefit. Assume, further, that only one party to the war has killer robots, thus precluding a reciprocity of risk which might be thought to justify their use (Altham 1983: 26; Hans-

son 2013: 101-4). The same *pro tanto* unjust distribution of risk that marks out policy F would also be imposed by the deployment of killer robots: a group is non-consensually, unfairly picked out for exposure to greater risk for the sake of others' welfare. This cannot be justified in terms of those individuals' reasons. Allowing any non-strictness in killer robots' tolerance levels would be to permit an unreasonable lack of precaution regarding country *A*'s non-combatants' lives. It could not be justified *to them* specifically, in terms of the reasons that they have. The tolerance level for killer robots must be strict, on pain of failure to satisfy the Precaution Thesis.

The objection is unsound. Although there is a sense in which the risk posed by killer robots is new, it is false that there is no *ex ante* benefit for country *A*'s non-combatants. The risk baseline for country *A*'s non-combatants prior to killer robots' deployment by country *B* is *not* the routine risk level of everyday life. Rather, it is the risk of unjust death occurring as a result of operations prosecuted by an all-human combat force. Call this latter level of risk r_1. Call the same non-combatants' risk of unjust death occurring as a result of operations by a combat force that includes killer robots, r_2. If $r_2 < r_1$, then Country *A*'s non-combatants enjoy *ex ante* a risk reduction. Here is a case to illustrate the structure of re-distribution that would actually occur. Suppose that, in the natural course of events, villagers in a valley are exposed to the non-trivial risk of death by flooding. To mitigate this risk, the council then builds a dam upstream to hold excess water. There is a sense in which the villagers face a new risk, namely that of the dam's collapse leading to lethal flooding. But so long as there is less risk of drowning as a result of the dam's collapsing than due to natural flooding events, there is *ex ante* benefit for the villagers in the kind of risk that matters. *Mutatis mutandis* for killer robots.

Reducing the level of risk is not yet decisive. There is also a question of equality in the distribution of risk, as Policy F brings out. Yet it is an extreme view indeed that takes the equal distribution of risk always to override the level of risk in policy preference. Further, the case at hand is not one that falls foul of plausible views about permissible risk distribution. For if $r_2 < r_1$, the deployment of killer robots is structurally akin not to policy F, but to the following policy G:

> *Policy G.* Eliminate the risk to 10 million of the population, while reducing the risk to the remaining 10 million to 1 in 1 million.

The no-risk pool is analogous to country *B*'s combatants, and the risky pool to *A*'s non-combatants. When compared with policy E, it is far from obvious that policy G is objectionable. For in G, *everyone* enjoys a reduction in risk; it is just that the reduction is not distributed equally. Indeed, G seems obviously welcome. Of course, there is a plausible egalitarian case that it would be unjust if the alternative policy H was available:

Policy H. Reduce the risk to each of the population to 1 in 2 million.

While both policies G and H result in 10 expected deaths, the egalitarian presumption that an equal distribution of risk is a fair one favours H. Policy H is justifiable to all too, not just to those who benefit from being in the low-risk pool. So, when it is available, the Precaution Thesis mandates policy H. In the killer robots' case, however, *policy H is not available*. The technology poses an unavoidable asymmetry, with the non-combatants of country A exposed to some risk while it is eliminated for country *B*'s soldiers. The choice for the remaining 10 million is thus not between r_2 with the overall risk distributed according to policy G, or r_3 according to H, and where $r_2 > r_3$. It is between r_1 with no technological intervention and distributed as risks in war currently are, or r_2 distributed in the same way as policy G, where $r_1 > r_2$. Why would country *A*'s non-combatants *not* reasonably agree to the latter? Because the distributional structure of policy G is defensible to everyone concerned, given the unavailability of an intervention akin to policy H, G *does* satisfy the Precaution Thesis. Nor need the tolerance level be strict in that scenario. So long as $r_2 < r_1$ for country *A*'s non-combatants, the policy is defensible. Clearly, they will want the differential between r_2 and r_1 to be as great as is reasonably possible, and the Thesis requires this. But reasonable failure to achieve no targeting malfunctions does not render the advent of killer robots undesirable. Contrary to the supposed objection, then, the Precaution Thesis *is* satisfied by a non-strict tolerance level.

With the distinction between the level and the distribution of risk in hand, it becomes clear that the issue here is an analogue in terms of risk of a familiar political debate: that of how to weight the often competing claims of the absolute value of goods distributed and of equality in their distribution. It is a strong egalitarian view indeed that accords lexical priority to the equal distribution of risk over the level of risk. In Derek Parfit's terms, only a 'pure Telic' egalitarianism would do so, one which takes inequality in itself to be bad and its badness to be the only or overriding moral consideration (Parfit 1997). Rather than being vulnerable to the 'levelling down' objection for goods, pure Telic egalitarianism about risk faces a levelling up objection: one should *increase* the level of risk faced by everyone to the level faced by the expectedly worst-off. Less dramatically—but pertinently to the case in hand—one should not decrease the level of risk faced by some, if one cannot do so for all. This last stricture is equally horrendous; Oskar Schindler-style rescues would be ruled out, for instance. For more plausible views about how to weight equality in distribution with the value of goods distributed, such as the prioritarianism of Rawls' difference principle, absolute improvements for the worst-off will be welcome, even if distribution is more unequal. Apply these views to risk, and to the case at hand. The absolute reduction in levels of risk for country *A*'s non-combatants that killer robots could bring about is

welcome. This applies even if the result is a more unequal distribution of risk, due to the elimination of risk for country *B*'s people.

So morality makes the following demands on the tolerance level for killer robots. The tolerance level is set by whichever is the more stringent of the following two conditions. First, the risks that killer robots pose to country *A*'s non-combatants must be less than that posed by all-human armies: it must be the case that $r_2 < r_1$. If $r_2 > r_1$, the introduction of the technology is an impermissible imposition of risk, and such killer robots ought not be deployed in war. (We are here committed to the falsity of what Jeff McMahan has termed the 'priority of combatants' thesis, viz. that Country *A*'s combatants are justified in pursuing tactics which increase the risk to Country *B*'s non-combatants, in order to decrease their risk. This commitment *raises* the bar for us. If the priority of combatants thesis is correct, then killer robots could be permissibly deployed even for some cases where $r_2 > r_1$.)[4] It is a deep and difficult question whether it is likely that r_2 will be greater or lesser than r_1. Most of the contributions to the policy debate by scientists and technologists consist in justifications (or assertions) of their best estimates regarding that relation (see, e.g., Arkin 2010, who claims in effect that it will likely be the case that $r_2 < r_1$; and Sharkey 2012a, 2012b, who argues the contrary). The second condition is this: the tolerance level requires that r_2 should be as low as is technologically feasible. Anything less would be a failure to take reasonable precautions. If it is possible to build robots that comply with a strict tolerance level, then that is obligatory. But if it is not possible, the best we can do is good enough.[5]

It is noteworthy that remotely-piloted weapons systems impose the same possible structures of redistribution of risk that killer robots would. Discussion of risk in the context of drones has had two foci hitherto. Debate has focused principally on the charge that justice requires reciprocity of risk in war. According to Henry Shue, one of the reasons for marking liability to attack according to the combatant/non-combatant distinction is that 'it allows for a "fair fight" by means of protecting the utterly defenceless from assault' (2004: 51, cited by Steinhoff 2006: 336). Paul Kahn argues that the combatants possess the right to injure each other 'just as long as they stand in a relationship of mutual risk' (Kahn 2002: 3). We have not addressed this view because there are convincing replies, most especially by Uwe Steinhoff (2006) and B. J. Strawser (2010). A second focus addresses whether drones result in disproportionately high levels of non-combatant casualties, and are therefore unjust, given

---

[4] See Kasher & Yadlin (2005) for its proposal, and Margalit & Walzer (2009), McMahan (2011), Luban (2014) and Zohar (2014) for rejection.

[5] This moral argument gives only *pro tanto* justification for a non-strict *legal* tolerance level. The incentivising effect of requiring legally a strict tolerance level may do much to reduce unjust deaths in war.

less destructive alternatives. Proponents of drone use argue that: in principle, they 'allow for an even higher level of fidelity to each of these moral and legal requirements [of proportionality, discrimination and noncombatant immunity]' (Johnson 2013: 177; also Strawser 2010 351-2); and in fact, drones have achieved this (Plaw 2013; contested by Braun & Brunstetter 2013, Kreps & Kaag 2012). Both points are in support of the thesis that the level of risk posed by drones, $r_{2(*)}$, is less than r_1. We have provided the argument for why this parallel condition must be met.

It is feasible that some killer robots will satisfy our conditions. This is most obvious for tactical environments in which non-combatants rarely go or are easily identified: air-to-air combat, underwater, in space, and in nearly uninhabited regions of land. Pertinently, killer robots will also be useful in tactical situations more akin to those that predominate in today's asymmetric and undeclared conflicts. Consider possible scenarios from two conflicts on-going at the time of writing: undeclared Russian forces operating in Eastern Ukraine, and 'Islamic State' in Syria and northern Iraq. It is feasible that, in the coming decade, killer robots will possess algorithms that accurately discriminate armoured vehicles and artillery pieces from soft-skinned vehicles; both tend to have highly distinctive heat and shape signatures. Such systems patrolling the border of Eastern Ukraine able to interdict infiltrating Russian armour and artillery would likely satisfy the moral requirements of the tolerance level. Similar possibilities could apply in Syria. There the vehicle of choice is a pick-up with heavy machine-gun or anti-aircraft gun mounted on the rear. Killer robots patrolling defined areas where IS are known to operate could likewise identify and attack only these. No doubt in each scenario combatants would learn quickly what camouflage or tactics worked against the new threat. No doubt vicious combatants would use human shields to deter attack or create outrage, and requiring killer robots' targeting software to incorporate a proportionality calculation. Neither of these observations undermine the claim that killer robots will likely satisfy morality's demands, as well as have real military utility.

7. Robots and respect

Two possible objections to our approach are important, which we develop and reply to in this penultimate section. It may be charged, first, that we have ignored the phenomenology. The prospect of being killed by a robot, rather than merely being killed, holds a distinct horror; the act feels different in a way that sets it apart from conventional ways of dying at the hands of people. While the distribution of risk is a legitimate moral concern, it hardly expresses the gravity of what is really at stake. Combatants may have no legitimate complaint against being attacked as such, but they have a complaint against being attacked *by a robot*.

'Tax averse' preferences provide a parallel. Some people would rather pay a higher monetary price for a product than pay less net, and have some of that go to the

government as tax (Sussman & Olivola 2011). Perhaps something similar is true of how one dies. Barbara Ehrenreich speculates in *Blood Rites* that heroic ideals central to the culture of war arose in evolutionary pre-history, in which our most urgent fear was of being eaten by animals (1997). Killer robots raise the spectre of a new form of 'inter-species' predation, in which we are again the prey. While many are willing to risk death in combat for any number of reasons, perhaps the prospect of being killed by a robot has an emotive quality that we are especially aversive too. The aversion may be sufficiently powerful that many would prefer a higher risk of death at human hands than a lower risk at an android's cannon, even though the physical pain would be qualitatively indistinguishable. The atavistic horror that the idea of being hunted by a robot gives rise to may well be what is expressed by the imprecise but frequently heard charge that it would be 'inhumane' to authorise robots to identify people as targets and attack them autonomously. No matter what utility there is in having such machines, for whatever kind of person-person quarrels we have, people should not die at the hands of a robot. It would be a kind of species disloyalty.

The aversion could be taken to have two possible implications. If sufficiently powerful and widespread, it may be in our interests collectively to forgo the technology. Similarly, the aversion may be such that the sense of injustice that would result from robots' unjust killings would be far greater than the same deaths at human hands, and which would permit neither justification nor excuse. If such a judgment was sufficiently widely shared, the tolerance level should be strict.

This is not implausible. But it is not, in our view, a judgment that will be sustained in the long term. The reason is not merely a clash of intuitions. Rather, there has been a repeated pattern historically at the arrival of a new technology. Its prospect is greeted with utopian and more typically dystopian visions of its social significance. As pragmatic reasons for its adoption become compelling, it becomes widely present, so that people have regular first-hand experience of how it works. On doing so, the technology loses its aura of moral singularity and becomes part of the fabric of life. Recall the arrival of self-checkout terminals in supermarkets, which have largely automated the payment process. Initial reluctance to use them, combined with clunky technology, has been relatively swiftly overcome through familiarity. It is not knowable whether killer robots will prove discontinuous with this pattern; perhaps their restricted contexts of use will prevent them ever being normalised. But *prima facie*, we expect them to. As the presence of robots in our lives generally becomes routine, there is reason to suppose that any current 'robot aversion' will dissipate in future. Nor can we see a non-question-begging reason to object morally to their normalisation.

The normalisation of robots in society also answers a second possible objection. Although the responsibility trilemma forms the heart of Sparrow's (2007) argument, his article contains a pregnant but undeveloped suggestion, originally proposed in justification of the trilemma but containing wider implications. He argues that kill-

ings in war must satisfy the demands of interpersonal respect; this generates the requirement to attribute responsibility. The point is owed to Thomas Nagel, from an influential early discussion. '[W]hatever one does to another person intentionally must be aimed at him as a subject, with the intention that he receive it as a subject. It should manifest an attitude to *him* rather than just to the situation, and he should be able to recognize it and identify himself as its object' (Nagel 1972: 136; Sparrow 2007: 76, fn. 30; Sparrow 2011: 124-5 develops the connection). Nagel applies his point principally to who may be targeted, if the demands of respect are to be satisfied. Sparrow supplements Nagel's with a parallel claim: the technology of killing must also satisfy the demands of respect. If killer robots fail to satisfy the requirement of respect, the implication is more serious than that there should be a strict tolerance level. Killer robots ought not to be used at all.

The likely normalisation of robots in society replies to this objection too. As robotic technology becomes more commonplace, it will no longer be a sign of disrespect to delegate some kinds of interaction to a machine rather than a person. Indeed, Nagel has anticipated our reply. In considering why it is appropriate to punch someone in the mouth who has insulted you, he remarks that 'in our culture', it is an insult to punch someone in the mouth, and not merely an injury. This reveals 'a perfectly unobjectionable sense in which convention may play a part in determining exactly what falls under an absolutist restriction [of use in war] and what does not' (1972: 135, fn. 7). Convention may play a large role in determining what constitutes offense. Take artillery, a paradigm example of a weapons technology used for indirect mass killing in war. Only a romantic longing for an idealised medieval code of chivalry supposes that it is ethically unacceptable on the contemporary battlefield. Familiarity with the weapon has rendered it no less disrespectful as a tool for killing than face-to-face combat, regardless of the temptation to have labelled it as such when introduced. There is no principled reason why, with familiarity, the same would not be true for killer robots.

6. The obligation to regulate

In conclusion, it should be noted what we have not argued. We have defended two necessary conditions for the just use of killer robots. Their fulfilment does not result in an all-things-considered judgment of their permissibility. Like drones, perhaps a technology that puts such great power in the hands of its owners makes the resort to war too likely. This 'threshold objection' is serious (see Brunstetter & Braun 2011. Steinhoff's 2013 argument that extreme asymmetries of risk justify terrorism is a version of this objection; for counter-considerations, see Beauchamp & Savulescu 2013). The same concentration of power raises a democratic worry, that military action may lack the consent of the people generally, in a way that is harder with mass mobilisation.

The same concentration of power also undermines a revisionist aim, of encouraging soldiers to refuse to fight in unjust wars, thereby reducing their incidence (McMahan 2013). Perhaps killer robots will impose excessive and widespread psychological damage, rendering them disproportionate tools of war (cf. IHRCRC/GJC 2012: 80-8). There are other objections too. All this is to show that the distribution of responsibility for and risk of robots' killings are not the only moral considerations.

We close with a practical implication of our argument. Governments in well-ordered societies have a responsibility to regulate emerging robotic technologies. Those who will face the risks of malfunctioning killer robots will not have consented. There will be a moral obligation to determine the appropriate tolerance level for killer robots, and to licence them for use only after a thorough testing and inspection process. International organisations likewise have a moral mandate to ensure that badly-ordered societies are nonetheless accountable for not deploying killer robots in ways that contravene the laws of war. Our position is well summarised in the following slogan. Regulate; don't ban.[6]

THOMAS W. SIMPSON
Blavatnik School of Government
University of Oxford, UK

VINCENT C. MÜLLER
Anatolia College/ACT
Thessaloniki, Greece

7. References

Altham, J. (1983) 'The Ethics of Risk', *Proceedings of the Aristotelian Society*, n.s. 84: 15-29.

Arkin, R. (2009) *Governing Lethal Behaviour in Autonomous Robots*. Boca Raton, FL: CRC Press.

Arkin, R. (2010) 'The Case for Ethical Autonomy in Unmanned Systems', *Journal of Military Ethics* 9/4: 332-41.

Beauchamp, Z., and J. Savulescu (2013) 'Robot Guardians: Teleoperated Combat Vehicles in Humanitarian Military Intervention', in B. J. Strawser (ed.) *Killing by Remote Control: The Ethics of an Unmanned Military*, 106-25. Oxford: OUP.

Benbaji, Y. (2009) 'The War Convention and the Moral Division of Labour', *Philosophical Quarterly*, 59/237: 593-617.

[6] We are grateful for comments and criticism to Ronald Arkin, Janina Dill, Cécile Fabre, Alex Leveringhaus, Jeff McMahan, Marco Meyer, Andreas Mogensen, Filippo Santoni de Sio, two anonymous referees, and audiences in Aarhus, Delft, London and Oxford.

Brunstetter, D., and M. Braun (2011) 'The Implications of Drones on the Just War Tradition', Ethics and International Affairs 25/3: 337-58.

Brunstetter, D., and M. Braun (2013) 'Rethinking the Criterion for Assessing CIA-targeted Killings: Drones, Proportionality and Jus Ad Vim', Journal of Military Ethics, 12/4: 304-24.

Coady, C. (2008) Messy Morality: The Challenge of Politics. Oxford: Clarendon.

Ehrenreich, B. (1997) *Blood Rites: Origins and History of the Passions of War.* London: Virago.

Fabre, C. (2012) *Cosmopolitan War*. Oxford: OUP.

Frowe, H. (2014) 'Non-Combatant Liability in War', in H. Frowe and G. Lang (eds.) *How We Fight*, 172-88. Oxford: OUP.

Hansson, S. (2013) *The Ethics of Risk: Ethical Analysis in an Uncertain World.* London: Palgrave Macmillan.

Hayenhjelm, M. (2012) 'What is a Fair Distribution of Risk?', in S. Roeser *et. al* (eds.). *Handbook of Risk Theory*, 909-29. Dordrecht: Springer.

Hayenhjelm, M. and J. Wolff (2012) 'The Moral Problem of Risk Impositions: A Survey of the Literature', *European Journal of Philosophy*, 20: E26-E51.

Human Rights Watch. 2012. *Losing Humanity: The Case Against Killer Robots*. <http://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf > accessed 15 August 2014.

International Human Rights and Conflict Resolution Clinic, Stanford, and Global Justice Clinic, NYU School of Law (2012) *Living Under Drones: Death, Injury and Trauma to Civilians From US Drone Practices in Pakistan* <http://www.livingunderdrones.org/wp-content/uploads/2013/10/Stanford-NYU-Living-Under-Drones.pdf> accessed 20 May 2015.

Johnson, R. (2013) 'The Wizard of Oz Goes to War: Unmanned Systems in Counter-insurgency', in B. J. Strawser (ed.) *Killing by Remote Control: The Ethics of an Unmanned Military*, 154-78. Oxford: OUP.

Kahn, P. (2002) 'The Paradox of Riskless Warfare', *Philosophy and Public Policy Quarterly*, 22/3: 2-8.

Kasher, A. and A. Yadlin (2005) 'Military Ethics of Fighting Terror: An Israeli Perspective', *Journal of Military Ethics*, 4/: 3-32.

Krebs, S. and J. Kaag (2013) 'The Use of Unmanned Aerial Vehicles in Contemporary Conflict: A Legal and Ethical Analysis', *Polity* 44: 260-85.

Lenman, J. (2008) 'Contractualism and Risk Imposition', *Politics, Philosophy and Economics*, 7/1: 99-122.

Luban, D. (2014) 'Risk Taking and Force Protection', in Y. Benbaji and N. Sussman (eds.) *Reading Walzer*, 277-301. Abingdon: Routledge.

Margalit, A. and M. Walzer (2009) 'Israel: Civilians and Combatants', *New York Review of Books*, 14 May 2009.

Matthias, A. (2004) 'The responsibility gap: Ascribing responsibility for the actions of learning automata', *Ethics and Information Technology*, 6/3: 175-83.

McMahan, J. (2009) *Killing in War*. Oxford: OUP.

McMahan, J. (2011) 'The Just Distribution of Harm Between Combatants and Non-combatants', *Philosophy and Public Affairs*, 38/4: 342-79.

McMahan, J. (2013) 'Foreword', in B. J. Strawser (ed.) *Killing by Remote Control: The Ethics of an Unmanned Military*, ix-xv. Oxford: OUP.

Müller, V. C. (Forthcoming) 'Autonomous Killer Robots are Probably Good News', in E. Di Nucci and F. Santoni De Sio (eds.) *Drones and Responsibility*. Farnham: Ashgate.

Nagel, T. (1972) 'War and Massacre', *Philosophy and Public Affairs*, 1/2: 123-44.

Nozick, R. (1974) *Anarchy, State and Utopia*. Oxford: Blackwell.

Parfit, D. (1997) 'Equality and Priority', *Ratio* 10/3: 202-21.

Plaw, A. (2013) 'Counting the Dead: The Proportionality of Predation in Pakistan', in B. J. Strawser (ed.) *Killing by Remote Control: The Ethics of an Unmanned Military*, 126-53. Oxford: OUP.

Rawls, J. (1999) *A Theory of Justice*, rev. ed. Cambridge, MA: Belknap Press.

Rodin, D. (2014) 'The Myth of National Self-Defence', in C. Fabre and S. Lazar (eds.) *The Morality of Defensive War*, 69-89. Oxford: OUP.

Scanlon, T. (1982) 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds.) *Utilitarianism and Beyond*, 103-28. Cambridge: CUP.

Singer, P. (2009) *Wired for War: The Robotics Revolution and Conflict in the 21st Century*. London: Penguin.

Sharkey, N. (2012a) 'The Evitability of Autonomous Robot Warfare', *International Review of the Red Cross*, 94/886: 787-99.

Sharkey, N. (2012b) 'Autonomous Robots and the Automation of Warfare', *International Humanitarian Law Magazine*, 2: 18-19.

Shue, H. (2004) 'Torture', in S. Levinson (ed.) *Torture: A Collection*, 49-60. Oxford: OUP.

Sparrow, R. (2007) 'Killer Robots', *Journal of Applied Philosophy*, 24/1: 62-77.

Sparrow, R. (2011) 'Robotic Weapons and the Future of War', in J. Wolfendale and P. Tripodi (eds.) *New Wars and New Soldiers: Military Ethics in the Contemporary World*, 117-33. Surrey: Ashgate.

Steinhoff, U. (2006) 'Torture—The Case for Dirty Harry and Against Alan Dershowitz', *Journal of Applied Philosophy*, 23/3: 337-53.

Steinhoff, U. (2013) 'Killing Them Safely: Extreme Asymmetry and Its Discontents', in B. J. Strawser (ed.) *Killing by Remote Control: The Ethics of an Unmanned Military*, 179-207. Oxford: OUP.

Strawser, B. J. (2010) 'Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles', *Journal of Military Ethics*, 9/4: 342-68.

Sullins, J. (2010) 'RoboWarfare: can robots be more ethical than humans on the battlefield?', *Ethics and Information Technology*, 12/3: 263-275.

Sussman, A. and C. Olivola (2011) 'Axe the Tax: Taxes are disliked more than equivalent costs', *Journal of Marketing Research*, 48: 91-101.

Thomson, J. (1990) *The Realm of Rights*. Cambridge, MA: Harvard University Press.

US Department of Defense (2013) *FY2013-2038 Unmanned Systems Integrated Roadmap* < http://www.dtic.mil/get-tr-doc/pdf?AD=ADA592015> accessed 15 August 2014.

Walzer, M. (2006) *Just and Unjust Wars: A Moral Argument with Historical Illustrations,* 4th ed. New York, NY: Basic Books.

Zohar, N. (2014) 'Risking and Protecting Lives: Soldiers and Opposing Civilians', in H. Frowe and G. Lang (eds.) *How We Fight*, 155-71. Oxford: OUP.