

How Just Could a Robot War Be?

Peter M. ASARO

HUMlab & Department of Philosophy, Umeå University

Center for Cultural Analysis, Rutgers University

peterasaro@sbcglobal.net

Abstract. While modern states may never cease to wage war against one another, they have recognized moral restrictions on how they conduct those wars. These “rules of war” serve several important functions in regulating the organization and behavior of military forces, and shape political debates, negotiations, and public perception. While the world has become somewhat accustomed to the increasing technological sophistication of warfare, it now stands at the verge of a new kind of escalating technology—autonomous robotic soldiers—and with them new pressures to revise the rules of war to accommodate them. This paper will consider the fundamental issues of justice involved in the application of autonomous and semi-autonomous robots in warfare. It begins with a review of just war theory, as articulated by Michael Walzer [1], and considers how robots might fit into the general framework it provides. In so doing it considers how robots, “smart” bombs, and other autonomous technologies might challenge the principles of just war theory, and how international law might be designed to regulate them. I conclude that deep contradictions arise in the principles intended to govern warfare and our intuitions regarding the application of autonomous technologies to war fighting.

Keywords. Just war theory, robots, autonomous systems.

1. Introduction

Just war theory is a broadly accepted theoretical framework for regulating conduct in war, that has been embraced by such esteemed and influential institutions as academia, the US military establishment (including the military academies¹), and the Catholic Church. It is also compatible with, if not actually a formulation of, the principles underlying most of the international laws regulating warfare, such as the Geneva and Hague Conventions.

This paper aims to illuminate the challenges to just war theory posed by autonomous technologies. It follows Michael Walzer’s [1] articulation of the theory, which has been the most influential modern text on just war theory. While there are compelling criticisms of Walzer’s formulation (*e.g.* [2]), it is his articulation which has had the most influence on the institutions and international laws regulating war.

¹ Walzer’s book was a standard text at the West Point Military Academy for many years, though it was recently removed from the required reading list.

Before we begin, I should clarify what I mean by robots and other autonomous systems. “Autonomy” is a rather contentious concept, and its relation to material technologies adds further complications. It is thus useful to think about a continuum of autonomy along which various technologies fall depending upon their specific capabilities. Most generally, any system with the capability to sense, decide and act without human intervention has a degree of autonomy. This includes simple systems, such as a landmine that “decides” to explode when it senses pressure. Obviously, systems with only the most rudimentary forms of sensors, decision processes and actions lack various aspects of full autonomy. The landmine does not decide where it will be placed, and its physical placement largely determines the consequences of its actions, thus it has much less “autonomy” than systems with more sophisticated means of sensing, deciding and acting. If we were to consider it as a moral agent, we would not be inclined to hold it morally responsible for its actions, but rather hold responsible those who placed and armed it. It thus occupies an endpoint in the continuum of autonomy and moral responsibility.

Certain kinds of “precision” weapons, such as “smart” bombs that use global-positioning systems (GPS) and sophisticated control mechanisms to deliver them accurately to a target. The selection of a target, and the determination of its location, value and risks, is still determined by human agents who control the weapons system, however. Thus we might wish to “blame” a smart bomb, or its design, for failing to reach a designated target, but not for the selection of the target. It thus represents a point further along this continuum, and shares this position with various kinds of guided weapons and automated anti-aircraft batteries (*e.g.* Patriot missile systems, and the Phalanx gun systems), and automatic anti-ballistic missile systems (*e.g.* Star Wars/SDI), that detect and destroy sensed threats without immediate human intervention, though are dependent on responsible human decisions as to when it is appropriate to activate such an armed system.

Still more autonomous are systems which use sophisticated sensor analysis to select appropriate targets on their own and make decisions about the appropriateness of various actions in response to its situation. The emerging technologies of robotic weapons platforms incorporate some or all of these features, using image processing to identify targets, and selecting from a broad range of offensive and defensive actions in their engagement of targets. These are technological capabilities which already exist, and are beginning to be implemented in various countries. These systems tend to be designed to seek permission from human authorities before using lethal force against a target, what the US military calls the “human-in-the-loop,” but this is not a technological necessity. We can identify the choice to use deadly force against a specific target as a critical threshold along the continuum of autonomy, and one which carries a greater moral burden in the design and use of such a technology. There are, however, systems with more autonomy than this.

As robotic technologies advance, it is possible that they will acquire moral capacities that imitate or replicate human moral capacities. While some systems might merely enact pre-programmed moral rules or principles, autonomous robotic agents might be capable of formulating their own moral principles, duties, and reasons, and thus make their own moral choices in the fullest sense of moral autonomy. There are many possibilities short of replicating a fully autonomous moral subject, such as agents with moral awareness but not the freedom of choice to act upon that awareness. While

this still remains in the realm of science fiction, it is not impossible in principle that a robot could achieve autonomy in a Kantian sense, in which it takes responsibility for its actions, reasons about them morally, and identifies itself with the moral quality of its own actions. At some point along the continuum, but probably before Kantian autonomy, will arise various questions about the moral responsibilities of others towards such autonomous systems, and in particular whether these systems have moral rights.

There are many degrees of autonomy along the continuum at which specific systems might fall, so we will consider the implications of these on the interpretation and application of just war theory to various situations. I should also say something at this point about the speculative nature of this work. Many people find the idea of robotic soldiers and robotic moral agents to be somewhat fantastical, the stuff of science fiction and not something that deserves serious consideration. My arguments are meant to cover technological possibilities that do not yet exist, and even some that shall perhaps never exist, yet I believe that it is important to develop our understanding of existing technologies in light of these hypothetical possibilities. There is a very great interest in building autonomous systems of increasing complexity, and a great deal of money is being invested towards this goal. It does not seem to be an unreasonable prediction that within the next decade we will see something very much like a robotic soldier being used.² If we consider the amount of time and effort it took moral and legal theorists to come to terms with the atomic bomb, then it makes sense to start thinking about military robots now, before they appear fully formed on the battlefield. Robots may not have the same potential to reshape global politics that the atomic bomb did, though indeed they may. Still, it is not unreasonable to expect that these technologies might find their way into other security applications, such as policing civilian populations. There is thus a definite necessity, and a certain urgency to establishing the moral framework in which we might judge the various applications of such technologies, as well as the ethics of designing and building them. This examination of just war theory is a part of that broader investigation.

As my concern in this analysis is with the general capabilities of technologies, primarily their ability to act autonomously, and not with the specific technologies used, I will not spend much time discussing how these technologies might work, apart from their degree of autonomy. One might also object that there already exist legal restrictions on the use of autonomous systems in combat, and these have succeeded thus far in keeping humans-in-the-loop of the most advanced military systems being developed, at least in the US, and thus there is no need for such an analysis. While it is true that humans are being kept “in the loop,” it is not clear to what extent this is only a contingent truth, and whether this restriction can resist pressures to extend the autonomy granted to military systems [3]. It thus seems reasonable to review the fundamental principles which underwrite existing prohibitions on the use of autonomous technologies, as well as to try to anticipate how they may need to be augmented or extended to address new technological possibilities as they begin to appear on the horizon.

² The military of South Korea already has plans to deploy autonomous robots armed with machine guns and live ammunition along the border with North Korea. The system is designed by Samsung, and will shoot at any human attempting to cross the DMZ.

2. Just War Theory

Walzer's [1] just war theory aims to provide a theoretical framework for debate about the morality of specific choices and actions with regard to war by establishing a small set of principles that effectively capture general moral sentiments. Rather than dismiss all war as immoral, it seeks to carefully distinguish those specific acts that are moral so as to deny authority to those who would abuse moral sentiments in taking a nation to war, or try to legitimate immoral acts during the conduct of a war. It establishes a rational framework for distinguishing just from unjust acts of, and in, war and takes a liberal approach which seeks to protect the rights of individuals and states from unjust harms. It derives its rational principles from reflecting on shared moral sentiments and intuitions, and from recognizing the conventional nature of war and the rules of war which tend to govern it. As McMahan [2] makes clear, there are numerous inconsistencies in how Walzer articulates the foundations of the theory in terms of individual rights, state rights, moral sentiments and conventional norms. His critique, however, aims to preserve most of the overall structure of just war theory by putting it on firmer foundations based in individual rights.

A key distinction in just war theory is drawn between what are just reasons for going to war, *jus ad bellum*, and what are just acts in the fighting of war, *jus in bello*. For Walzer, the two are completely independent of one another. Thus, for Walzer, the actions of soldiers on both sides of a war can be just if they observe the overriding principles of *discrimination* and *proportionality*. The substantive critiques of Walzer challenge this independence, and I believe that, as a moral question, proportionality depends in important ways upon the reasons for going to war. Despite this, the distinction is highly relevant to the practice and legal regulation of war and reflects two distinct moral questions, even if the moral character of how a war is fought cannot be fully determined without considering the reasons for fighting.

3. Autonomous Technology and *jus ad bellum*

There are at least two significant ways in which autonomous technologies might influence the choice of a nation to go to war. The first of these is that such systems could directly threaten the sovereignty of a nation. As such it would challenge just war theory, or the theory might have to be extended to cover this possibility. The second way is one of the reasons many people fear the development of autonomous killing machines, like robotic soldiers, though it is not really a problem for just war theory itself. This is the belief that these technologies will make it easier for leaders who wish to start a war to actually start one, in short that autonomous technologies would lower the *barrier to entry* to war, or be directly responsible for starting a war intentionally or accidentally.

3.1 Autonomous Technologies Challenging Sovereignty

It seems that autonomous technologies offer a great potential for starting wars accidentally, and perhaps even doing so for their own purposes. While this latter potential seems more like science fiction than a real possibility, it is nonetheless useful to consider how just war theory can help aid our thinking about such a situation. The

theory accepts only a few very specific causes for war as being just. In principle, only aggression against a nation's sovereignty by another nation is a just cause. Strictly speaking, the aggressor has acted unjustly and the defender has a right to self-defense, and thus may justly fight against the aggression—though history is rarely so clear cut. By extension of this principle, other third party nations may justly join in the fight on the side of nation defending itself against the aggressor, though they are not necessarily *required* to do so, *e.g.* if doing so would threaten their own existence or sovereignty. There are only a few and highly circumscribed exceptions to this principle, those being a pre-emptive strike against an immediately impending aggression, and a humanitarian intervention to stop severe human rights abuses or genocide.

With these general principles firmly in mind, we can imagine new ways in which autonomous technologies could impact upon sovereignty. First, there is the case in which an autonomous technology “accidentally” starts a war. This could be the result of human manipulation, a genuine technical error, or perhaps even by the purposeful intention of the technology. It could also turn against the nation that created it, resulting in one sort or another of a “robot revolution.”

3.1.1. Accidental War

The idea of an “accidental” war is closely related to our conception of the sovereignty of states, though is not truly a threat to sovereignty. An “act of war” is considered to be an intentional act committed by one state against another. Thus, an unintentional act which is interpreted as an act of war could lead to an accidental war. The possibility of an accidental war has always existed, and generally the decisions to go to war are based on intentions that pre-exist any specific act of war, which is only the proximate cause or a token of justification. Autonomous technological systems introduce new dangers, however, in that they might act in unanticipated ways that are interpreted as acts of war.

There was a common fear throughout the Cold War era that the complex technological control systems for nuclear arms might malfunction, unintentionally starting a nuclear war that nobody could stop. To the extent that all large complex technological systems are prone to unpredictable errors in unforeseeable circumstances, the systems of control for autonomous robotic armies will be too. Further, to the extent that robots are used to patrol dangerous areas, contentious borders and political hot spots, it seems quite likely that some of their actions might be interpreted as acts of war, though no political or military official specifically orders such an act. While this is a real danger, it is not altogether unlike the threat posed by rogue officers and soldiers who accidentally or purposely commit such acts despite not being properly authorized by their chain of command, though they are likely to be aware of the possibilities for misinterpretation.

The wars that might result from such accidents cannot be just from the perspective of the unintentional aggressor, who then has an obligation to stand down from that aggression. While the state that is harmed by such unintentional acts has a genuine grievance, and has a right to defend itself, if it declares war in response, it will not be a just war unless the aggressor continues to pursue a war with further acts of aggression. Often, however, groups within one or both states are interested in having a war and will seize upon such incidents as opportunities to escalate hostilities and justify a full-scale war. In any case, it seems that just war theory provides the means necessary to interpret

and consider cases of unintentional acts of war, in which both the human chain of command and the autonomous technology do not intend the act of war in the relevant sense.

More complex are cases in which the autonomous technologies have intentions of their own—when they are near the Kantian moral-agent end of the autonomy spectrum. Again, such agents may act unintentionally and the situation would not be unlike those in which human acts unintentionally. However, a new kind of problem arises when autonomous technologies begin to act on their own intentions and against the intentions of the states who design and use them. These situations present many problems. First, it may be difficult to distinguish a genuine intention from a technical error, which casts doubt on just what the intention behind the act is. Further, such a display of incongruent intention might indicate that the autonomous system is no longer under the control of the state, or individual, which produced or employed it. As such, it might not be appropriate to attribute its actions as being representative of the state, *i.e.* it is a rogue agent. It is not clear what responsibility a state has for the actions of rogue agents that it creates or supports, *i.e.* whether it is liable to be attacked, but the rogue agents themselves are capable of taking some of the responsibility in virtue of their moral autonomy and are certainly liable to be attacked.

These possibilities also bring into focus a new kind of question, namely whether it is wise, or just, to build and install such automated systems in the first place, given the kinds of risks they engender. This question has been asked most pointedly with automated nuclear defense systems. Here the stakes are very high, and it seems morally wrong to leave the ultimate decision to launch a nuclear assault up to an automatic process rather than a human, who might quite reasonably fail to act out of a moral sense of duty, and thus avert a total nuclear war.³ In fact, we might want to design military robots in a way that allows them to refuse orders that they deem to be illegal, unjust or immoral, though researchers are only beginning to think about how we might do that.⁴ To the extent that autonomous systems begin acting on their own intentions, however, then we might be concerned about their acts of aggression towards their own state, as well as towards other states.

3.1.2. Robot Revolutions

The notion of a robot revolution is as old as the stage play in which the word “robot” was first coined, Capek’s *R.U.R.* [4], in which worker robots all over the world join in a global revolution and overthrow the humans. While this may seem like a fanciful bit of science fiction, we can ask serious questions about the moral status of such revolutions according to just war theory. Let us imagine a situation in which a nation is taken over by robots—a sort of revolution or civil war. Would a third party nation have a just cause for interceding to prevent this?

³ There is a rich literature on the morality of “Doomsday” devices in nuclear deterrence, see [5] on this. The deterrence literature is preoccupied with strategic considerations, and with the credibility of the deterrent threat, and thus sees the amoral automaticity as a strategic advantage. Morally, this seems contemptible because it provides for no moral reflection of the consequences of such an action. If we take the notion of autonomous moral agents seriously, then perhaps a machine could make a moral determination without human intervention, something not considered during the Cold War.

⁴ The roboticist Ronald Arkin [6] of Georgia Tech is working on just such a robot control architecture, under contract from the US Army.

In such cases just war theory might be of two minds, depending upon the moral status and autonomy of the robots in question. On the one hand, it is a violation of sovereignty to interfere in the civil war of another nation. On the other hand, it is legitimate to intervene to aid a state that is being threatened by a foreign power. Thus a great deal turns upon whether the robots who lead the revolution are seen as autonomous moral agents with a political right to revolt, or if they are the non-autonomous agents of some other morally autonomous agent, or if they are a non-moral or amoral system under the control of no autonomous agents but simply threaten a state. A further possibility is that the robots might represent a humanitarian crisis if they were seeking to completely eradicate or enslave the humans, though intervention would clearly be just in such a case.

Even if we consider robotic moral agents as being different than humans in important ways, the question regarding their right to revolution ultimately turns on whether they are entitled to a right to self-determination. For humans this right is tied up in other rights to individual freedom, to not be killed, to not be tortured, *etc.* While it remains to be seen if these rights are separable, due to different technological systems which only partially replicate human mental and moral capacities, we can answer this question based on the relevant portion of those rights. Assuming that we have a theory of what is required for someone to be entitled to a right to self-determination, *e.g.* a sufficiently sophisticated moral autonomy, then the robot rebels will either have this right or they will not. If they do, then just war theory will treat them just as it treats human rebels seeking to take control of their own country. If they do not, then the theory will treat them as agents of another power. If that power has legitimate claims to revolution, then there is no right for third parties to intervene. If they are agents of a foreign or otherwise illegitimate power, or of no autonomous moral agency at all—a sort of man-made disaster—then the threat they pose will justify intervention by outside powers who seek to protect the state from the robots. If in any of these cases the robots represent a humanitarian threat, then third party nations may also intervene on humanitarian grounds.

Thus we can see again that just war theory is able to deal with many of the more fanciful cases in a fairly traditional way. It leaves open, however, the critical question of whether or when machines might be due some or all of the rights of humans, as this is completely outside the scope of theory of just war. It is also not clear that it can always be easily determined whether machines are autonomous and, if not, on whose commands or intentions they are acting.

3.2 Technologically Lowering the Barriers of Entry to War

I believe that one of the strongest moral aversions to the development of robotic soldiers stems from the fear that they will make it easier for leaders to take an unwilling nation into war. This is readily apparent in light of recent history, including the 1991 Persian Gulf War, the 1999 war in Kosovo, and the 2003 invasion of Iraq. These events have brought into sharp relief the complex relationships between the political requirements on national leaders, the imagery and rhetoric of propaganda and the mass media, and the general will of citizens in the processes of deciding when a democratic nation will go to war.

Irrespective of the underlying justness of the motives, when the leadership of a state decides to go to war, there is a significant propaganda effort. This effort is of particular importance when it is a democratic nation, and its citizens disagree about whether a war is worth fighting, and there are significant political costs to a leader for going against popular sentiments. A central element of war propaganda is the estimation of the cost of war in terms of the lives of its citizens, even if that is limited to soldiers, and even if those soldiers are volunteers. A political strategy has evolved in response to this, which is to limit military involvement to relatively “safe” forms of fighting in order to limit casualties, and to invest in technologies that promise to lower the risks and increase the lethal effectiveness of their military.

We can see these motivations at work in the NATO involvement in Kosovo in 1999, in which NATO limited its military operations to air strikes.⁵ The political pressure to avoid casualties among a nation’s own soldiers is thus often translated into casualties among innocent civilians, despite this being fundamentally unjust. Technologies can shift risks away from a nation’s own soldiers and can thereby add political leverage in both domestic politics, through propaganda, and in diplomatic efforts to build alliances among nations. Thus, the technology functions not only in the war itself, but in the propaganda, debate and diplomacy that brings a nation into war. In this regard, it is primarily the ability of the technology to limit risks to the nation that possesses it, and its allies, that allows it to function in this way. Certainly the replacement of soldiers by robots could achieve this in a new and somewhat spectacular way, perhaps by eliminating the need for any soldiers from that nation to actually go to the battle zone.

Given these facts, it is quite reasonable to conclude that the introduction of any technology that can limit the risks to a nation’s soldiers and civilians would serve a similar function. In some sense, all military technologies that work well serve this function, to some degree or when taken altogether, whether it is better airplanes, or better body armor, or better bombs, or better communications, or better strategies, or better robots, or even better battlefield surgery techniques. Indeed, recent US media has spent a great deal of energy trumpeting the technological sophistication of the US military in this way. The ultimate aim of all military technology is to give an advantage to one’s own soldiers, and this means limiting their risks while making it easier to kill enemy soldiers and win the war. So in general, all military technological development aims at this same objective. Moreover, even with the most sophisticated machinery, and guarantees of extremely low casualties, most citizens in most countries are still averse to starting an avoidable war, and are nearly always averse to starting an unjust war.

⁵ It is important to note that Walzer, in his introduction to the third edition of [1] in 1999, has criticized this particular decision as being unjust because of the nature of the war that resulted, namely an air campaign that disproportionately harmed innocent civilians and their property, rather than the military forces it was seeking to suppress. Despite efforts to avoid directly bombing civilians, the air strikes intentionally targeted civilian infrastructure (so-called dual-use targets) such as bridges, roads, power stations, water purification plants, *etc.*, which greatly impacted the lives and safety of civilians. Moreover, while the warring military factions in Kosovo were not able to conduct major military operations, they were not eliminated or significantly hurt by the air strikes either, and indeed switched to guerilla and urban fighting tactics which further imperiled civilians. Walzer does not dispute that the underlying cause of NATO involvement was just. He simply argues that NATO should have committed to a ground war which would have saved many innocent civilians from harm, even though it would have increased the risks for the NATO soldiers.

Even if robots did make it easier for a nation to go to war, this in itself does not decide whether that war is just or not. There is, however, a deeper question of political justice lurking here that concerns whether it is desirable to make it practically easier to go to war or not. If we assume that only nations fighting just wars will utilize such technologies, then it would not necessarily be unjust or immoral to develop those technologies. However, history instructs us that all wars involve at least one unjust (or badly mistaken) nation, and so the chances that such technologies will enabling future injustices is a real and legitimate concern. Moreover, it is likely that obviously just wars do not need their barriers lowered, and so this function tends to aid the propaganda of aggressors more than that of just defenders. If we agree that the majority of wars are in fact unjust on one side, then any technologies that lower the barriers to entry of war are empirically more likely to start wars period, even if one side has just cause to enter it. Still, this is an argument against militarization in general, and not specifically about autonomous systems and robots, even if they are a dramatic example of it. From an empirical perspective, it is also important to consider why these specific technologies are being heavily invested in rather than other technologies, and if it is primarily due to this propagandistic function, then this should raise concern.

4. Autonomous Technology and *jus in bello*

Walzer claims that just war theory is largely indifferent to the kinds of technology that are used in battle. As far as he is concerned, there are individuals who have a right not to be killed, the innocent civilians, and those who have given up that right by taking up arms, the uniformed combatants. As it is morally permissible to kill uniformed combatants, it does not matter much how one goes about killing them (assuming that one recognizes their right to surrender, *etc.*). However, there are a number of international conventions which do limit the use of specific kinds of weapons, such as chemical, biological and nuclear weapons, as well as landmines, lasers designed to blind soldiers, and other sorts of weapons. There are various reasons for the existence of these treaties, and several principles which determine what kinds of technologies are permissible as weapons of war. In this section, we will consider how autonomous technologies might challenge the standards of *jus in bello*.

Despite Walzer's claims that his theory does not care about the technologies used for killing, he does discuss several specific technologies in terms of how they changed the conventional standards of war. In particular, the use of submarines, aerial bombing and the atomic bomb, all relatively new technologies, changed the accepted conventions of warfare during World War II.

The clearest example of this is the way in which the specific technologies of submarine warfare in World War II served to repeal a centuries-old naval warfare convention. The convention held that there was a moral duty to rescue the surviving crew of a sinking enemy ship once a battle was over. Over the long history of European naval warfare, this convention made sense to all participants, since combat usually occurred in open seas, often hundreds or even thousands of miles from safe harbors, and disabled or sinking ships generally had significant numbers of survivors. From the development of submarines through World War I, this convention held for submarines as it had for all other ships. It was decided during World War II, however, that this convention no longer applied to submarines. The reason for this was that requiring a

submarine to surface and conduct rescue operations would make it too vulnerable to detection from radar and attack from airplanes armed with torpedoes. Additionally, the small submarines (with crews of less than 50) could not spare space for survivors on-board, adequate guards to take officers prisoner, or space to store much rescue equipment (ships sunk by submarines often had more than 1,000 people on board), all of which made rescue efforts challenging and impractical. They could, however, right upset lifeboats and provide food and water, as well as pick up men from the sea and put them in their own lifeboats, but these activities were considered too risky.

While there were some specific and dramatic events that led to the abandonment of this particular convention for submarines, it was largely due to the fact that obedience to the convention was so risky that it would render submarine warfare impractical. The official abandonment of the convention occurred when the German Admiral Doenitz issued the *Laconia* Order in 1942, which expressly directed submarines to not engage in any form of assistance to survivors [1]. Doenitz was tried at Nuremberg for a war crime in issuing this order, but was acquitted of the charge by the judges. The legal decision rested primarily on the fact that because both sides assented to the new convention in practice, the old convention was effectively annulled and submarines no longer had a moral obligation to rescue crews, despite the fact that it was sometimes safe for a submarine to rescue survivors. Walzer believes this is the correct moral interpretation, and accounts for it in his theory as a valid use of the principle of military necessity. That is, it became a military necessity to forgo the convention of rescue in order for submarine warfare to be an effective naval strategy, though this clearly makes naval warfare a more brutal affair.

I believe this deference to military necessity presents a significant weakness in just war theory, as formulated by Walzer, when viewed in the context of technological development. That is, why should we not say that submarines should not be used at all if they cannot be used in a manner which conforms to the conventions of just war? If we cannot argue this way, then there would seem to be a certain kind of impotence to using just war theory to argue against any technology that has the potential to change conventions via military necessity. Not only does this position mean that we have to accept all new technologies and the new conventions that arise from them, but also that we cannot judge the morality of devising various sorts of weapons. The question is: If I can only defend myself with an indiscriminate and disproportionate weapon, because that is the only militarily effective weapon I have, then was I acting unjustly when I chose to arm myself with that weapon rather than another weapon which could be discriminate and proportionate? Do I have a moral duty to invest in weapons that will not tend to put me in positions where I may be motivated to act unjustly (indiscriminately and disproportionately) in the future? If so, could robot soldiers be such a technology?

This failure in Walzer's formulation stems from its loose foundational considerations. While Walzer is willing to assert individual rights as the basis for prohibitions against killing civilians, he mistakenly asserts that soldiers forgo their rights not to be killed by merely taking up arms. Further, he seems to believe that the restrictions on the use of weapons against soldiers, and the rights of soldiers to surrender, rescue, medical aid, *etc.*, are a matter of convention between states who see these conventions as being in their mutual interests. Thus, there is no firm moral foundation to prevent states from abandoning these conventions when they are deemed

not to be in their mutual interest. A convention depends only upon both sides assenting to it, as in Walzer's analysis of the Doenitz decision which rests upon the fact that both sides of the conflict observed the same convention.

Apart from convention, Walzer might appeal to moral sentiments in determining the morality of certain military strategies and technologies. To the extent that it derives its principles from moral sentiments, just war theory is an attempt to describe sentiments that occur *after* an event. In the case of many new technologies, we do not really know how that technology will function in the complex socio-technical system of war. Submarines, radar and airplanes armed with torpedoes had never been used together in warfare before WWII, so nobody really knew how they would be used. Indeed, the German admiralty only abandoned the sea rescue convention for submarines in 1942, well into the war. Thus it seems that if we cannot predict military necessity very well, just war theory as it stands cannot tell us much about which technologies might be best left undeveloped.

The more critical issue that just war theory faces is that the conventionalist and sentimentalist interpretations both fail to secure a moral foundation for restrictions on actions against combatants *in bello*. Most generally, if we accept that the combatants on a just side of a war have not waived their rights not to be killed [2], then no conventional agreement between states can waive or trump that right. Similarly, if sailors have a moral right to be rescued after their ship is sunk, then neither the demands of military necessity, nor the technological limitations of submarines, nor the conventional agreements between belligerent navies can waive that right, even if it makes it impractical to respect it. The practicality issue comes into play only when we come to consider the legal restrictions on combat, not its fundamental morality. Thus, we might accept a certain degree of immorality in our laws because of the impracticality of judging and enforcing moral justified laws [2].

The real question then becomes one of what moral rights individuals have against the use of specific technologies, and of the moral duties of states in the development of arms that might be used in hypothetical future wars. Again there is the distinction between fundamental morality and practical law, but it seems possible in principle to develop a moral foundation for arms control and limitations on the design and use of various technologies. While it is well beyond the scope of this essay, it will remain a topic for further research.

4.1 Distinguishing Civilians & Combatants

On Walzer's interpretation of just war theory, the most fundamental distinction made by just war theory is that between combatants and civilians. While this distinction fails, like military necessity, to find solid moral grounding, it proves quite useful in establishing practical laws for regulating war. This distinction makes it legally permissible, at least sometimes, for combatants to kill enemy combatants. It also makes it possible to say that it is almost never legally justified for combatants to kill innocent civilians. There are, of course, shades of grey. Even combatants retain certain rights, like the right to surrender, and not to be killed unnecessarily. There are also cases in which it is legally permissible to kill civilians—but these cases must meet a very strict and limiting set of conditions, and even those are contentious. Further problems arise in guerrilla and insurgent warfare, in which combatants pose as civilians.

In this section I want to consider several different aspects of the problem of distinguishing civilians and combatants as it relates to autonomous systems. First, the practical ability of autonomous technologies to draw this distinction correctly is crucial. On the one hand, it has been argued that this ability makes the use of such systems *morally required* if they are available. What is more surprising is that it is human rights groups, such as Human Rights Watch [7] in demanding the use of only “smart” bombs in civilian areas, who have made this argument. On the other hand, it is the fear of indiscriminate violence, perhaps mixed with impoverished cultural and social intelligence, that makes robotic soldiers seem particularly dangerous and morally undesirable.

The relevance of the civilian-combatant distinction to robotic soldiers is that if they are to be autonomous in choosing their targets, they will have to be able to reliably distinguish enemy combatants from civilians. It seems that this capability will remain the most difficult theoretical and practical problem facing the development of such robots. While there are technologies for picking out humans based on computer vision, motion and heat patterns, it is extremely difficult to identify particular people, or even types of people, much less to categorize them reliably into groups such as “friend” or “foe,” the boundaries of which are often poorly defined and heavily value-laden.

In keeping with the Human Rights Watch argument, there is a line of reasoning which asserts that advanced technologies have the potential to be superior to human capabilities. Arkin [6] argues that if we can achieve the proper discriminatory capabilities in robots, they may very well be *morally superior* to human soldiers. The argument maintains that if machines are better able to discriminate civilians from combatants, then it will make fewer mistakes than humans. Moreover, because it is a machine, it will not feel the psychological and emotional stress of warfare, and thus will not be inclined to commit war crimes or atrocities as humans under such pressure might. As such, there is not only a moral obligation to use such systems when they are available, but also to build them (insofar as war is taken as an unavoidable feature of human civilization). This all depends, of course, on the actual abilities of the technology, and the abilities of combatants to fool such systems into misidentification.

4.2 “Push-Button” Wars

Walzer notes that a radical transformation in our understanding of the war convention occurred with the rise of the modern nation-state. Before this, warfare was largely conducted by individuals who freely chose to participate in a given war, and a given battle. With the rise of the modern nation-state came the power to recruit and conscript individuals into standing armies. Because of this, nearly all individual soldiers lost their freedom to choose which wars and which battles they would fight in, even if they had the choice of whether to volunteer for military service. The consequences for moral estimations of conduct in war were thus reshaped by our knowledge that many of the actual combatants in war are not there freely, and thus deserve a certain degree of moral respect from their own commanders as well as from enemy commanders. While it is permissible to kill them, they still have the right to surrender and sit-out the rest of the war as a prisoner. It is also morally required that commanders seek out ways of winning battles that minimize killing on both sides. That is, the lives of the enemy still have moral weight, even if they weigh less than the civilians and combatants on one’s own side. And the lives of one’s own soldiers also

count in a new way. Whereas it might be moral to lead a group of soldiers on a suicidal charge if they all volunteer for that charge, ordering conscripted soldiers into such a charge is usually immoral. In the same way, it is deemed highly honorable to throw oneself on a grenade to save one's comrades, but not to throw one's comrade onto the grenade—the autonomy of the individual soldier to choose his or her fate has moral implications.

The use of autonomous systems may similarly change our conception of the role of soldiers in war, by fully realizing a “push button” war in which the enemy is killed at a distance, without any immediate risk to oneself. This approach to war could be deemed unjust by traditional conventions of war because those doing the killing are not themselves willing to die. This principle is fundamental because it powerfully influences our sense of fairness in battle, and concerns the nature of war as a social convention for the settling of disputes. Insofar as it can serve this purpose, both sides must essentially agree to settle the dispute through violence and, by the norms of the convention, the violence is to be targeted only at those who have agreed to fight, *i.e.* the combatants. Thus it is immoral to kill civilians, who have not agreed to fight. This convention is only abandoned in a “total war” in which no actions are considered unjust because the stakes of losing are so high. By fighting a war through pressing a button, one does not fully become a combatant because one has not conformed to the norms of war in which both sides agree to risk death in settling the dispute. The limitations of such a conventionalist notion of just war have been noted above, however, and there would seem to be no deeper moral obligation for a just combatant to risk their own lives in defense of their state.

We could imagine a war in which both sides sent only robots to do the fighting. This might be rather like an extremely violent sporting contest in which the robots destroy each other. For this to actually count as a war, and not merely a sport, however, political decisions would have to be made as a result of this competition, such as ceding territory. While it might seem unlikely that a nation would simply give up its territory or autonomy once its robots were destroyed, this is not an unreasonable or impossible outcome. It might also be deemed moral to fight to the last robot, whereas it is generally not deemed moral to fight to the last human. While many nations have surrendered after the crushing defeat of their armies but before the actual conquest of their lands, it would seem likely that a state might continue fighting with humans after its robots have been destroyed, rather than simply capitulate at that point. In general, I think it is fair to say that an exclusively robotic war might even be a highly preferable way of fighting to what now exists. In its most extreme form, we could even imagine a decisive war fought without a single human casualty.

Such a war would not be completely without its costs and risks, however. First, such a war would have to take place somewhere, and it seems likely that the destruction of natural resources and civilian property would be highly likely in most locations. As the most common military objectives are to hold cities and towns, there is both the risk of harming civilians in the course of fighting, and the problems of holding towns, and thus controlling and policing civilian populations with robots. There would also be the cost in terms of the time, money and resources devoted to building up these robot armies.

At the other extreme lies the completely asymmetric “push-button” war. Thanks to science fiction and the Cold War, it is not hard to imagine an autonomous military

system in which the commander needs only to specify the military action, and press a button, the rest being taken care of by a vast automated war machine. We could even imagine a civilian government that has completely replaced its military with a fully automated system, perhaps designed and staffed by civilian technicians, but one that did not require any uniformed soldiers to operate. Such a system would, I believe, seriously challenge the conventional concept of war.

In a completely asymmetric war, in which one side offers no legitimate uniformed combatants in battle, but only robots, our moral sentiments could be profoundly upset. If one nation fights a war in which its soldiers never appear on the battlefield, offering no opportunity for them to be killed, then the combatants are all machines and the humans are all civilians. As in a guerrilla war, one side presents no legitimate human targets to be killed. A legitimate army would not have any opportunity to reciprocally kill the soldiers of their opponents in such a situation (and could only inflict economic damage on their robots). This could thereby be interpreted as a fundamental violation of the war convention itself, like showing up for a duel in armor or sending a proxy, and thereby as a nullification of the associated conventions. Seen another way, such a situation might also be presented as an argument in favor of terrorism against the civilians who sit behind their robotic army. It could be argued that because such an army is the product of a rich and elaborate economy, the members of that economy are the next-best legitimate targets. This possibility should alert us to unsuitability of conventions and moral sentiments, rather than individual rights, as a basis for just war theory, since we would not want a theory of just war which legitimizes terrorism.

If we instead see the foundations of just war as deriving from individual rights, it would be unreasonable to insist that a nation fighting for a just cause is obliged to let an unjust aggressor kill its citizens even though it has the technological means of preventing this. Indeed, outside of Walzer's interpretation of the moral equality of soldiers, we do not expect a technologically superior nation to refrain from using its available technologies simply because they give too great of an advantage, nor do we expect a larger army to use special restraint in fighting a smaller army out of a moral sense of fairness. Similarly, as long as the robot army is no more likely to cause unjust harms than a human army, it would seem to offer a superior military advantage in limiting the risks to one's own citizens.

There is a compelling rationale for a nation desiring to defend itself without risking human lives. That is, a nation could quite reasonably decide that it does not want its children to be trained as soldiers or sent to war, and so develops a technological solution to the problem of national defense that does not require human soldiers, namely a robot army. In such a case it would not seem to be immoral to develop and use that technology, and we might go even further and say it is morally required for that nation to protect its children from becoming soldiers if it is within their technological capacity to do so. If an aggressor invaded this nation, I do not think many people would raise a moral objection to their using robot soldiers to defend themselves.

Of course, the push-button war is already available in a certain sense, namely for those countries with superior air forces and a willingness to bomb their enemies. The practical consequence of such wars is the asymmetric war in which one side is so obviously technologically powerful that it does not make much sense for the opposition to face it in the traditional manner. The result is often guerrilla warfare, and sometimes terrorism. The advent of robot armies may further exacerbate such situations, but

would not seem to be fundamentally different. Their development and employment should, however, take into consideration that these are likely responses to the use of robot armies, even if they are not morally just responses.

5. Conclusions

Ultimately, just war theory concludes that the use of autonomous technologies is neither completely morally acceptable, nor is it completely morally unacceptable under Walzer's [1] interpretation of just war theory. In part this is because the technology, like all military force, could be just or unjust, depending on the situation. This is also, in part, because what is and is not acceptable in war, under this interpretation, is ultimately a *convention*, and while we can extrapolate from existing conventions in an attempt to deal with new technologies, like autonomous killing machines, this process can only be speculative. It is up to the international community to establish a new set of conventions to regulate the use of these technologies, and to embody these in international laws and treaties. Such a process can be informed by Walzer's theory, but his approach is to appeal to conventional practices as the ultimate arbiter of military necessity when it comes to technological choices. In light of this, we may wish to extend or revise the theory of just war to deal more explicitly with the development and use of new military technologies. In particular, we might seek to clarify the moral foundations for technological arms control, perhaps upon individual rights or another solid moral ground. Such a theory might also begin to influence the practical control of autonomous weapons systems through international laws and treaties. I believe that this would be a promising approach for further work.

References

- [1] M. Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, Basic Books, NY, 1977.
- [2] J. McMahan, The Sources and Status of Just War Principles, *Journal of Military Ethics*, 6(2), 91-106, 2007.
- [3] J. S. Canning, A Concept of Operations for Armed Autonomous Systems: The difference between "Winning the War" and "Winning the Peace", presentation at the Pentagon, 2007.
- [4] K. Capek, *Russum's Universal Robots (RUR)*, 1921.
- [5] L. Alexander, The Domsday Machine: Proportionality, Prevention and Punishment, *The Monist*, 63, 199-227, 1980.
- [6] R. C. Arkin, Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture, Georgia Institute of Technology, Technical Report GIT-GVU-07-11, 2007.
- [7] Human Rights Watch, International Humanitarian Law Issues in the Possible US Invasion of Iraq, *Lancet*, Feb. 20, 2003.