This is the author's accepted manuscript of a book chapter that has been published in Edward Downs (Ed.) *The Dark Side of Media & Technology* (pp. 274-285). Peter Lang, 2019. Please do not circulate.

The Killer App:

Drones and Autonomous Machines

David J. Gunkel – Northern Illinois University

There are at least two ways to interpret the title to this chapter. "Killer app" is Silicon Valley speak for an application that provides proof of concept for a technology or an ensemble of technologies. Understood in this way, the drone is the killer app of a number of related technological innovations: remote telepresence, augmented reality, HD imaging, and wireless data communications. But we can also read the title in a more literal fashion—understanding technologies like drones and other autonomous machines as applications that can kill. Need to neutralize enemy combatants and terrorists? Need to locate and subdue a criminal suspect? Need to decide who lives and who dies in a fatal self-driving car accident? There's an app for that.

Responses to these killer apps have pulled in two seemingly opposite directions. On the one hand, the drone, or what the US military calls an Unmanned Aerial Vehicle (UAV), has been celebrated as a remarkable innovation that is perfectly designed for current global conflicts. "They are," Mark Bowden (2013) writes, "remarkable tools, an exceedingly clever combination of existing technologies that has vastly improved our ability to observe and to fight. They represent how America has responded to the challenge of organized, high-level, stateless terrorism—not timidly, as bin Laden famously predicted, but with courage, tenacity, and ruthless ingenuity." On the other hand, Human Rights Watch and various Legal research centers have been highly critical of the lethal capabilities of this new weapon system and its mode of deployment:

> As covert drone strikes become the norm, actions or conduct by individuals that, in other circumstances, would lead to investigation or detention are increasingly blurring into a basis for lethal targeting. The result is that an ever-greater number of individuals are vulnerable to lethal targeting, and accordingly a larger number of civilians are at risk of either being killed or harmed as a result of collateral damage, or due to mistaken beliefs about their identity or associations. (Center for Civilians in Conflict, 2012, p. 75)

The following chapter does not take sides in this debate but asks questions that remain largely unasked by both sides involved. The questions are: "When autonomous machines kill, who (or maybe even 'what') is responsible?" Who, in other words, is to be praised for successful operations undertaken by autonomous machines? And; Who or what can or should be blamed for mistakes or failures? This is an entirely different kind of inquiry and one that can get us thinking and talking about the larger social opportunities and challenges regarding new forms of autonomous or semi-autonomous technology like drones and related systems.

Default Setting

As with all such questions, there is a kind of standard response that we might call the default setting. A default is a mode of behavior or a value that is already assigned and operative without needing to think about it or deliberately deciding to do so. It is, in other words, the "normal way" of doing things. And according to the normal way of doing things, we recognize that drones are technologies, and technologies are just tools or instruments that are used more or less appropriately by human beings. This is what is called the "instrumental theory of technology" and it informs those common sense opinions like "Drones don't kill people. People kill people." In other words, it is not the technology that is to blame; it is how the technology comes to be used or misused that really matters.

The instrumentalist theory, as Andrew Feenberg (1991) writes, "offers the most widely accepted view of technology. It is based on the common sense idea that technologies are 'tools' standing ready to serve the purposes of users. Technology is deemed 'neutral,' without

valuative content of its own" (p. 5). Technology, therefore, is essentially without intrinsic value; it is a neutral tool. What ultimately matters is not the technology *per se*, but how it comes to be used and for what purpose. And the current debate about drones in domestic airspace show us how widespread this way of thinking is. If used for the purposes of finding a lost child or pursuing a violent criminal, drones are (it seems) perfectly acceptable. But if used to spy on people and their activities, then, so the argument goes, there should be some restrictions and even prohibitions. This means, in other words, that the drone has no inherent moral status in and of itself. It is neither good nor bad. What matters is how it comes to be used. "Morality," as J. Storrs Hall (2001) explains, "rests on human shoulders, and if machines changed the ease with which things were done, they did not change responsibility for doing them. People have always been the only 'moral [and legal] agents.'"

Consequently, mobilizing this default instrumental theory of technology has distinct advantages. It affirms that technology is just a tool of human action and decision making and locates responsibility in a widely accepted and intuitive subject position—in the hands of the human user of the tool. This explanation conforms to the most common and accepted view we have of technology. It therefore appears to be "normal" and largely unremarkable. But there are also problems with taking this approach. Unlike a hand tool or even a personal computer, the drone does not have a single and easily identifiable user. It is always deployed within a complex network of operators, managers, and commanders (cf. Currier, 2015 for the complexities of this "Kill Chain") and is therefore exposed to what Martha Nissenbaum (1996, p. 25) calls "the many hands problem." Although connecting drone activities to individual human action and decision making is entirely reasonable and expedient, the complexity of the technological system makes the identification or assignment of responsibility difficult and potentially obscure.

Distributed Responsibility

In response to these problems, alternative theories of social action have been proposed and operationalized. F. Allan Hanson (2009, p. 91), for instance, introduces something he calls "extended agency theory," which is a kind of extension/elaboration of the "actor-network

theory" initially developed by Bruno Latour (2005). According to Hanson, the best and most expedient way to respond to technological systems, like drones, is to formulate what he calls a "joint responsibility," where "moral agency is distributed over both human and technological artifacts" (Hanson, 2009, p. 94).

According to this way of thinking, actions undertaken with technological systems like a drone are the product of a network of interacting agents: the operators in the field who actually fly and control the UAV; the unit commanders and managers who make decisions and issue orders; the civilian lawmakers and leaders who establish policy; and the technological object itself, which is not neutral, but helps to shape and influence what actions are possible. This latter aspect is a form of technological determinism, which recognizes that technology is never neutral but actively contributes to the way something comes to be understood, deployed, and utilized. As Marshall McLuhan (1995) once argued, in direct opposition to the instrumentalist theory, "our conventional response to all media, namely that it is how they are used that counts, is the numb stance of the technological idiot" (p. 18).

Similar proposals have been advanced and advocated by Deborah Johnson and Peter Paul Verbeek for dealing with innovation in information technology. "When computer systems behave," Johnson (2006, p. 202) writes, "there is a triad of intentionality at work, the intentionality of the computer system designer, the intentionality of the system, and the intentionality of the user." Verbeek (2011, p. 13), for his part, makes a comparable assertion: "I will defend the thesis that ethics should be approached as a matter of human-technological associations. When taking the notion of technological mediation seriously, claiming that technologies are human agents would be as inadequate as claiming that ethics is a solely human affair." For both Johnson and Verbeek, responsibility is something that is distributed across a network of interacting components and these networks include not just other human persons, but organizations, natural objects, and technologies.

This hybrid formulation—what Verbeek calls "the ethics of things" and Hanson terms "extended agency theory"—has advantages and disadvantages. To its credit, this approach appears to be attentive to the exigencies of life in the 21st century. None of us, in fact, make decisions or act in a vacuum; we are always and already tangled up in networks of interactive

elements that complicate the assignment of responsibility and decisions concerning who or what is able to answer for what comes to pass. And these networks have always included others—not only other human beings but institutions, organizations, and even technological components like the robots and algorithms that increasingly help organize and dispense with social activity. This combined approach, however, still requires that someone decide and answer for what aspects of responsibility belong to the machine and what should be retained for or attributed to the other elements in the network. In other words, "extended agency theory," will still need to decide *who* is able to answer for a decision or action and *what* can be considered a mere instrument (Derrida, 2005, p. 80).

Furthermore, these decisions are (for better or worse) often flexible and variable, allowing one part of the network to protect itself from culpability by instrumentalizing its role and deflecting responsibility and the obligation to respond elsewhere. This occurred, for example, during the Nuremberg trials at the end of World War II, when low-level functionaries tried to deflect responsibility up the chain of command by claiming that they "were just following orders." But the deflection can also move in the opposite direction, as was the case with the prisoner abuse scandal at the Abu Ghraib prison in Iraq during the presidency of George W. Bush. In this situation, individuals in the upper echelon of the network deflected responsibility down the chain of command by arguing that the documented abuse was not ordered by the administration but was the autonomous action of a "few bad apples" in the enlisted ranks. Finally, there can be situations where no one or nothing is accountable for anything. In this case, moral and legal responsibility is disseminated across the elements of the network in such a way that no one person, institution, or technology is culpable or held responsible. This is precisely what happened in the wake of the 2008 financial crisis. The bundling and reselling of mortgage-backed securities was considered to be so complex and dispersed across the network that, in the final analysis, no one was able to be identified as being responsible for the collapse.

Machine Ethics

A third alternative comes in the form of something that goes by the name "machine ethics." And there has, in fact, been a number of recent proposals addressing this innovation. Wendell Wallach and Colin Allen (2009, p. 4), for example, not only predict that "there will be a catastrophic incident brought about by a computer system making a decision independent of human oversight" but use this fact as justification for developing "moral machines," advanced technological systems that are able to respond to morally challenging situations. Michael Anderson and Susan Leigh Anderson (2011) take things one step further. They not only identify a pressing need to consider the moral responsibilities and capabilities of increasingly autonomous systems but have even suggested that "computers might be better at following an ethical theory than most humans," because humans "tend to be inconsistent in their reasoning" and "have difficulty juggling the complexities of ethical decision-making" owing to the sheer volume of data that need to be taken into account and processed (Anderson & Anderson, 2007, p. 5).

These proposals, it is important to point out, do not necessarily require that we first resolve the "big questions" of AGI (Artificial General Intelligence), robot sentience, or machine consciousness. As Wallach (2015, p. 242) points out, these kinds of machines need only be "functionally moral." That is, they can be designed to be "capable of making ethical determinations…even if they have little or no actual understanding of the tasks they perform." But would this even apply in the case of drones? We are, in fact, told by both official government sources and the press that drones are not the "robotic killing machines" of science fiction. They are always tethered to and under the control of a human operator. This statement is correct, but not entirely accurate. The fact is that most drone operations can be pre-programmed and automated. Algorithms residing on the downlink computer are able to take control of flight operations, draw down and analyze large sets of intelligence data, and even perform target acquisition and discernment. In fact, these automated systems are designed to do just about everything except pull the trigger. "As the layers of software pile up between us and our machines," Colin Allen (2011, p. 1) argues, "they are becoming increasingly independent of our direct control. In military circles, the phrase 'man on the loop' has come to

replace ‘man in the loop,’ indicating the diminishing role of human overseers in controlling drones and ground-based robots that operate hundreds or thousands of miles from base.” The driverless car presents us with another notable example. In fact, the term “driverless” is technically incorrect. The autonomous vehicle, whether the Google Car or one of its competitors, is not driverless; the vehicle is controlled by an autonomous system that is designed to make decisions without direct human involvement. This point was recently acknowledged by the National Highway Traffic Safety Administration (NHTSA), which in a 4 February 2016 letter to Google, stated that the company’s Self Driving System (SDS) could legitimately be considered the legal driver of the vehicle: “As a foundational starting point for the interpretations below, NHTSA will interpret ‘driver’ in the context of Google’s described motor vehicle design as referring to the SDS, and not to any of the vehicle occupants” (Hemmersbaugh, 2016). Although this decision is only an interpretation of existing law, the NHTSA explicitly states that it will “consider initiating rulemaking to address whether the definition of ‘driver’ in Section 571.3 [of the current US Federal statute, 49 U.S.C. Chapter 301] should be updated in response to changing circumstances” (Hemmersbaugh, 2016). Consequently, as we develop machines with increasing levels of autonomy and confront questions concerning the assignment of moral/legal accountability, it becomes increasingly important to consider developing a kind of functional morality—or at least some capability for responsible decision making—that is situated in the mechanism itself.

Doing so, however, presents both opportunities and challenges. On the positive side it can help sort out complex questions of moral accountability by both recognizing and seeking to develop artificial autonomous agents. This is not science fiction. There is precedent for this way of thinking. We already live in a world populated by artificial agents who are considered persons under the law, namely, the limited liability corporation. Corporations are, according to both national and international law, legal persons. And they are considered “persons” (which is, we should recall, a moral/legal classification and not an ontological category) not because they are conscious entities like we assume ourselves to be, but because social circumstances make it necessary to assign agency and responsibility to these artificial entities for the purposes of social organization and jurisprudence. Consequently, if entirely artificial and human fabricated

entities, like Google or IBM, are legal persons with associated social responsibilities, it would be possible, it seems, to extend the same moral and legal considerations to an AI or robot like the Google car, IBM's Watson, or UAVs. The question, it is important to point out, is not whether these mechanisms are or could be "natural persons" with what is assumed to be "genuine" moral status; the question is whether it would make sense and be expedient, from both a legal and moral perspective, to treat these mechanisms as responsible entities in the same way that we currently do for corporations, organizations and other human artifacts (see Gunkel 2018).

But there are, on the negative side, some significant problems. First, this proposal requires that we rethink everything we thought we knew about ourselves, technology, and ethics. It entails that we learn to think beyond the human exceptionalism (the assumption that it is only human individuals who can be considered responsible agents), technological instrumentalism, and many of the other -isms that have helped us make sense of our world and our place in it. In effect, it calls for a thorough reconceptualization of who or what should be considered a legitimate center of moral/legal concern and why.

Second, robots that are designed to follow rules and operate within the boundaries of some kind of programmed restraint—like Knightscope's security robots and Google's and Uber's self-driving vehicles—might turn out to be something other than what is typically recognized as a responsible agent. Terry Winograd (1990), for example, warns against something he calls "the bureaucracy of mind," "where rules can be followed without interpretive judgments" (pp. 182–183).

> When a person views his or her job as the correct application of a set of rules (whether human-invoked or computer-based), there is a loss of personal responsibility or commitment. The 'I just follow the rules' of the bureaucratic clerk has its direct analog in 'That's what the knowledge base says.' The individual is not committed to appropriate results, but to faithful application of procedures (Winograd 1990, p. 183).

Mark Coeckelbergh (2010) paints a potentially more disturbing picture. For him, the problem is not the advent of "artificial bureaucrats" but "psychopathic robots" (p. 236). The term "psychopathy" has traditionally been used to name a kind of personality disorder characterized by an abnormal lack of empathy which is masked by an ability to appear normal in most social situations. Functional morality, like that specified by Anderson and Anderson and Wallach and Allen, intentionally designs and produces what are arguably "artificial psychopaths"—robots that have no capacity for empathy but which follow rules and in doing so can appear to behave in morally appropriate ways. These psychopathic machines would, Coeckelbergh (2010) argues, "follow rules but act without fear, compassion, care, and love. This lack of emotion would render them non-moral agents—i.e. agents that follow rules without being moved by moral concerns—and they would even lack the capacity to discern what is of value. They would be morally blind." (p. 236)

Efforts in "machine ethics" (or whatever other nomenclature comes to be utilized to name this development) effectively seek to widen the circle of moral subjects to include what had been previously excluded and instrumentalized as mere neutral tools of human action. This is, it is important to note, not some blanket statement that would turn everything that was a tool into a moral subject. It is the recognition, following Marx, that not everything technological is reducible to a tool and that some devices—what Marx called "machines" and Langdon Winner (1997) calls "autonomous technology"—might need to be programmed in such a way as to behave reasonably and responsibly for the sake of respecting human individuals and communities. This proposal has the obvious advantage of responding to moral intuitions: if it is the machine that is making the decision and taking action in the world with little or no direct human oversight, it would only make sense to hold it accountable (or at least partially accountable) for the actions it deploys and to design it with some form of constraint in order to control for possible bad outcomes.

But doing so has considerable costs. Even if we bracket the questions of AGI, super intelligence, and machine consciousness; designing robotic systems that follow prescribed rules might provide the right kind of external behaviors but the motivations for doing so might be lacking. "Even if," Noel Sharkey (2012, p. 121) writes in a consideration of autonomous

weapons, “a robot was fully equipped with all the rules from the Laws of War, and had, by some mysterious means, a way of making the same discriminations as humans make, it could not be ethical in the same way as is an ethical human. Ask any judge what they think about blindly following rules and laws” (p. 121). Consequently, what we actually get from these efforts might be something very different from (and maybe even worse than) what we had hoped to achieve.

Conclusion

Drones and autonomous machines are not coming, they are already here. As Ronald Arkin (who wrote what many consider to be the agenda-setting textbook on the subject) has argued: “The trend is clear: Warfare will continue and autonomous robots will ultimately be deployed in its conduct” (Arkin, 2009, p. 29). The question—the critical question for all of us—is to decide how to respond to this development. And as one might anticipate, the range of possible responses extends across a rather broad spectrum bounded by two opposing positions. On the one side, there are international efforts to control or ban the development and use of autonomous weapons. In 2009, for instance, Jürgen Altmann, Peter Asaro, Noel Sharkey, and Rob Sparrow organized the International Committee for Robot Arms Control (ICRAC), calling “upon the international community for a legally binding treaty to prohibit the development, testing, production and use of autonomous weapon systems in all circumstances” (ICRAC, 2017). In 2013, ICRAC partnered with 63 other international and national NGOs from 28 countries on the Campaign to Stop Killer Robots (2017). The Campaign’s website explains the problem and the solution they advocate in the following way:

> Giving machines the power to decide who lives and dies on the battlefield is an unacceptable application of technology. Human control of any combat robot is essential to ensuring both humanitarian protection and effective legal control. The campaign seeks to prohibit taking a human out-of-the-loop with respect to targeting and attack decisions. A comprehensive, pre-emptive prohibition on the development, production and use of fully autonomous weapons—weapons that

> operate on their own without human intervention—is urgently needed. This could be achieved through an international treaty, as well as through national laws and other measures.

This proposal, and others like it (e.g. "Future of Life Institute," 2017), sound entirely reasonable. First, they follow a well-established precedent that has proven to be successful with restricting the development, production, and use of other kinds of lethal military technology. This is, for instance, the situation with the Chemical Weapons Convention (CWC)—a multilateral treaty that bans chemical weapons and requires their destruction within a specified period of time. Since its launch in April of 2013, the Campaign to Stop Killer Robots has petitioned and worked with the United Nations to develop similar international agreements that would do something like this for fully autonomous weapon systems.

Second, the Campaign's mission and efforts are legitimated by and seek to ensure the success of the instrumental theory of technology. In effect, the Campaign argues that advanced weapon systems, no matter how sophisticated their design or operations, must always be tethered to and remain under human control, and there should always be a human being in-the-loop who is able to take responsibility and to be held accountable for targeting and attack decisions. The Berlin Statement from ICRAC (2017) advances a similar instrumentalist position: "We believe that it is unacceptable for machines to control, determine, or decide upon the application of force or violence in conflict or war. In all cases where such a decision must be made, at least one human being must be held personally responsible and legally accountable for the decision and its foreseeable consequences."

On the other side of the issue is Ron Arkin and others, who, following the promise of machine ethics, argue that military robots—assuming that we design and program them properly—might be better at following the rules of military engagement than fallible human soldiers and therefore could make armed conflict more and not less humane. Among Arkin's reasons why autonomous robots "may be able to perform better than humans" in the "fog of war," are: (1) Robots do not need "to have self-preservation as a foremost drive" and therefore "can be used in a self-sacrificing manner if needed." (2) Machines can be equipped with better

sensors that exceed the limited capabilities of the human faculties. (3) “They can be designed without emotions that cloud their judgment or result in anger and frustration with ongoing battlefield events.” And (4) “they can integrate more information from more sources far faster before responding with lethal force than a human possible could in real-time” (Arkin, 2009, pp. 29–30). According to the argument that Arkin develops, autonomous robots offer the global community a “technological fix” to the unavoidable problems and complications that result from armed conflict.

In between these two extremes there are clearly a number of possible hybrid positions that try to split the difference and negotiate some kind of synthetic, middle ground. But like all hybrid solutions, this might sound good in theory—i.e. you do not need to choose sides—but the devil is in the details of its actual practice. In any event, the time to start thinking about these issues and developing possible solutions is now, before these devices are widely deployed and operational. As was remarked in the Future of Life Institute’s (2017) open letter to the UN, which was signed by 116 leaders in the AI/robotics field, “we do not have long to act. Once this Pandora’s Box is opened, it will be hard to close.” It is, therefore, not too soon to begin planning for and developing a response to the opportunities and challenges of autonomous military robots. And this response needs to come not just from technology experts and politicians; it must and needs to include involvement from all citizens who care about the current state of and future possibilities for armed conflict.

References

Allen, C. (2011). The future of moral machines. *The New York Times*, Retrieved December 25, from https://opinionator.blogs.nytimes.com/2011/12/25/the-future-of-moral-machines/

Anderson, M., & Anderson, S. L. (2007). The status of machine ethics: A report from the AAAI symposium. *Minds & Machines, 17*(1), 1–10.

Anderson, M., & Anderson, S. L. (2011). *Machine ethics*. Cambridge: Cambridge University Press.

Arkin, R. (2009). *Governing lethal behavior in autonomous robots*. Boca Raton, FL: Chapman & Hall/CRC Press.

Bowden, M. (2013). The killing machines: How to think about drones. *The Atlantic*. Retrieved from https://www.theatlantic.com/magazine/archive/2013/09/the-killing-machines-how-to-think-about-drones/309434/

Campaign to Stop Killer Robots. (2017). Retrieved from http://www.stopkillerrobots.org

Center for Civilians in Conflict and Human Rights Clinic. (2012). *The civilian impact of drones: Unexamined costs, unanswered questions*. Retrieved from https://civiliansinconflict.org/wp-content/uploads/2017/09/The_Civilian_Impact_of_Drones_w_cover.pdf

Coeckelbergh, M. (2010). Moral appearances: Emotions, robots, and human morality. *Ethics and Information Technology, 12*(3), 235–241.

Currier, C. (2015). The kill chain. *The intercept: The drone papers*. Retrieved from https://theintercept.com/drone-papers/the-kill-chain/

Derrida, J. (2005). *Paper machine* (R. Bowlby, Trans.). Stanford, CA: Stanford University Press.

Feenberg, A. (1991). *Critical theory of technology*. New York, NY: Oxford University Press.

Future of Life Institute. (2017). An open letter to the united nations convention on certain conventional weapons. Retrieved from https://futureoflife.org/autonomous-weapons-open-letter-2017

Gunkel, David J. (2018). *Robot rights*. Cambridge, MA: MIT Press.

Hall, J. S. (2001). Ethics for machines. *KurzweilAI.net*. Retrieved July 5, from http://www.kurzweilai.net/ethics-for-machines

Hanson, F. A. (2009). Beyond the skin bag: On the moral responsibility of extended agencies. *Ethics and Information Technology, 11*(1), 91–99.

Hemmersbaugh, P. A. (2016). NHTSA letter to Chris Urmson, Director, self-driving car project, Google, Inc. Retrieved from https://isearch.nhtsa.gov/files/Google-compiled response to 12 Nov 15 interp request-4 Feb 16 final.htm

ICRAC. (2017). International Committee for Robot Arms Control. Statements. Retrieved from https://www.icrac.net/statements/

Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. *Ethics and Information Technology, 8*(4), 195–204.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. New York: Oxford University Press.

McLuhan, M. (1995). *Understanding media: The extensions of man*. Cambridge, MA: MIT Press.

Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics, 2*(1), 25–42.

Sharkey, N. (2012). Killing made easy: From joysticks to politics. In K. Abney, P. Lin, & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robots* (pp. 111–128). Cambridge, MA: MIT Press.

Verbeek, P. P. (2011). *Moralizing technology: Understanding and designing the morality of things*. Chicago: University of Chicago Press.

Wallach, W. (2015). *A dangerous master: How to keep technology from slipping beyond our control*. New York: Basic Books

Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. Oxford: Oxford University Press.

Winner, L. (1977). *Autonomous technology: Technics-out-of-control as a theme in political thought*. Cambridge, MA: MIT Press.

Winograd. T. (1990). Thinking machines: Can there be? Are we? In D. Partridge & Y. Wilks (Eds.), *The foundations of artificial intelligence: A sourcebook* (pp. 167–189). Cambridge: Cambridge University Press.