The Human Role in Autonomous Weapon Design and Deployment

M.L. Cummings
Duke University

Abstract

There has been increasing debate in the international community as to whether it is morally and ethically permissible to use autonomous weapons, which are weapon systems that select and fire upon a target with no human in the loop. Given the tightly coupled link between emerging technology and policy development in this debate that speaks to the very core of humanity, this paper explains how current automated control systems, including weapons systems, are designed in terms of balancing authority between the human and the computer. The distinction between automated and autonomous systems is explained, and a framework is presented for conceptualizing the human-computer balance for future autonomous systems, both civilian and military. Lastly, specific technology and policy implications for weaponized autonomous systems is discussed.

Keywords: autonomous weapons, supervisory control, drones, unmanned aerial vehicles, killer robots

Introduction

Because of the increasing use of Unmanned Aerial Vehicles (UAVs, also commonly known as drones) in various military and para-military (i.e., CIA) settings, there has been increasing debate in the international community as to whether it is morally and ethically permissible to allow robots (flying or otherwise) the ability to decide when and where to take human life. In addition, there has been intense debate as to the legal aspects, particularly from a humanitarian law framework.

In response to this growing international debate, recently the United States government released the Department of Defense (DoD) 3000.09 Directive[1] that sets a clear policy for if and when autonomous weapons would be used in US military and para-military engagements. This US policy asserts that only "human-supervised autonomous weapon systems may be used to select and engage targets, with the exception of selecting humans as targets, for local defense …".

This statement implies that outside of defensive applications, autonomous weapons will not be allowed to independently select and then fire upon targets without explicit approval from a human supervising the autonomous weapon system (a control architecture known as human supervisory control). The defense caveat in this policy is needed because the United States currently uses highly automated systems for defensive purposes, e.g., Counter Rocket, Artillery, and Mortar (C-RAM) systems and Patriot anti-missile missiles.

Due to the time-critical nature of such environments (e.g., soldiers sleeping in barracks within easy reach of shoulder-launched missiles), these automated defensive systems cannot rely upon a human supervisor for permission because of the short engagement times and the inherent human neuromuscular lag which means that even if a person is paying attention, there is approximately a half second delay in hitting a firing button, which can mean the difference for life and death for the soldiers in the barracks.

So as of now, no US UAV (or any robot) will be able to launch any kind of weapon in an offensive environment without human direction and approval. However, the 3000.09 Directive does contain a clause that allows for this possibility in the future. This caveat states that the development of a weapon system that independently decides to launch a weapon is possible but first must be approved by the Under Secretary of Defense for Policy (USD(P)); the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD(AT&L)); and the Chairman of the Joint Chiefs of Staff.

While not all stakeholders are happy with this policy that leaves the door open for what used to be considered science fiction, it is worth noting that as of now, the United States is the only nation that has put forth such a written policy. There are many nations who are actively engaged in research attempting to develop advanced autonomous platforms, including lethal ones, yet other than the United States, none have made any formal public statements about intended future use and related policies.

Many opponents of such uses of technologies call for either an outright ban on autonomous weaponized systems, or in some cases, autonomous systems in general. Others will acquiesce that the growth of this technology is inevitable, but that we need to make sure we increase the level of human supervision of such systems. However, both approaches are somewhat naive in their understanding of the technology. Autonomy is not a discrete state, it is a continuum and versions have been in the US inventory for some time.

Present-day UAVs use the very same guidance, navigation and control technology flown on commercial aircraft. Tomahawk missiles, which have been in

[1] http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf

the US inventory for more than 30 years, are highly automated weapons that can navigate by themselves with no GPS and still maintain accuracies of less than a meter. Global Hawk UAVs can find their way home and land on their own without any human intervention in the case of a communication failure.

The growth of the civilian UAV market is also a critical consideration in the debate as to whether these technologies should be banned outright. There is an $82B industry emerging for the commercial use of drones in agricultural settings, cargo delivery, first response, commercial photography, and the entertainment industry (Association for Unmanned Vehicle Systems International (AUVSI) 2014). So it is an important distinction that UAVs are simply the platform for weapon delivery (autonomous or conventional), and that they have many peaceful and commercial uses independent of military applications.

Given that such advanced automation is pervasive across civilian and military technologies, this paper will explain how current supervisory control systems, including weapons systems, are designed in terms of balancing authority between the human and the computer. A framework will be presented for how to conceptualize the human-computer balance for future autonomous systems, both civilian and military, and the specific implications for weaponized autonomous systems will be discussed.

Balancing the role between a human and a computer

Human supervisory control (HSC) is the process by which a human operator intermittently interacts with a computer, receiving feedback from and providing commands, often remotely, to a system with varying degrees of embedded automation (Sheridan 1992)(Figure 1). Not only do all UAVs, both military and civilian, operate at some level of supervisory control as depicted in Figure 1, but so do nuclear power plants, automated trains, and commercial passenger planes.

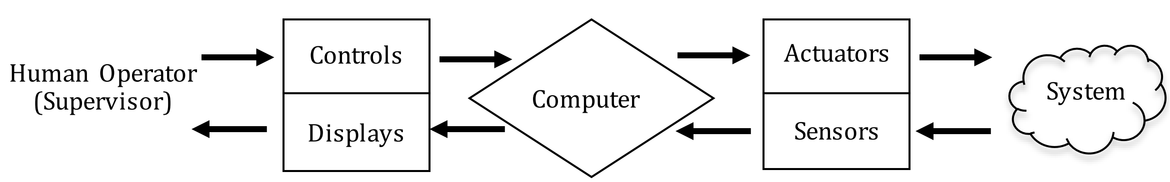

Figure 1: Human Supervisory Control (Sheridan 1992)

In developing any HSC system that involves the integration of human decision making in concert with an automated system, the question often arises as to where, when, and how much humans and automation should be in the decision making loop? Allocating roles and functions between the human and computer is critical in defining efficient and safe system architectures.

Getting this balance just right is not obvious for most system designers. The predominant engineering viewpoint across these systems is to automate as much as possible, and minimize the amount of human interaction. Indeed, many controls engineers see the human as a mere disturbance in the system that can and should be designed out. Others may begrudgingly recognize the human must play a role in such systems, either for regulatory requirements or low probability event intervention (such as problems in nuclear reactors).

But how do we know what is the right balance between humans and computers in these complex systems? Most engineers have little to no training in human supervision of complex systems and do not know how to address the inherent variability that accompanies all human performance. They desire a set of rules and criteria that reduce the ambiguity in the design space, which for them typically means reducing the role of the human or at least constraining human behavior in an attempt to control it. Yet this is exactly what opponents of autonomous systems fear the most – that the human will be designed out of the system, which could create a significant moral and ethical shift in authority.

A historical look at the role allocation debate

Human Factors engineers have been addressing the human-computer role allocation conundrum since the early 1950's when radar was just coming on line and there was much discussion as to how to design what is now our current national air traffic control system (Fitts 1951). In part to help traditional engineers understand the nuances of how humans could interact with a complex automated system in a decision-making capacity, Levels of Automation (LOAs) have been proposed.

LOAs generally refer to the role allocation between automation and the human, particularly in the analysis and decision phases of a simplified information processing model of information acquisition, analysis, decision, and action phases (Parasuraman 2000, Sheridan and Verplank 1978, Parasuraman, Sheridan, and Wickens 2000). Sheridan and Verplank (Sheridan and Verplank 1978) initially proposed that such LOAs could range from a fully manual system with no computer intervention to a fully automated system where the human is kept completely out of the loop, and this framework was later expanded to include ten LOAs (Table 1) (Parasuraman 2000).

Table 1: Levels of Automation (Sheridan and Verplank 1978, Parasuraman, Sheridan, and Wickens 2000)

Automation Level	Automation Description
1	The computer offers no assistance: human must take all decision and actions.
2	The computer offers a complete set of decision/action alternatives, or
3	narrows the selection down to a few, or
4	suggests one alternative, and
5	executes that suggestion if the human approves, or

6	allows the human a restricted time to veto before automatic execution, or
7	executes automatically, then necessarily informs humans, and
8	informs the human only if asked, or
9	informs the human only if it, the computer, decides to.
10	The computer decides everything and acts autonomously, ignoring the human.

For LOA scales like that in Table 1, at the lower levels the human is typically actively involved in the decision-making process. As the levels increase, the automation plays a more active role in decisions, increasingly removing the human from the decision-making loop. This scale addresses authority allocation, i.e., who has the authority to make the final decision. Other taxonomies have proposed alternate but similar LOAs, attempting to highlight less rigid and more dynamic allocation structures (Kaber et al. 2005, Endsley 1987) as well as address the ability to humans and computers to coach and guide one another (Riley 1989).

In terms of weapons systems today, in keeping with the DoD 3000.09 Directive, there are no US offensive weapons that operate above LOA 5. Whether fired from manned or unmanned aircraft (or from a ship as in the case of a Tomahawk missile), an operator in the supervisory control loop may use a computer to assist in the targeting process, but a human always tells the computer when to fire and what to hit. It is important to note though that the US (as well as other countries) do have defensive weapons that operate at LOAs 6 and above, as discussed previously. These higher levels are often called management by exception because the operator only intervenes in the firing process that is commenced by the computer. However, the only difference between an offensive or defensive automated weapon in the US inventory is the label we choose to give it and the human-driven policies that guide their deployment.

Automated vs. Autonomous Systems

Despite their seemingly advanced technological status, the weapon systems described previously, as well as drones today, are more automated than autonomous. This seemingly nuanced description is far from trivial and is critical for the debate about future lethal autonomous systems. An automated system is one that acts according to a pre-programmed script for a task with defined entry/exit conditions. Take, for example, a house thermostat. There is a set of predetermined rules that guide the system (i.e., turn heat on at 68 degrees and keep the temperature there). The sensors on the thermostat are highly reliable, with a simple feedback loop that looks for rises or falls in the temperature, and there is very little chance for error or failure. Automated weapons work in a similar fashion, e.g., home in on a set of GPS coordinates (or a laser designation), correct flight path when a disturbance is detected, and repeat until impact.

Autonomous systems represent a significant leap in complexity over automated systems primarily because of the role of probabilistic reasoning in such systems. An autonomous system is one that independently and dynamically determines if, when, and how to execute a task. Such a system will also contain many feedback loops but the script is not so clearly defined. In terms of the thermostat example, an autonomous thermostat is one that, for example, anticipates your arrival home and heats the house to your preference just prior to getting home. Such a system would need to guess at your arrival time, which could mean accessing traffic and weather reports as well as your smartphone calendar (indeed such systems are just now coming on the market now). These systems will likely be correct most of the time, but your calendar or traffic reports may not be updated to reflect reality, causing the system to either heat too early or late.

While such an example is benign, the same logic applies to autonomous weapons systems. They will have to independently assess and reason about the world with their own sensors and have to check databases much more complex than a smartphone. For example, an autonomous UAV that launches its own weapons in the future will need to determine if a suspected target is friendly or hostile, and whether the currently battle conditions meet the Rules of Engagement (ROE) for weapons release, and if so, what are the chances of collateral damage and a whole host of other possible issues.

Effectively such systems will ultimately have to make best guesses, just like humans do, particularly for target detection and identification, which remains a very difficult technological problem even by today's standards. It is this environment of having to make best guesses in the presence of significant uncertainty is what truly distinguishes an autonomous system from an automated one. And this difference is not discrete, but rather a continuum. The Global Hawk UAV, for example, is a highly automated UAV but has some autonomy on board that allows it to take itself home and land when it senses a loss of communication.

Given that we have limited autonomy in some (but not all) UAVs, it is only a matter of time that this will increase. However, the role allocation problem as discussed previously will only get harder with autonomous systems because of the critical role that uncertainty and the need for guessing plays. Because of the increasing need to reason probabilistically in these systems (called stochastic systems to distinguish them from the more rule-based deterministic automated systems), the LOA framework and others like it discussed previously provide little guidance in terms of balancing the human and computer role (Cummings and Bruni 2009, Defense Science Board 2012). To address this gap, in the next section I will discuss a new framework to think about role allocation in autonomous systems, which highlights some of the obstacles to future autonomous weapons system.

The Skills, Rules, Knowledge, & Expertise Framework for Autonomous Systems

Instead of thinking about role allocation as a function of whether the human or the computer is best suited for the task, Figure 2 depicts role allocation for both automated and autonomous systems, based on what kind of reasoning is needed for that task, independent of who (the human and/or the computer). This depiction is an extension of Rasmussen's SRK (skills, rules, and knowledge-based behaviors) taxonomy (Rasmussen 1983).

In this taxonomy, skill-based behaviors are sensory-motor actions that are highly automatic, typically acquired after some period of training (Rasmussen 1983). Indeed, Rasmussen says, "motor output is a response to the observation of an error signal representing the difference between the actual state and the intended state in a time-space environment (p. 259)."

In Figure 2, an example of skill-based control for humans is the act of flying an aircraft. Student pilots spend the bulk of their training learning to scan dials and gauges so that they can instantly recognize the state of an aircraft and adjust if the intended state is not the same as the actual state (which is the error signal controls engineers are attempting to minimize.) Once this set of skills is acquired, pilots can then turn their attention (which is a scarce resource, particularly under high workload), to higher cognitive tasks.

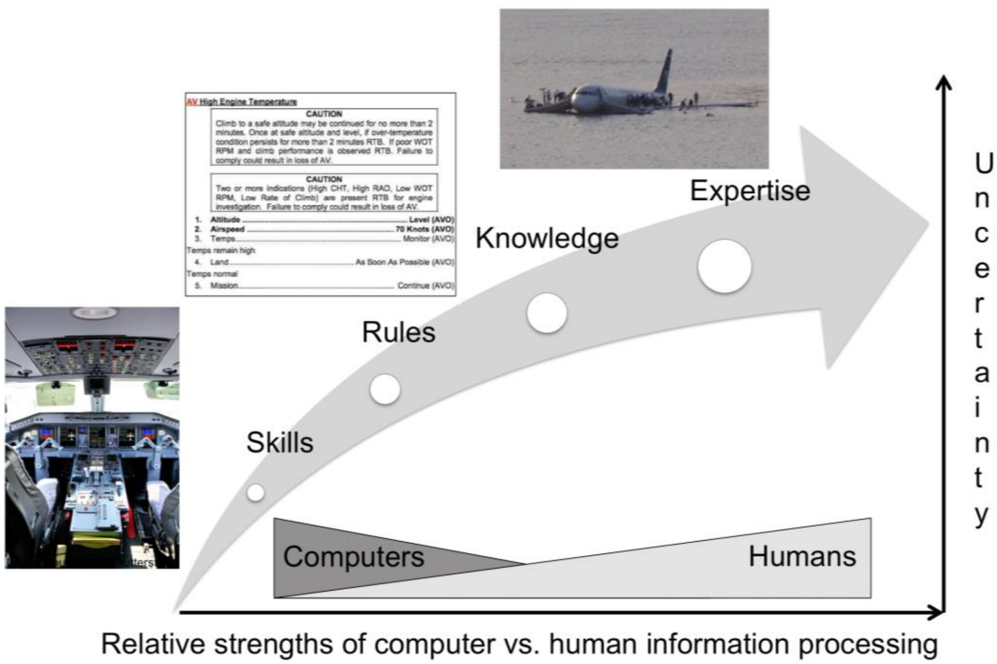

Figure 2: Role Allocation for Information Processing Behaviors (Skill, Rule, Knowledge, Expertise) and the Relationship to Uncertainty

Up the cognitive continuum in Figure 2 are rule-based behaviors, which are effectively those actions guided by subroutines, stored rules, or procedures. Rasmussen likens rule-based behavior to following a cookbook recipe (Rasmussen

1983) (p. 261). Difficulties for humans in rule-based environments often come from recognizing the correct goal in order to select the correct procedure or set of rules.

In Figure 2, in keeping with the aviation piloting example, pilots spend significant amounts of time learning to follow procedures. So for example, when a fire light illuminates or some other sub-system indicates a problem, pilots recognize that they should consult a manual to determine the correct procedure (since there are far too many procedures to be committed to memory), and then follow the steps to completion. Some interpretation is required, particularly for multiple system problems, which is common during a catastrophic failure like the loss of thrust in one engine. And as mentioned previously, recognizing which procedure to follow is not always obvious, particularly in warning systems where one aural alert can indicate different failure modes.

For Rasmussen, the highest level of cognitive control is that of knowledge-based behaviors, where mental models built over time aid in the formulation and selection of plans for an explicit goal (Rasmussen 1983). The landing of USAIR 1549 in 2009 in the Hudson River, as depicted in Figure 2, is an example of a knowledge-based behavior in that the Captain had to decide whether to ditch the aircraft or attempt to land it at a nearby airport. Given his mental model, the environment, and the state of the aircraft, his very fast mental simulation made him choose the ditching option, with clear success.

I added a 4th behavior to the SRK taxonomy, that of expertise, to demonstrate that knowledge-based behaviors are a prerequisite for gaining expertise in a particular field, and this cannot be achieved without significant experience. Moreover, judgment and intuition, concepts that often make traditional engineers uncomfortable since they lack a mathematical formal representation, are the key behaviors that allow experts to quickly assess a situation in a fast and frugal method (Gigerenzer, Todd, and Group 1999), without necessarily and laboriously comparing all possible plan outcomes.

The last, but certainly not the least important, addition to the SRK taxonomy is my representation of uncertainty in the arrow that increases in size as behaviors move from skill-based to expertise. Uncertainty occurs when a situation cannot precisely be determined, with potentially many unknown variables in a system that itself can be highly variable.

As discussed previously, reasoning under uncertainty is a hallmark characteristic of an autonomous system, and can arise from exogenous sources such as the environment, i.e., birds in the general vicinity of an airport that *might*, on rare occasion, be ingested in an engine. However, uncertainty can also be introduced from endogenous sources, either from human behaviors but also from computer/automation behaviors. As evidenced by the Air France 447 crash in 2009 where the pitot-static system gave erroneous information to the pilots due to icing, sensors can degrade or outright fail, introducing possibly unknown uncertainty into a situation.

Automation and Skill-Based Tasks

When considering role allocation between humans and computers, it is useful to consider who or what can perform the skill, rule, knowledge, and expertise-based behaviors required for a given objective and associated set of tasks. For many skill-based tasks like flying an aircraft, automation in general easily out performs humans. By flying, I mean the act of keeping the aircraft on heading, altitude, and airspeed, i.e., keeping the plane in balanced flight on a stable trajectory.

Ever since the introduction of autopilots and more recently, digital fly-by-wire control, computers are far more capable of keeping planes in stable flight for much longer periods of times than if flown manually by humans. Vigilance research is quite clear in this regard, in that it is very difficult for humans to sustain focused attention for more than 20-30 minutes (Warm, Dember, and Hancock 1996), and it is precisely sustained attention that is needed for flying, particularly for long duration flights.

There are other domains where the superiority of automated skill-based control is evident, such as autonomous trucks in mining industries. These trucks are designed to shuttle between pickup and drop off points and can operate 24/7 in all weather conditions since they are not hampered by reduced vision at night and in bad weather. These trucks are so predictable in their operations that some uncertainty has to be programmed into them or else they repeatedly drive over the same tracks, creating ruts in the road that make it difficult for manned vehicles to negotiate.

For many domains and tasks, automation is superior in skill-based tasks because, given Rasmussen's earlier definition, such tasks are reduced to motor memory with a clear feedback loop to correct errors between a desired outcome and the observed state of the world. In flying and driving, the bulk of the work is a set of motor responses that become routine and nearly effortless with practice. The automaticity that humans can achieve in such tasks can, and arguably should, be replaced with automation, especially given human limitations like vigilance, fatigue, and the neuromuscular lag (Jagacinski and Flach 2003).

The possibility of automating skill-based behaviors (and as we will later see, all behaviors) depends on the ability of the automation to sense the environment, which for a human happens typically through sight, hearing and touch. This is not trivial for computers, but for aircraft, through the use of accelerometers and gyroscopes, inertial and satellite navigation systems, and engine sensors, the computer can use it sensors to determine with far greater precision and reliability, whether the plane is in stable flight and how to correct in microseconds if there is an anomaly.

This capability is why military and commercial planes have been landing themselves for years far more precisely and smoothly than humans. The act of landing requires the precise control of many dynamic variables, which the computer can do repeatedly without any influence from a lack of sleep or reduced visibility. The same is true for cars that can parallel park by themselves.

However, as previously mentioned, the ability to automate a skill-based task is highly dependent on the ability of the sensors to sense the environment and make adjustments accordingly, correcting for error as it arises. For many skill-based tasks, like driving, vision (both foveal and peripheral) is critical for correct environment assessment. Unfortunately, computer vision still lags far behind human capabilities in many respects, although there is significant research underway in this area. Ultimately this means that for a skill-based task to be a good candidate for automation, uncertainty should be low and sensor reliability high, which is difficult for computer vision applications.

Thus, drones are very effective platforms because basic flight control, a skill-based task, is ideally suited for automation. However, driverless cars will not be so easy to automate because they require significant reliance on computer vision, which is not yet reliable under the much more variable operational environments as compared to flying.

Rule-Based Tasks and Autonomy

As depicted in Figure 2, skill-based behaviors and tasks can be the easiest to automate, since by definition they are highly rehearsed and automatic behaviors with inherent feedback loops. Rule-based behaviors for human, however, require higher levels of cognition since interpretation must occur to determine that given some stimulus, a set of rules or procedures must be applied to attain the desired goal state.

By the very nature of their if-then-else structures, rule-based behaviors are also potentially good candidates for automation but again, uncertainty management is key. Significant aspects of process control plants, including nuclear reactors, are highly automated because the rules for making changes are well established and based on first principles, with highly reliable sensors that accurately represent the state of the physical plant.

Path planning is also very rule-based in that given rules about traffic flow (either in the air or on the road), the most efficient path can be constructed. However, uncertainty in such domains makes it a less ideal candidate for automation. When an automated path planner is given a start and end goal, for the most part the route generated is the best path in terms of least time (if that is the goal of the operator). However, many possibilities exist that automation may not have information about that cause such a path to be either suboptimal or even infeasible, such as in the case of accidents or bad weather.

While fast and able to handle complex computation far better than humans, computer optimization algorithms, which work primarily at the rule-based level, are notoriously brittle in that they can only take into account those quantifiable variables identified in the design stages that were deemed to be critical (Smith, McCoy, and Layton 1997). In complex systems with inherent uncertainties (weather impacts, enemy movement, etc.), it is not possible to include a priori every single variable that could impact the final solution.

Moreover, it is not clear exactly what characterizes an optimal solution in such uncertain scenarios. Often, in these domains, the need to generate an optimal solution should be weighed against a satisficing (Simon et al. 1986) solution. Because constraints and variables are often dynamic in complex environments, the definition of optimal is also a constantly changing concept. In those cases of time pressure, having a solution that is good enough, robust, and quickly reached is often preferable to one that requires complex computation and extended periods of times, which may not be accurate due to incorrect assumptions.

Another problem for automation of rule-based behaviors is similar to one for humans, which is the selection of the right rule or procedure for a given set of stimuli. Computers will reliably execute a procedure more consistently than any human, but the assumption is that the computer selects the correct procedure, which is highly dependent on the sensing aspect.

It is at the rule-based level of reasoning where we start to see the shift between needing automated versus autonomous behaviors. Some higher-level reasoning starts at this level, but the uncertainty also starts to grow as well, especially in the presence of an incomplete rule set. The Global Hawk UAV works at a rule-based level when it is able to land itself when it loses communication. However, it is not yet been demonstrated that such aircraft can reason under all situations it might encounter, which would require a higher level of reasoning, discussed next.

Knowledge-Based Tasks & Expertise

The most advanced form of cognitive reasoning occurs in domains where knowledge-based behaviors and expertise are required. These settings are also typically where uncertainty is highest, as depicted in Figure 2. While rules may assist decision makers (whether human or computer) in aspects of knowledge-based decisions, such situations are by definition vague and ambiguous such that a mathematically optimal or satisficing solution is not available. Weapons release from any platform, manned or unmanned, requires knowledge-based reasoning, and autonomous weapon systems will have to be able to achieve this level of reasoning before they can be safely deployed.

It is precisely in these situations where the human power of induction is critical. Judgment and intuition are critical in these situations, as these are the weapons needed to combat uncertainty. Because of the aforementioned brittleness problems in the programming of computer algorithms and the inability to replicate the intangible concept of intuition, knowledge-based reasoning and true expertise, for now, are outside the realm of computers. However, there is currently significant research underway to change this, particularly in the machine learning (sometimes called artificial intelligence) community, but progress is slow.

IBM's Watson, 90 servers each with a 3.5 GHz core processor (Deedrick 2011), is often touted as a computer with knowledge-based reasoning, but people confuse the ability of a computer to search vast databases to generate formulaic responses with knowledge. For Watson, which leverages natural language processing and pattern matching through machine learning, uncertainty is low. Indeed, because Watson leverages statistical reasoning, it can bound answers with confidence intervals.

This example highlights the probabilistic nature of knowledge-based reasoning. Whether humans or computers do it, both are guessing with incomplete information based on prior probabilities about an outcome. While the consequences for an autonomous thermostat guessing wrong about your arrival time are relatively trivial, the same cannot be said for autonomous weapons.

Another major limitation of such approaches, and an inherent limitation of much machine learning research, is the brittleness aspect in that it is not clear whether a computer can detect a pattern or event it has never seen before, or that is slightly different than a pattern it has seen before. In a recent major 'breakthrough', a machine learning algorithm was able to successfully recognize cats in images with 15.8% accuracy, which was reported to be an improvement of 70% over the current state- of-the-art (Le et al. 2012). This is critical when considering the viability of autonomous weapon systems. UAVs are excellent platforms for obtaining an image but letting the UAV decide whether a target is in that image with a high degree of certainty is still several years away.

Conclusion

There is no question that UAVs/drones will become commonplace in both military and commercial settings. Indeed, robots of all shapes, sizes, and capabilities will become part of our everyday landscape. But as these systems start to grow in numbers and complexity, it will be critical for engineers and policy makers to address the role allocation issue. To this end, this paper presented a taxonomy for understanding what behaviors can be automated (skill-based), what behaviors can be autonomous (rule-based), and where humans should be leveraged, particularly in cases where inductive reasoning is needed and uncertainty is high (knowledge-based). It should be noted that these behaviors do not occur in discrete stages with clear thresholds, but rather are on a continuum.

Because computers cannot yet achieve knowledge-based reasoning, especially for the task of target detection and identification where uncertainty is present, autonomous weapons simply are not achievable with any guarantees of reliability. Of course, this technological obstacle may not stop other nations and terrorist states from attempting to build such systems, which is why it is critical that policy makers understand the clear technological gap between what is desired and what is achievable.

This raises the question of technical competence for policy makers who must approve the use of autonomous weapons. The United States has designated the Under Secretary of Defense for Policy, the Under Secretary of Defense for Acquisition, Technology, and Logistics, and the Chairman of the Joint Chiefs of Staff as decision makers with the authority to approve autonomous weapons launches. Such systems will be highly sophisticated with incredibly advanced levels of probabilistic reasoning never before seen in weapon systems. It has been well established that humans are not effective decision makers when faced with even simple probabilistic information (Tversky and Kahneman 1974). So this begs the question whether these four individuals, or any person overseeing such complex systems who is not a statistician, will be able effectively judge whether the benefit of launching an autonomous weapon platform is worth the risk.

And whether we *can* build such weapons should be followed by the discussion of whether we *should*. Since the invention of the long bow, soldiers have been trying to increase the distance for killing, and UAVs in their current form are simply another technological advance along this continuum. However, autonomous weapons represent an entirely new dimension where a computer, imbued with probabilistic reasoning codified by humans with incomplete information, must make life and death decisions with even more incomplete information in a time critical setting. As many have discussed (Cummings 2004, Human Rights Watch 2012, Anderson and Waxman 2013, International Committee for the Red Cross 2014), autonomous weapons raise issues of accountability as well as moral and ethical agency, and the technical issues outlined here further highlight the need to continue this debate.

References

Anderson, K., and M. Waxman. 2013. Why a Ban Won't Work and How the Laws of War Can. Stanford, CA Hoover Institution.

Association for Unmanned Vehicle Systems International (AUVSI). 2014. *The Economic Impact of Unmanned Aircraft Systems Integration in the United States.* . AUVSI 2014 [cited June 2 2014]. Available from http://www.auvsi.org/econreport.

Cummings, M. L., and S. Bruni. 2009. "Collaborative Human Computer Decision Making." In *Handbook of Automation*, edited by Shimon Nof. New York: Springer.

Cummings, M.L. 2004. "Creating Moral Buffers in Weapon Control Interface Design." *IEEE Technology and Society* no. 23 (3):28-33.

Deedrick, Tami. 2013. *It's Technical, Dear Watson*. IBM Systems Magazine 2011 [cited November 6 2013]. Available from http://www.ibmsystemsmag.com/ibmi/trends/whatsnew/It%E2%80%99s-Technical,-Dear-Watson/

Defense Science Board. 2012. The Role of Autonomy in DoD Systems. Department of Defense.

Endsley, M. 1987. The application of human factors to the development of expert systems for advanced cockpits. Paper read at Human Factors Society 31st Annual Meeting, at Santa Monica, CA.

Fitts, P.M. 1951. Human Engineering for an Effective Air Navigation and Traffic Control system. Washington, DC: National Research Council.

Gigerenzer, Gerd, Peter M. Todd, and The ABC Research Group. 1999. *Simple Heuristics That Make Us Smart*. Edited by Stephen Stich, *Simple Heuristics That Make Us Smart*. Oxford, NY: Oxford University Press, Inc.

Human Rights Watch. 2012. Arms: New Campaign to Stop Killer Robots. Human Rights Watch

International Committee for the Red Cross. 2014. Report of the ICRC Expert Meeting on 'Autonomous weapon systems: technical, military, legal, and humanitarian aspects' 26-28 March 2014. Geneva, Switzerand: Red Cross.

Jagacinski, R. J., and J. M. Flach. 2003. *Control Theory for Humans: Quantitative Approaches to Modeling Performance*. New Jersey: Lawrence Erlbaum Associates, Publishers.

Kaber, D.B., M.C. Wright, L.J. Printzel III, and M.P. Clamann. 2005. "Adaptive Automation of Human-Machine System Information-Processing Functions." *Human Factors* no. 47 (4):730-741.

Le, Quoc V., Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeffrey Dean, and Andrew Y. Ng. 2012. Building High-level Features Using Large Scale Unsupervised Learning. Paper read at Twenty-Ninth International Conference on Machine Learning, at Edinburgh, Scotland.

Parasuraman, Raja. 2000. "Designing automation for human use: empirical studies and quantitative models." *Ergonomics* no. 43 (7):931-951.

Parasuraman, Raja, Thomas B. Sheridan, and Chris D. Wickens. 2000. "A Model for Types and Levels of Human Interaction with Automation." *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* no. 30 (3):286-297.

Rasmussen, J. 1983. "Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models." *IEEE Transactions on Systems, Man, and Cybernetics* no. 13 (3):257-266.

Riley, V. 1989. A General Model of Mixed-Initiative Human-Machine Systems. Paper read at Human Factors Society 33rd Annual Meeting, at Denver, CO.

Sheridan, Thomas B. 1992. *Telerobotics, Automation and Human Supervisory Control*. Cambridge, MA: The MIT Press.

Sheridan, Thomas B., and W. Verplank. 1978. Human and Computer Control of Undersea Teleoperators. Cambridge, MA: Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT.

Simon, Herbert A., R. Hogarth, C.R. Piott, H. Raiffa, K.A. Schelling, R. Thaier, A. Tversky, and S. Winter. 1986. Decision Making and Problem Solving. In *Report of the Research Briefing Panel on Decision Making and Problem Solving*, edited by National Academy of Sciences. Washington D.C.: National Academy Press,.

Smith, P.J., C.E. McCoy, and C. Layton. 1997. Brittleness in the design of cooperative problem-solving systems: The effects on user performance. Paper read at IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans.

Tversky, A., and D. Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* no. 185 (4157):1124-1131.

Warm, J.S., W.N. Dember, and P.A. Hancock. 1996. "Vigilance and Workload in Automated Systems." In *Automation and Human Performance: Theory and Applications*, edited by R. Parasuraman and M. Mouloua. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.