Draft of chapter to appear in "Autonomous Weapons Systems: Law, Ethics, Policy", edited by Nehal Bhuta, Susanne Beck, Robin Geiss, Claus Kress and Hin Yan Liu.

The human control of weapons: a humanitarian perspective[1]

Noel Sharkey
University of Sheffield, UK

Abstract

There is an ongoing technological transformation in warfare with ever more control of weapons being delegated to computer systems. There is considerable international concern among states and civil society about where humans fit into the control loop. Rather than move to a point where computer programs control the weapons, it is proposed here that hi-tech nations should set the specific goal of developing weapons that enable zero civilian casualties and facilitate combatant surrender. With the right balance between the best of human abilities and the best of computer functionality we could progress towards this goal and have significantly greater humanitarian effect. The psychological literature on human decision-making provides a foundation for the type of control required for weapons. A human control classification is provided that reframes autonomy/semi-autonomy in terms of levels of supervisory control. This allows for greater transparency in command and control and the allocation of responsibility.

There is considerable and increasing international discussion and debate about whether or not we should allow the decision to kill a human to be delegated to autonomous weapons systems; systems that, once activated, can track, identify and attack targets with violent force without further human intervention. The discussion has ranged from moral and legal[2] to technical and operational[3] to issues about international security[4].

It seems clear that for the foreseeable future[5], we cannot guarantee that fully autonomous weapons systems will be able to comply with International Humanitarian Law (IHL). Apart from problems with the principles of distinction and proportionality in determining the legitimacy of targets, autonomous weapons systems are, by definition, less predictable than other weapons systems. In addition,

[1] Some portions of this paper have previously appeared in Noel Sharkey, Towards a principle for the human supervisory control of robot weapons, *Politica & Società*, 2/2014, 305-324.

[2] See P. Asaro, On Banning Autonomous Weapon Systems: Human Rights, Automation and the Dehumanisation of Lethal Decision-Making, *International Review of the Red Cross*, n. 94 (2012), pp. 687-709; and C. Heyns, Report of the special Rapporteur on extrajudicial summary or arbitrary executions», to the *Human Rights Council Twenty-third session* (2013).

[3] see: N. Sharkey, The Evitability of Autonomous Robot Warfare», *International Review of the Red Cross*, n. 94 (2012), pp. 787-799; and N. Sharkey, Saying – No! to Lethal Autonomous Targeting», *Journal of Military Ethics*, n. 4 (2010), 9, pp. 299-313.

[4] Concerns have been expressed that unknown combating algorithms controlling autonomous weapons would interact in unpredictable ways. This could make it impossible for weapons reviews to guarantee compliance with IHL. N. Sharkey, The Automation and Proliferation of Military Drones and the Protection of Civilians», *Journal of Law, Innovation and Technology,* n. 3 (2011), 2 pp. 229-240.

[5] In the context of this paper, forseeable future means that it follows from an analysis of the current state of the technology, the ongoing research projects and the current empirical evidence from the technology. Any departure from a forseeable future analysis is dependent on speculation about the future without clear supporting evidence.

the US Department of Defense has pointed out a number of computer problems for the use of autonomous weapons systems.[6]

Some argue that such weapons could be used legally in certain very limited circumstances while others argue that at some point in the future, they may be able to fully comply with IHL. But these arguments are about an IHL compliant technology that no one yet knows how to create. There is nothing wrong with technological ambitions or a general research agenda in civilian domains, but there is less room for such conjecture when discussing autonomous technologies of violence.

For example, robot soccer is seen as a great research challenge and a chance to test robotics technology within a real world application. The ultimate aim is to develop a team of autonomous humanoid robots that will beat human world champions by 2050. No one knows if this will work but the challenge enables the development of new methods of robot control and sensing that can be applied elsewhere.[7] So success in the ultimate aim is not vital to reap technological benefits. If the enterprise fails, we may invent a different kind of sport for humans and robots to play together (and still keep the old sport specifically for humans) with new rules of engagement to give robots an equal chance of victory.

In contrast, if our thinking, our strategies and our funding are directed towards developing autonomous weapons systems and it turns out to be that making them IHL compliant is unsuccessful, what will we do with that weapons technology? What if we get involved in serious conflicts? We may then have to change what IHL compliance means and modify the rules of engagement to give the new weapon a place. This is what happened in the past with aerial bombardment and submarine warfare.

The limitations of the technology is partly why technologically capable states such as the UK and the US have made it clear that there will be a human in the loop for lethality decisions. In the UK, , the Parliamentary Under Secretary of State, Lord Astor of Hever, said: "the MoD [Ministry of Defence] currently has no intention of developing systems that operate without human intervention". And: "let us be absolutely clear that the operation of weapons systems will always be under human control".[8]

When the US Department of Defence (DoD) issued the first policy document on autonomous weapons, they stated: "Autonomous and semi-autonomous weapons systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force".[9]

[6] US Department of Defense, *Autonomy in Weapon Systems*, Directive 3000.09, November 21 2012 p14 points to potential problems with autonomous weapons: human error, human-machine interaction failures, malfunctions, communications degradation, software coding errors, enemy cyber attacks, infiltration into the industrial supply chain, jamming, spoofing, decoys, other enemy countermeasures or actions, unanticipated situations on the battlefield.

[7] For a fuller discussion see Edoardo Datteri, Guglielmo Tamburrini (2013) Robotic weapons and democratic decision-making, In Hilgendorf E., Guenther J-P (Eds) Robotik und Gesetzegebung

[8] 26 March 2013 cf http://bit.ly/1lZMQyW _14. Last accessed June 19 2014.

[9] Op cit note note 4

What has not been made absolutely clear from the UK is exactly what type of human control will be employed. Nor has the US DoD made any attempt to define "appropriate levels of human judgment." Without addressing these points, and they are not easy to address, there is no transparency in the operation of such computerized weapons. To say that there is a human in the control loop does not clarify the degree of human involvement. It could simply mean a human programming a weapons system for a mission or pressing a button to activate it or it could mean exercising full human judgment about the legitimacy of a target before initiating an attack.

Moreover, the terms 'autonomous' and 'semi-autonomous' weapons do not help to clarify the control issue. For example the US Navy used three levels while the US Army used ten.[10] This could be very confusing for a military commander having to work with several systems at different levels. We end up with long technological discussions about what are the levels of semi-autonomy and what they mean in terms of computing and robotics.

The US DoD science board Task Force agree: "The Task Force reviewed many of the DoD-funded studies on "levels of autonomy" and concluded that they are not particularly helpful to the autonomy design process." They recommended that, 'the DoD abandon the use of "levels of autonomy'." This is because, "they focus too much attention on the computer rather than on the collaboration between the computer and its operator/supervisor to achieve the desired capabilities and effects."[11]

An alternative approach to the classification of autonomy and semi-autonomy is to turn the discussion on its head and reframe autonomy in terms of the type and quality of human control afforded by different types of computerized weapons systems. An examination of scientific research on human supervisory control allows us to develop a classification consisting of five levels of control.[12] Table 1 lists five levels for the human supervisory control of weapons.[13]

Table 1: A classification for levels of human supervisory control of weapons

[10] For some discussion and references see N. Sharkey, «Cassandra or the False Prophet of Doom: AI Robots and War», *IEEE Intelligent Systems*, n. 4, (2008), 23, pp. 14-17.

[11] Task Force Report, 'The Role of Autonomy in DoD Systems', Department of Defense – Defense Science Board, July 2012, p. 48, available at: http://www.fas.org/irp/agency/dod/dsb/autonomy.pdf

[12] This is adapted from early work on general (non-military) supervised control with 10 levels of human supervisory control: T.B. Sheridan and W. Verplank, «Human and Computer Control of Undersea Teleoperators», *Man-Machine Systems Laboratory*, Department of Mechanical Engineering, MIT, Cambridge MA 1978.

[13] These levels should not be considered as absolute; they are a work in progress and are under discussion with military advisors.

1. human engages with and selects target and initiates any attack
2. program suggests alternative targets and human chooses which to attack
3. program selects target and human must approve before attack
4. program selects target and human has restricted time to veto
5. program selects target and initiates attack without human involvement

If used correctly, such a classification affords transparency about who is in control and when. It also moves away from technical jargon to plain language.

The strengths and limitations of each of the methods depend on what weapon is being controlled and more importantly how much it inhibits the possibility of human reasoning. There is no point in designing an interface to a weapons system that does not take human capabilities into account.

The control of weapons mediated by computer programs requires the human and machine to operate together in a way that should optimize the strengths of both. Computers are better and more efficient at some task than humans while humans are better at other tasks. Examples are provided in Table 2.

Table 2: examples of tasks computers and humans are better at[14]

Computers	**Humans**
calculate numbers	Deliberative reasoning
search large datasets	Perceive novel patterns
respond quickly to control tasks	meta-cognition
perform repetitive routine tasks	reasoning inductively
deductive reasoning	applying diverse experience to novel tasks
sort data	exercise meaningful judgment

The ideal partnership

We are embarking down a path where human control of weapons is being increasingly ceded to computer programs. It is vital that we take the opportunity to use the evolution of technology to ensure that the partnership between human and machine increases rather than diminishes the ability of humans to ensure the legitimacy of the targets of attack. Rather than making more and more hi-tech weapons with the aim of more effective killing and destruction of targets, would it not be better to create hi-tech weapons with the aim of having greater humanitarian impact?

[14] See also M.L. Cummings, Automation Bias in Intelligent Time Critical Decisions Support Systems, *American Institute of Aeronautics and Astronautics*, AIAA 3rd Intelligent Systems Conference Chicago, 2004.

Some may argue that making more precise weapons with greater accuracy is in effect enabling less collateral damage. But this is at best a 'side effect' to the goal of more effective killing with less ammunition. It is proposed here that greater humanitarian impact should be the explicit goal of new technological weapons; with greater technology should come greater responsibility towards civilians, civilian infrastructure and those hors de combat. The ideal would be to reduce collateral damage to zero and facilitate better opportunities for combatants to surrender.

The question then is, what might such humanitarian impact look like? With the right configuration of human and machine partnership we would ideally expect a commander (or operator) to be able to perform as suggested in the following list:

1. have full contextual and situational awareness of the target area at the time of initiating a specific attack

2. be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack such as changes in the legitimacy of the targets.

3. have active cognitive participation in the attack

4. have sufficient time for deliberation on the nature of targets, their significance in terms of the necessity and appropriateness of an attack, and likely incidental and possible accidental effects of the attack.

5. have a means for the rapid suspension or abortion of the attack.

Sharre[15] has correctly suggested that this list could rule out a large number conventional weapon currently in use. But the point here is not to look back at 'old' weapons but to look forward to upgrade our sensibility to civilian harm as a result of technological developments. The current operating procedure is that the commander has *sufficient* situational awareness of the target area at the *time of planning an attack* rather than *full* contextual and situational awareness at the *time of the attack* as stated in point 1 of the list.

However, if one of the reasons for using advanced technology to apply violent force is genuinely to reduce or eliminate harm to civilians, others *hors de combat* and civilian infrastructure, then striving for full contextual and situation awareness at the time of attack is a way forward. It does not necessarily require a change in the law of armed conflict. It requires a change in the acceptability and responsibility standards for compliance with the law.

[15] P. Sharre. Autonomy, "killer robots" and Human Control in the use of force, Part II, Just Security July 2014 http://justsecurity.org/12712/autonomy-killer-robots-human-control-force-part-ii/

These statements run the risk of condoning the use of distance killing such as carried out by drone (remotely piloted aerial vehicle) strikes. That is not the intention here and that should be clear from the outset. There have been dubious uses of drones that are at best legally questionable.

What is being suggested is a perilous path fraught with dangers of misuse of the technology to expand the battlefield to areas outside of official conflict zones with the excuse of more humanitarian killing. It would require a considerable tightening of the current legal regime to ensure greater accountability and compliance with existing laws. The stated aim should be to raise the bar on humanitarian standards in conflict rather than to make entering into conflicts easier.[16]

If modern technology is employed for supervisory control of weapons, it should be possible for commanders to have active participation during attacks rather than simply in planning them. Even with a cruise missile, it should be possible to use advanced camera systems to view targets and ensure that their legitimacy has not changed since launch.[17] This means satisfying point 2 in the above list providing that there is an inbuilt facility to rapidly suspend or abort the attack (point 5).

None of this is easy in the delicate human-computer balancing act. It is vital that we use human reasoning at its best when we consider targeting with violent force. When humans fail at human-computer tasks, it can simply mean that they are being asked to perform in a mode of operation that is not well suited to human psychology. This needs to be part of the equation of ensuring efficient and meaningful human supervisory control of weapons. If we get the balancing act right military objectives could be met while at the same time having better and more predictable compliance with International Humanitarian Law. Getting it wrong could result in considerable humanitarian problems.

Getting reasoning right

One of the most well established distinctions in human psychology is between *automatic* and *controlled* processes. It follows from more than 100 years of research on dual processing (1890).[18] Automatic processing refers to fast responses that are always cued automatically such as those required for playing tennis or riding a bicycle.

[16] We should also be very wary of attempts to take such technological developments into the civilian world without due consideration of the potential for violations of human rights including the right to life, privacy and dignity and the right to peaceful protest.

[17] Note that there is no intention here to legitimate the use of armed current s in warfare. These have been deployed into conflict with insufficient foresight and regulation about how they might be best be used. And the use by the intelligence services is, at best, legally questionable.

[18] W. James, *The Principles of Psychology*, vol. 1, Holt, New York 1890. There is debate in the psychological literature about the underlying brain mechanisms (e.g. W. Schneider and J.M. Chen, Controlled and Automatic Processing: Behavior, Theory and Biological Mechanisms, *Cognitive Science*, n. 27 (2003), pp. 525-559.and whether it is possible to create a unified model of the dual processes (e.g. S.B.T. Evans and K.E. Stanovich, Dual-Process Theories of Higher Cognition: Advancing the Debate», *Perspectives on Psychological Science*, n. 8 (2013), 3, pp. 223-241. However the distinction

Controlled refers to slower deliberative processes that we need in order to make a thoughtful decision such as determining the winner of a competition or judging a murder trial.

Daniel Kahneman, winner of the Nobel Memorial prize for his work on reasoning, used the dual process theory as an explanation of human decision making.[19] The terms automatic and deliberative processes will be used here for consistency and clarity.

The *deliberative* processes always come into play after the automatic and are thus slower: Kahneman calls them "lazy processes". They will go along with the automatic processes unless there is something surprising or irregular and/or we are operating in novel circumstances or performing tasks that require vigilance and/or deliberation.

A downside of deliberative processes is that they require attention and free memory space. If a distraction takes our attention away or requires memory resources, automatic processes dominate In fact anything that impacts on memory capacity or attention such as stress or being pressured to make a quick decision could incapacitate deliberative reasoning.

The upside of automatic processes is that they do not require active control or attention. Normally both systems operate seamless together and we don't even notice their interplay. However, the distinction should become clear by working through the following example devised by Kaheman:[20]

Example of automatic v deliberative processing

Task. 1: go down each column in turn and whisper to yourself lower or upper case for each word. Task 2: Repeat the exercise but this time whisper whether each word is to the right or left of its column.	
LEFT left right RIGHT RIGHT left LEFT right	upper lower LOWER upper UPPER lower LOWER upper

This example illustrates the distinction between automatic and deliberative processes by showing them in conflict with one another. The task requires deliberative reasoning because whispering upper/lower or right/left is unusual when reading columns of

[19] D. Kahneman, *Thinking, Fast and Slow*, Penguin Books, London 2011. He refers to the two processes as System 1 and System 2, these are exactly the same as the terms automatic and deliberative used here for clarity and consistency.
[20] *Ibid*.

words. But you will find that one column was significantly easier than the other and the easy column was different for both tasks. This is because we cannot help but automatically read the actual words and this interferes with the deliberative processes.

The relevance to weapons control is that both types of reasoning have different advantages and disadvantages. The advantage of automatic decision processes is that they can be trained through repetition and practice on routine tasks. They are needed for fast reaction in sports, and for riding a bicycle, driving a car or in military routines. In fact, automaticity is used anytime for routine decisions that have to be made rapidly for *predictable* events (and the word 'predictable' is highly important here). For automatic processes to work well, it is vital to have an environment that contains useful cues that, via practice, have been (over) rehearsed.

For the right tasks, automatic reasoning can be a better option than deliberative reasoning and it is not inherently bad. When initiated by well-practiced cues, it reduces much of the tedium in our lives and saves us from a life of indecision.

Members of the armed forces rehearse and are over-trained for many routine tasks that require automatic action on order. Fast automatic response can be trained with well practiced cues. These can be useful in military contexts such as when someone shouts "fire in the hole" – a warning that should prompt those hearing it to immediately take cover. The question to ask about automatic reasoning is, does a given domain afford enough regularity to be learnable as an automatic process? When it comes to human supervised targeting, the unpredictable and unanticipated circumstances in a dynamically changing environment play to the weakness of automatic reasoning.

Four of the properties of automatic reasoning from Kahneman *ibid* illustrate how it would be problematic for the supervisory control of weapons. Automatic reasoning:

1. **Neglects ambiguity and suppresses doubt**: automatic processes are all about jumping to conclusions. They are guided by experience. An unambiguous answer pops up immediately and does not allow doubt. Automatic reasoning does not search for alternative interpretations and does not examine uncertainty. So if something looks like it might be a legitimate target in ambiguous circumstances, automatic reasoning will be certain that it is legitimate.

2. **Infers and invents causes and intentions:** automatic reasoning is adept at finding a coherent causal story to link together fragments of available information. Events including people (or even inanimate objects such as robots) are automatically attributed with intentions that fit the causal story. For example, if a human operator is seeking out patterns of behaviour to

determine a lethal drone strike, then seeing people load bales of hay or shovels onto a truck could initiate a causal story that they were loading rifles for an attack. This relates to *assimilation bias* in the human supervisory control literature.[21]

3. **Is biased to believe and confirm:** the operation of automatic reasoning has been shown to favour the uncritical acceptance of suggestions and maintains a strong bias. Thus if a computer system suggests a target to an operator, automatic reasoning alone would make it highly likely that it would be accepted. This is known as automation bias in the supervisory literature.[22] When people seek out information to confirm a prior belief, this is confirmation bias.[23]

4. **Focuses on existing evidence and ignores absent evidence:** automatic reasoning builds a coherent explanatory story without considering any evidence or contextual information that might be missing. This is why Kahneman uses the term WYSIATI or What you see is all there is. It facilitates the feeling of coherence that makes us confident to accept information as true whether it is or not. This is a problem if a more detailed analysis of the context of a target showed that it was not in fact legitimate. For example, an ununiformed man firing a rifle in the vicinity of an army platoon may be deemed to be a hostile target with WYSIATI. But some deliberation and a quick scan around might reveal that he had actually just killed a wolf that had taken one of his goats.

What these properties of automatic reasoning show is that in the context of supervised control of lethal targeting, things could go badly wrong. It may work well for many instances and seem OK but not when there is contradictory information of target legitimacy. Contradictory evidence could remain unseen or be disbelieved. Doubt and uncertainty will be suppressed as will any notion that there is more evidence that cannot be seen.

In normal operation both automatic and deliberative processes operate smoothly together. The point here is that it is vitally important that deliberative reasoning is enabled in the design of supervisory control for weapons systems. Although this is also subject to error and flaws, it does as good a job as can be done with uncertainty and

[21] J.M. Carroll and M.B. Rosson, «Paradox of the Active User», in J.M. Carroll (eds.), *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*, MIT Press, Cambridge MA 1987, pp. 80-111.
[22] K.L. Mosier and L.J. Skitka, «Human Decision Makers and Automated Decision Aids: Made for Each Other?», in M. Mouloua (eds.), *Automation and Human Performance: Theory and Applications*, Lawrence Erlbaum Associates, Inc. Mahwah NJ 1996, pp. 201-220.
[23] C.G. Lord, L. Ross and M. Lepper, «Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence», *The Journal of Personality and Social Psychology*, n. 47 (1979), pp. 1231-1243.

doubt.

If a supervisory weapons operator is distracted by another task or if they are stressed, their attentional capacity may be low. Many experimental studies have demonstrated that a small amount of interference to our attention or memory can disable the deliberative system. For example, when people are asked to do a task, such as verifying the truth or falsity of a statement, and at the same time add numbers, they will believe anything they are told.

In such circumstances weapon supervisors trying to decide on the necessity or legitimacy of an attack, may not be reasoning at an acceptable level. This is one reason why, in what is known as on-the-loop control, having a single operator controlling multiple weapons systems could be disastrous. They would not be able to use their deliberative reasoning and could simply catch the downsides of automatic reasoning if there were problems or irregularities.

3. Deliberative reasoning meets supervisory control of weapons=

Having now discussed some of the relevant processes of human reasoning, we return to consider how they relate to the human supervisory weapons control framework introduced in Table 1 in the introduction of the chapter. [24] It is understood that some of the requirement proposed here are difficult to carry out in a military context but that should not stop them being attempted to as much as conceivable. They represent ideals that should be genuinely aspired to.

Level 1: *a human deliberates about a target before initiating any attack*. Here it is critically important to towards the ideal of adhering to the strict requirements of *deliberative human control* as described in the previous section: a human commander (or operator) must have full contextual and situational awareness of the target area at the time of a specific attack and be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack. There must be active cognitive participation in the attack and sufficient time for deliberation on the nature of the target, its significance in terms of the necessity and appropriateness of attack, and likely incidental and possible accidental effects of the attack. There must also be a means for the rapid suspension or abortion of the attack.

Level 2: *a computer program provides a list of targets and a human chooses which to attack.* This type of control could be acceptable if shown to meet the requirement of deliberative human control. A human in control of the attack would have to be in a position to assess whether an attack is necessary and appropriate, whether all (or indeed any) of the suggested alternatives are permissible objects of attack, and to

[24] There is a potential sub-class of defensive autonomous weapons that may lie outside of this framework. These are weapons that Sense and React to Military Objects (SARMO weapons), cf Noel Sharkey, Towards a principle for the human supervisory control of robot weapons, *Politica & Società*, 2/2014, 305-324.

select the target which may be expected to cause the least civilian harm. This requires deliberative reasoning. Without sufficient time or in a distracting environment the illegitimacy of a target could be overlooked.

A rank ordered list of targets is particularly problematic as there would be a tendency to accept the top ranked target unless sufficient time and attentional space is given for deliberative reasoning.

Level 3: *a program selects target and human must approve before attack is unacceptable.* This type of control has been experimentally shown to create what is known as *automation bias* in which human operators come to accept computer generated solutions as correct and disregard or don't search for contradictory information.

Cummings (2004) conducted a study on an interface designed for supervision and resource allocation of in-flight GPS guided Tomahawk missile.[25] The task for operators was to decide which candidate missile from pool of 8-16 would be the correct one to redirect to a time-critical emergent target.

The impact on the speed and accuracy of decision making was tested for two different methods to redirect Tomahawks in real time.

1. The computer provided the operator with ranked recommendations including the most "optimal" missile given the situation. This is type 3 in the above classification.
2. The computer filtered all those missile that were not candidates because of physical restraints (not enough fuel etc). The operator had to decide which missile. This equates to type 2 in the above classification.

Cummings also manipulated the computer recommendations so that half the time they were correct and half the time they were wrong. The result was that Type 3 operators made significantly faster decisions overall and their accuracy was equal to the slower Type 2 operators when the computer recommendations were correct. However, when the computer recommendation were wrong, the Type 3 operators had a significantly decreased accuracy. This is known as *automation bias*: operators are prepared to accept the computer recommendations without seeking any disconfirming evidence.

Level 4: *a computer program selects target and the human has restricted time to veto*. This is unacceptable because it does not promote target identification. Providing only a short time to veto would reinforce automation bias and leave no room for doubt or deliberation. As the attack will take place *unless* a human intervenes, this undermines well-established presumptions under international humanitarian law that promote civilian protection.

The time pressure will result in operators falling foul of all four of the downsides of

[25] M.L. Cummings, «Automation Bias», cit.

automatic reasoning described above: neglects ambiguity and suppresses doubt, infers and invents causes and intentions, is biased to believe and confirm, focuses on existing evidence and ignores absent evidence. An example of the errors caused by fast veto came in the 2004 war with Iraq when the U.S. Army's Patriot missile system engaged in fratricide, shooting down a British Tornado and an American F/A-18, killing three pilots.[26]

Level 5: *a computer program selects targets and initiates attacks without human involvement.* As argued above, such weapons systems could not comply with international law except in very narrowly bounded circumstances and is thus entirely unacceptable.

This classification of levels of human control is just a beginning. We need to map out exactly the role that the human commander/supervisor plays for each supervised weapons systems. Research is urgently needed to ensure that human supervisory interfaces make provisions to get the best level of human reasoning needed to comply with the laws of war in all circumstances.

4. Conclusions

There has been general agreement on the inadequacy of autonomous weapons systems to comply with international humanitarian law into the foreseeable future. While there seems to be a big push by some states to develop weapons that could be used autonomously, states such as the US and UK have made it clear that, at least for the time being, computerized weapons systems will always be under human control. However, it has not been made clear about what exactly is meant by human control and how meaningful it will be. Thus one aim of this article was to pull apart and examine the ideal conditions for the control of computerized weapons by humans.

Both humans and computer systems have their strengths and weaknesses and the aim of designing effective supervisory control systems for weapons control is to exploit the strengths of both. In this way it should be possible not only to gain better legal compliance, but also to ensure that the partnership between human and machine is more humanitarian than machines operating alone. Rather than making more hi-tech weapons with the aim of more effective killing and destruction of targets, it would better serve humanity to create hi-tech weapons with the explicit aim of eliminating civilian casualties.

To do this we must urgently begin to develop a principle for the meaningful human control of weapons that is founded on an understanding of the process of human reasoning. Reframing autonomy in terms of human control will eliminate some of the technical complexity of engineering jargon and make it clear who is in control, where

[26] M.L. Cummings, «Automation and Accountability in Decision Support System Interface Design», *Journal of Technology Studies*, vol. 32 (2006), pp. 23-31.

and when. This will not only make the control of weapons transparent for state weapons reviews and make the control clear to commanders, it could clarify who is responsible for mishaps and potential crimes.

The strict requirement for Level 1 control of weapons as specified in Section 3 may seem overly idealistic to some. But it could be achievable if there was a will to push technological developments in that direction. It is certainly more achievable than creating autonomous weapons that could comply with International Humanitarian Law. If states are set to continue fighting wars and initiating conflicts the least they can do is to aspire to the goal of creating weapons that will produce zero civilian casualties.