15th ICCRTS
"The Evolution of C2"

C^2 Design for Ethical Agency over Killing in War

Topic: C^2 Approaches and Organization

Patrick Hew
Defence Science and Technology Organisation

Point of Contact:

Patrick Hew
Defence Science and Technology Organisation
Canberra ACT 2600
Australia
Patrick.Hew@dsto.defence.gov.au

Abstract: The 1980s and 1990s saw substantial efforts in automated target recognition, "smart" and "brilliant" munitions, and related technologies. Research then slowed, on the ethical concern of accountability. Concerns have returned with the rapid fielding of unmanned systems post-2001, and the spectre of "killer robots".

This paper develops a C^2 design for assigning responsibility for killing in war. The key is to integrate supervisory control with the theory of intelligent agents. Supervisory control is where a machine closes a control loop, and a supervisor intermittently programs the machine. It is informally known as "on the loop", versus the human being "in" the control loop. The result is an engineering definition for the *ethical agent*, responsible for the consequences of a *lethal agent*.

An ethical agent is characterized by its ability to conduct supervisory control over itself, a capability unique to humans for the foreseeable future. Ethical killing in war thus requires a human "on" the firing loop; further, it is neither necessary nor sufficient for a human to be "in" the loop. The distinction between "on the loop" and "in the loop" roles should, therefore, be central to C^2 and combat systems design.

Introduction

Western ethics on warfare require that someone be held responsible for the deaths that occur [1]. This applies to enemies killed under Just War theory, the prosecution of war crimes in the event of non-combatant casualties, or investigations into so-called friendly-fire incidents. The current, implied axiom is that the "someone" is a human being. This axiom invites re-examination from a systems engineering perspective. What are the properties of a human being that enable them to be held responsible? Could the critical activities be distributed across humans and machines? Or, if there are duties that must be assigned to a human being, how does the overall system support the human in this capacity?

This paper establishes some necessary conditions for an *ethical agent*, one that can be held responsible for killing in war. The key contribution is to establish the conditions in solution-independent terms, as a basis for C^2 design. In particular, we shall see that an *ethical agent* must have the capacity to conduct supervisory control over itself. This capacity is unique to humans at this time, and cannot (yet) be implemented in machines. Hence, for now and into the foreseeable future, ethical killing in war requires a human to be "on the loop"; further, it is neither necessary nor sufficient for a human to be "in" the firing loop. The distinction between "on the loop" and "in the loop" roles should thus be central to C^2 and combat systems design.

The question is significant to C^2 thinking in two ways. The first is in reinvigorating automated target recognition, "smart" and "brilliant munitions" and related technologies. The 1980s and 1990s saw substantial efforts [2-4], on an expectation of high potential benefit [5].[1] Research slowed at the turn of millennium; a key reason being a perceived gap in ethical accountability [6]. However, current concepts make exorbitant manpower demands, for humans "in the loop" to process, exploit and disseminate the raw information into tailored intelligence [7].[2] As interest in automated target recognition and tracking renews [8], the need for appropriate C^2 approaches and organisation becomes increasingly acute.

Secondly, the explosion in unmanned systems post-2001 has prompted debate about the ethics of so-called "killer robots" [9, 10]. There is currently much confusion as to whether "killer robots" are ethically equivalent to "brilliant munitions", or represent something new [11, 12]. Clarifying this situation, particularly with respect to C^2 approaches and organisation, is necessary if Western military forces are to leverage the potential of robotics and automation.

[1] In 1992, the Strategic Technologies for the Army Report (STAR) identified robot vehicles (air or ground) for C^3I/RISTA missions and brilliant munitions for attacking ground targets as two of six advanced system concepts having particularly high-potential benefits for the US Army (C^3I/RISTA = Command, Control, Communications, Computing, Information / Reconnaissance, Intelligence, Surveillance and Target Acquisition). Computer Science, Robotics, and Artificial Intelligence was one of eight Technology Focus Areas forecast to have advances that could be fielded into Army systems by 2020.

[2] The United States Air Force estimates that it will have to recruit, train and support an additional 2500 intelligence, surveillance and reconnaissance airmen.

A Model for Killing

A successful model needs to integrate the technical and philosophical perspectives of killing in war. We do so by extracting the nodes and activities, and studying whether and how they can be assigned to humans or machines (if at all). The critical constructs are of lethal agents, supervisory control and then self-supervising agents.

Lethal Agents

In general usage, an *agent* is defined as someone or something that acts or has the power to act [13]. In Artificial Intelligence research, an *intelligent agent* is an autonomous entity that observes and acts upon an environment (it is an agent) and directs its activity towards achieving goals (it is rational) [14]. There are no restrictions on an agent's construction (mechanical, electronic, biological, software …), nor whether it is unitary or a networked assemblage of components, nor whether it is mobile or stationary. Hence, for example, an unmanned aircraft system is best regarded as a collection of agents, each assembled from components (human or artificial) housed by the airframe, on the ground, or elsewhere.

To emphasise, an intelligent agent is characterized by its closing a loop from sensors to effectors. Our particular interest is in *lethal agents*, agents that close a firing loop from sensor to weapon. That is, a lethal agent integrates the following three subsystems: (Figure 1):

1. A collection of *weapons*,
2. A collection of *sensors*,
3. A *decision* subsystem that uses the sensors to find targets of interest, and selects targets for prosecution with a weapon.

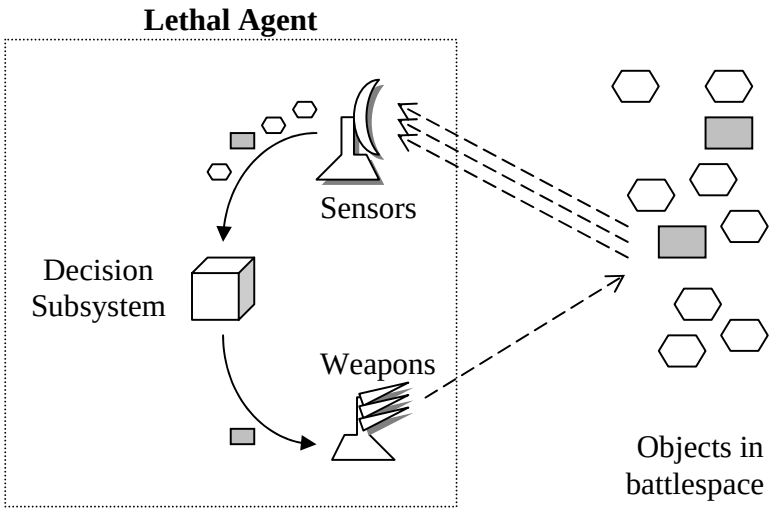

Figure 1: Lethal Agent – Finds and Prosecutes Targets.

A soldier with a rifle is a lethal agent. So too is an Improvised Explosive Device (IED), albeit a far less sophisticated one.[3] For comparison, Table 1 gives examples of rules used in current, real-world systems.

Table 1: Target Discrimination Rules in Current Lethal Agents.

Lethal Agent	Target Discrimination Rule
Aegis Air-Warfare Combat System (Auto Special)	Target has a radar signature matching «…» and is on a trajectory of «…».
Phalanx / Centurion Close-In Weapon System	Target has a radar and infrared signature matching «…» and is on a trajectory of «…».
SGR-1A (Sentry Robot on the Korean Demilitarised Zone)	Target is within «…» and has a colour optical signature matching «…».
Captor Mine	Target has an acoustic signature matching «…» and is on a trajectory of «…».
Improvised Explosive Device – Pressure Plate	Target is directly above me and has a weight exceeding «…».
Improvised Explosive Device – Passive Infrared	Target is within my line of sight and has an infrared signature exceeding «…».

Just War theory and the Laws of Armed Conflict require that lethal agents be proportionate and discriminate [9, 11]. There is, however, no requirement for them to be perfect, merely that they be used as militarily necessary and with minimum human suffering [12, 15]. Historically, the locus of responsibility has been assigned to a human being [1], but as described earlier, we seek to understand the reasoning, in terms amenable to systems engineering [16]. We do so by looking at the structures that deploy and supervise a lethal agent.

Supervisory Control and Autonomy

The definition of an artificial intelligent agent used the word "autonomous", a term which we have yet to define. We do so through the notion of *supervisory control*. *Supervisory control* is where one or more operators are intermittently programming and receiving information from an artificial intelligent agent [17].[4] We can thus quantify autonomy as the time between the operator providing supervision to the agent, ranging from zero to infinite as autonomy increases from low to high (Figure 2). Informally, when autonomy is low, the supervisor is "in the loop", while "on the loop" is high autonomy. Here the

[3] Specifically, an IED would be modelled as a "simple reflex agent", as it follows a simple "if-then" rule. Russell and Norvig define a range of classes for artificial intelligent agents, to recognise more sophisticated capabilities and "intelligence".

[4] Sheridan's definition of supervisory control has "one or more human operators are intermittently programming and continually receiving information from a computer that itself closes an autonomous control loop through artificial effectors to the controlled process or task environment." The proposed revision deletes the requirement for operators to be "human" and "continually" receiving information, and matches in the definition of an artificial intelligent agent.

"loop" refers to the agent's operations, so being "in the loop" corresponds to continuous supervision, while "on the loop" is more intermittent [18].[5]

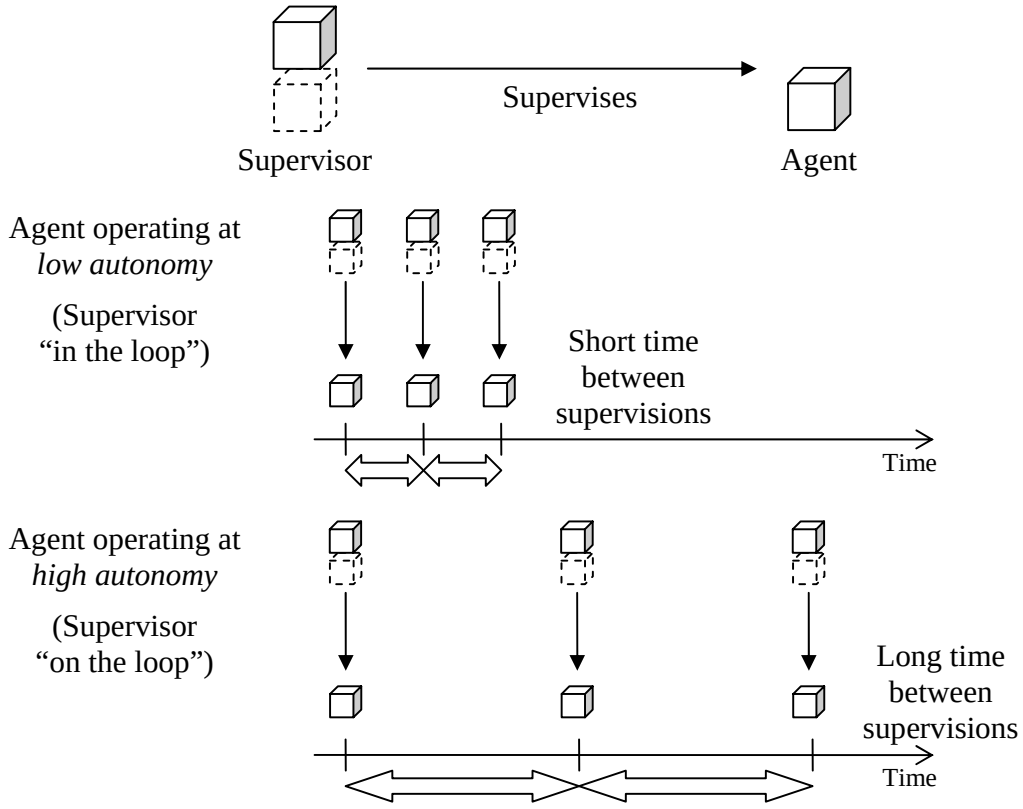

Figure 2: Autonomy of an Agent – Measured as Time between Supervisions.

The technical definition of autonomy appears to have no place for "intentionality", "freedom" or "free will", concepts important to the philosophies of action and agency [9]. We will address this by accepting the technical definition here, and revisiting the philosophers' perspective in due course.

In defining supervisory control, we implicitly defined a *supervisor* (short for *supervising agent*). A *supervisor* is an agent that has supervisory control over a subordinate agent(s); it will intermittently reprogram its subordinates, using information that it has gathered from the environment or taken from the subordinate agents. As for all agents, a supervisor can be human or artificial, without restriction. Similarly, in line with the earlier definition of supervisory control, a supervisor's subordinates can be operating on a spectrum of autonomy, ranging from zero to infinite.

We can then turn our attention to the supervisor of a lethal agent (Figure 3). That is, a lethal agent has some capacity to find and prosecute targets, under some mitigating

[5] The USAF Unmanned Aircraft Systems Flight Plan 2009-2047 mentions "Man on the loop" synonymously with supervisory control on p14. Section 4.6 puts it thus: 'Increasingly humans will no longer be "in the loop" but rather "on the loop" – monitoring the execution of certain decisions.'

conditions and up to some level of performance. A supervisor could be assigned to measure the lethal agent's performance, conduct its own battlespace appreciation, and make changes to the lethal agent in light of these inputs.

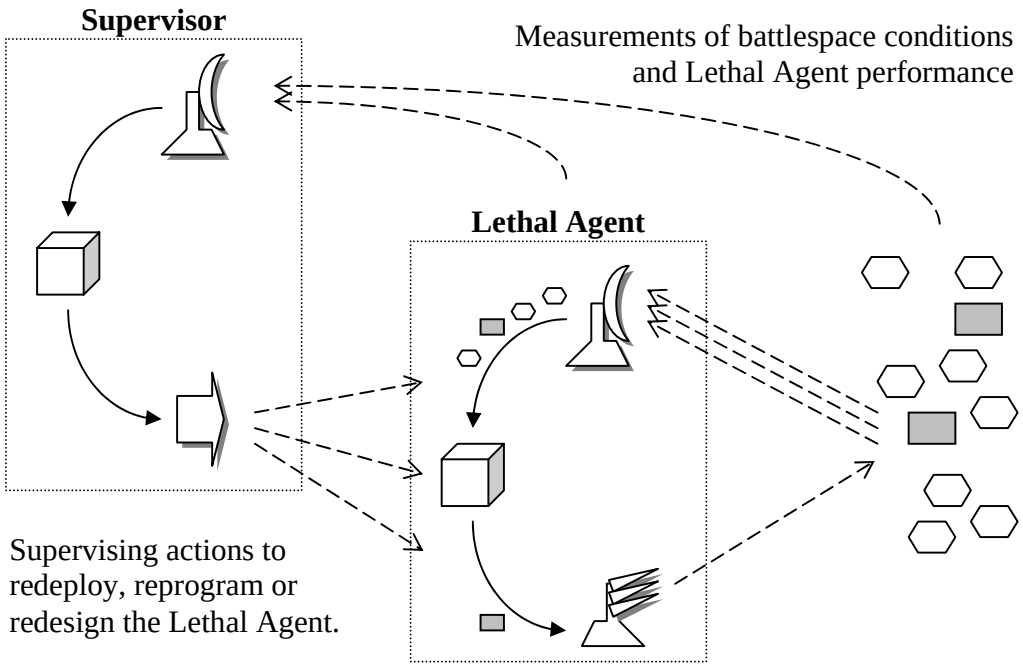

Figure 3: Supervisor and Lethal Agent.

Supervision Chains and Self-Supervising Agents

The performance of a lethal agent is, in some sense, a function of both the lethal agent and that of its supervisor. We are thus interested in the *supervision chain*, consisting of supervisors to the lethal agent (Figure 4). Each supervisor is in supervisory control of its subordinates, with the subordinates operating at some autonomy. As we ascend through the chain, the supervisors are less concerned with the lethal agent's performance, and more concerned with the performance of the supervisors.[6] Cyberneticists would recognize this as a reformulation of the Viable System Model [19]. At each level in the supervision chain, the agent will scan the environment, audit the system under supervision, and adapt it to maintain viability.

[6] Suppose we had a lethal agent that distinguished targets based on their colour. We could envisage a supervisor that held a database of known target types, and that periodically extracted the "recognition colour" to be used for each target. It would then be the supervisor's supervisor's job to decide which targets were entered into the database.

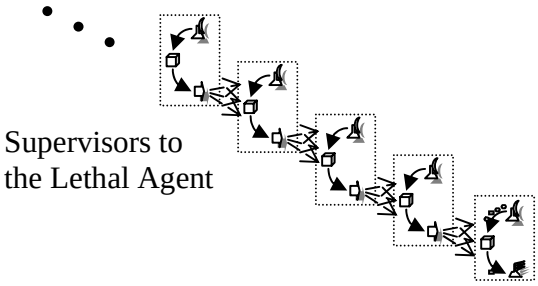

Figure 4: Supervision Chain, supervising a Lethal Agent.

In the real world, supervision chains do not continue forever. For our abstract analysis, we terminate our supervision chains with a *self-supervising agent*. A self-supervising agent has the capacity to conduct supervisory control on itself (Figure 5). We admit the possibility that self-supervising agent might be decomposed into a closed structure of agents, mutually supervising each other. The idea of self-supervising agents is related to the concept of self-organising systems [20-22],[7] self-regulating adaptive systems and meta-adaptivity [23-25].[8] However, the term "self-supervising" is more precise, and consistent with the definition of supervisory control. Similarly, while a philosopher might say that a "self-supervising agent" is actually "autonomous", we use autonomy and self-supervising as defined here and above, for technical clarity.

[7] A system that can perform "Supervisory control on itself" is also a "self-organising system", but the converse does not hold. The term "self-organising system" has, unfortunately, become ambiguous on this very point. De Wolf and Holvoet supply a working definition, "Self-organisation is a dynamical and adaptive process where systems acquire and maintain structure themselves, without external control." They then give the following example of a self-organising system: "Plugging in a PnP device in a computer can be considered as normal data input. A self-organising behaviour could be the autonomous configuration of drivers by the computer system. If a user has to install the drivers himself then there is no self-organisation." The ambiguity here is that the computer's operating system embeds an algorithm that specifies the driver for the PnP device. On point, the operating system is installed by the user, and the act of installing the operating system constitutes external control.
Put alternately, there have been proposals to engineer self-organising systems towards some desired global system behaviour. If this is the case, then the "self-organising system" is actually under supervisory control, towards that behaviour. The intervals between supervisory control may be huge, and the interventions tiny, but control is nonetheless being exercised. In contrast, "supervisory control on itself" requires that the desired global system behaviour is intrinsic to the "self-" of "self-organising".

[8] "Self-regulating adaptive systems" and "meta-adaptive systems" have a capacity to adapt, and a capacity to assess and modify this adaptive behaviour. That is, instead of adapting under some fixed logic, the logic of adaptation can evolve over time. However, this logic of adaptation is itself fixed. "Self-regulating adaptive systems" and "meta-adaptive systems" thus correspond to the first and second levels of supervision immediately above the lethal agent.

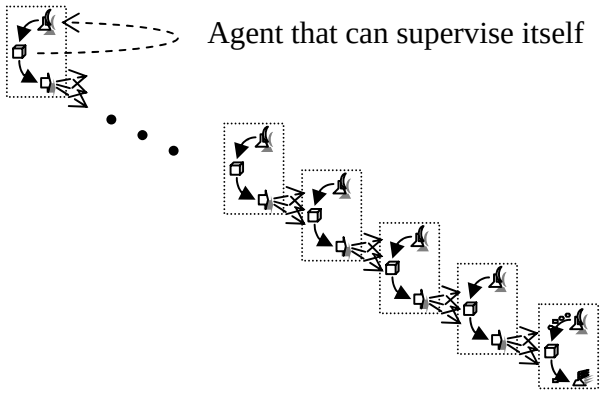

Figure 5: Supervision Chain is terminated by a Self-Supervising Agent.

It will be necessary to be able to talk in terms of a *tempo* of supervision. Tempo is defined as the reciprocal of *autonomy*. That is, if an agent is operating at high autonomy, then the supervisor is operating at a low tempo; conversely if the agent is operating a low autonomy, then the supervisor is operating at a high tempo. The tempo of supervision cascades down the supervision chain – the more frequent a supervisor provides intervention, the more frequent the subordinates may have to act, setting up the cascade. The overall tempo of a supervision chain is thus set by the self-supervising agent at the top; subordinates might work at a higher tempo, but not (in general) at a lower tempo.

In general, a lethal agent may have multiple supervision chains (Figure 6). For instance, we might have a supervision chain that periodically upgrades the sensors, another one for the weapons, and a third chain that assembles the lethal agent from these components. The chains may have different depths, and the supervisors could apply their supervision for low to high autonomy. However, as before, each chain will be capped by a self-supervising agent.

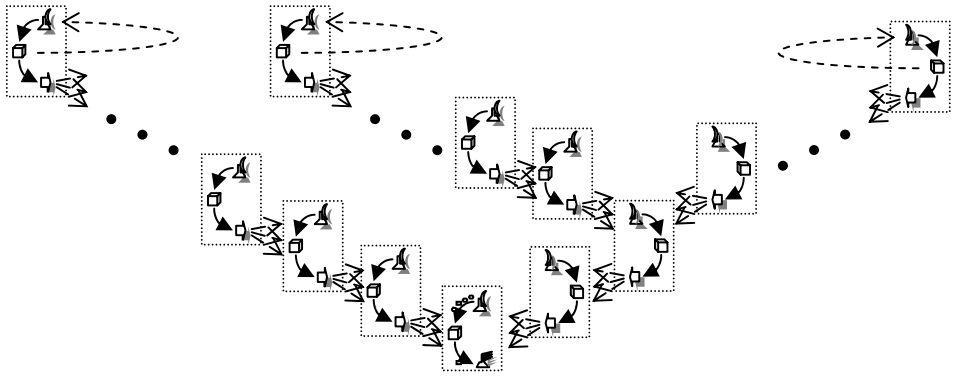

Figure 6: Lethal Agent may have multiple Supervision Chains (at different tempos).

Finally, the supervision chains need not be static. Indeed, the act of modifying a supervision chain can itself be regarded as an instance of supervisory control. We

especially note that a supervision chain might grow or shrink over time, notably from inserting or removal a self-supervising agent as a new capstone (Figure 7).

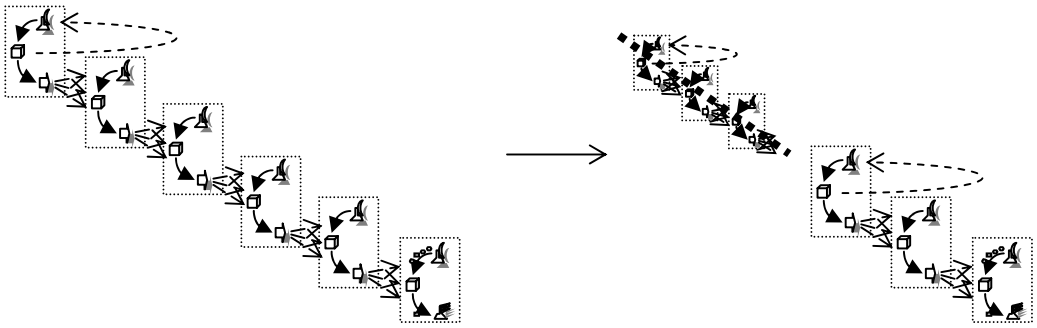

Figure 7: Supervision Chains can evolve over time.

To summarise, any lethal agent is associated with one or more supervision chains, each capstoned by a self-supervising agent. A real-world system might implement these agents as machines, or that assign agency to skilled personnel. The elements might all be packaged into one physical unit, or distributed over a network. With this technical perspective of killing in war, we can turn to the philosophical view.

Engineering for Ethical Agency

Proposed Definition – The Ethical Agent

We are now in a position to assign responsibility for the deaths that occur in war. The following definition is proposed: that the *ethical agent associated with a lethal agent* is identified as the self-supervising agent capstoning the supervision chain with the fastest tempo. There are two parts to this definition:

1. **Capstones a supervision chain**. The first condition requires that the ethical agent supervises itself, with no higher supervisor.[9] Intuitively, an ethical agent can look to themselves to "know what is right". This aligns with precedent, in that the "supervisory control on oneself" is the opposite of "just following orders". Western military forces empower their soldiers to reject unlawful or illegal orders, and rule that obedience to a superior officer is *not* a valid defence with respect to war crimes [26]. If "just following orders" is not a valid defence, it follows that there is an expectation that competent, adult soldiers have some innate capacity to generate their own orders; in other words, to supervise themselves.

   It is not enough for an agent to adhere to an "ethical standard" or "ethical code" for it to be an ethical agent. An algorithm for selecting targets and scheduling weapons is an instance of an ethical standard, but an agent following such an algorithm would merely be a lethal agent; regardless of how discriminate or proportionate. We might

[9] Nothing precludes a self-supervising agent from gathering information from other agents. However, this is a peer-to-peer relationship, not supervisory control.

have mechanisms for improving the algorithm, which themselves embed an ethical standard. In this case, the agents of improvement are merely the supervision chain. In contrast, the ethical agent would have an algorithm for revisiting algorithms, and can apply it to itself.

2. **Fastest tempo**. The second condition handles the case of multiple supervision chains. A key example is where a contractor designs and constructs an artificial lethal agent, and then a warfighter uses it in the field. Here, the warfighter is supervising the lethal agent at a higher tempo, and would thus be held as the ethical agent.[10] As a cross-check, we can look at the limiting case of supervision at faster tempos. The lethal agent is then operating at low to zero autonomy, with the ethical agent increasingly "in the loop".

For any given wartime casualty, we can identify the lethal agent, trace the supervision chains, apply the criteria and hence identify the ethical agent. The ethical agent may change from time to time, as the lethal agent or supervisory arrangements change, but the mechanism for identifying the agent remains the same. An observer may not agree with the ethics of a given ethical agent – an observer may in fact find them repugnant – but the agent can nonetheless be identified.

With the definition for an ethical agent, we can uniquely characterize the agent responsible for deaths under the Western ethics of war. We can now look at the requirements on combat systems engineering to support that responsibility.

Ethical Agency – A Uniquely Human Capability?

At time of writing, the capacity to self-supervise is unique to adult human beings, and arguably a necessary step towards Strong Artificial Intelligence [27, 28].[11] To emphasise: self-supervision is different from supervising at infinite autonomy. If a human programs and deploys an artificial agent, but then never visits it again, then the agent is being supervised at infinite autonomy. To be self-supervising, the agent needs to have a program that can rewrite programs, and the program needs to be able to take itself as its own input (without self-destructing!).[12] While artificial self-supervising systems have been conceived of in fiction [9, 10], there are no such systems operational today, and expectations are low for the near future.

As a consequence, for the foreseeable future, the ethical agent must be a human being. That is, ethical killing in war requires that there be a human "on the loop", in a supervision chain overseeing the lethal agent. In constructing the ethical agent, we made no assumptions about the nature of the lethal agent. Moreover, there is no apparent ethical requirement for humans to be "in the loop".

[10] We note that we have been seeking a single locus of responsibility. However, each of the self-supervising agents has the capacity to be held responsible; we have merely set a heuristic for selecting amongst them. Future work could revisit the heuristic; for instance, to assign weights of responsibility.

[11] See especially Hofstadter's proposal that "I" and "self" arise from "strange loops", constructed in the human brain, and stimulated by experience. Kurzweil counsels against a sole dependence on the "Consciousness is just a machine reflecting on itself" school, but recognizes that the perspective is at-least self-consistent.

[12] A *self-replicating* program can write an exact copy of itself. The programs that we seek do not write an exact copy, but still preserve a rewriting capacity.

We acknowledge the difference between *ethical* and *effective*. For any particular firing loop, a human may perform better than conceivable machine solutions, or they may not. Hence, the logic of effectiveness may call for a human "in the loop" (or not). The logic of ethics, however, calls for a human to be "on the loop".

Requirements from Combat Systems Engineering

We can now state the requirement from combat systems engineering. The engineering goal is to support the human, as they execute the following duties:

1. **Supervisory control over the lethal agent**. To revisit, at some tempo, the construction, deployment and programming of the lethal agent. In particular, to think about whether the robot has the "right" program, and to reprogram the robot to suit.
2. **Supervisory control over themselves**. To revisit, at some tempo, the basis on which they are supervising the lethal agent.

These requirements are consistent with effective support to command and control in the large [29].[13] The human is looking to dynamically craft the lethal agent as their instrument, while controlling the risks of non-combatant or friendly casualties. Commanders need not be "in the loop", micromanaging the agent's every action; rather they should be "on the loop" of employing the agent to achieve the mission.

The engineering goal can be expressed as an alternate school, of alerting the supervising human to potentially "wrongful" behaviours [30],[14] or possibly even blocking the human [31]. It may be desirable to ensure that the human is aware and conscious of their responsibilities, and not buffered from the reality of killing [32].

Nonetheless, command and control system should generate time for the human being to execute their duties. This can be thought of as the difference between *fighting-in-the-now* versus *fighting-in-the-future*. Continuously watching a full-motion video feed ("Predator crack" [33]) to find, fix and track a target is fighting-in-the-now. To fight-in-the-future is to think about courses of action to take if a target is found, or if a target acts in a certain way. Fighting-in-the-now leads to large spikes in cognitive load, and is thus highly stressful [34]. The engineering goal is to foster the human to fight-in-the-future.

The dimensions of control extend beyond the lethal agent itself; for instance, to deconflict the lethal agent away from non-combatant or friendly traffic [35].[15] There is a premium on architectures facilitating the dynamic assembly and upgrading of lethal agents [36],[16] notably data and algorithms for automated target recognition [2, 3], or weapons including less-than-lethal options [37]. The human may be the ethical agent to multiple lethal agents, again motivating attention on the control architectures [38, 39]. Conversely, the control of a given lethal agent might be contested (hostile action modelled as supervisory control!). The question of multiple agents raises interest in synchronisation [40].

[13] Pigeau and McCann define *command* as "the creative expression of human will necessary to accomplish the mission" and *control* "those structures and processes devised by command to enable it and to manage risk".
[14] See for example what Arkin calls a Responsibility Advisor.
[15] As demonstrated, for instance, in the use of a Joint Fires Area to mitigate the risk of fratricide between air and ground forces.
[16] What Alberts and Hayes called *edge applications*.

Finally, the human can change the degree of autonomy to which they release the lethal agent. The engineering goal is to support the human away from the extreme cases, namely: micromanaging the agent at zero autonomy, or deactivating the agent entirely.

Conclusion

Many systems can kill, in the sense of closing a firing loop. Only human beings have the ability to think about the rightness of killing, and the rightness of their own thinking. This factor ought to be central to the design of combat systems. The key is to place the humans "on the loop", as ethical agents in supervisory control over the lethal agent. Ethically speaking, it is neither necessary nor sufficient for humans to be "in the loop", immersed in every firing decision. Follow-on research should focus on C^2 design principles for supporting commanders to be "on the loop".

Acknowledgement

Colin Allen, Susan Blood, Richard Brabin-Smith, Lydia Byrne, John Canning, Gerhard Dabringer, Tony Dekker, Mary Cummings, Christian Enemark, Stephan Fruehling, Veronica Gray, John Hawley, Alex Kalloniatis, Gina Kingston, Elizabeth Kohn, Stephanie Koorey, Ed Lewis, Patrick Lin, Jeffrey Malone, Jeremy Manton, Gary Millar, Jeffrey Nachem, John O'Neill, Jon Rigter, Jason Scholz, Noel Sharkey, Robert Sparrow, Alan Stephens, Mike Sweeney, Richard Taylor, Paul Whitbread. The views and opinions expressed in this article are those of the author and do not necessarily represent those of the Australian Department of Defence.

References

[1] R. Sparrow, "Killer Robots," *Journal of Applied Philosophy,* vol. 24, no. 1, pp. 62-77, 2007.

[2] "Special Edition on Automatic Target Detection and Recognition," *IEEE Transactions on Image Processing,* vol. 6, no. 1, January, 1997.

[3] J. A. Ratches, C. P. Walters, R. G. Buser *et al.*, "Aided and Automatic Target Recognition Based Upon Sensory Inputs From Image Forming Systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, no. 9, pp. 1004-1019, September, 1997.

[4] R. F. Arnone, "Educating Our Bullets: A Roadmap to Munitions Centrality," US Army War College, Carlisle Barracks, PA, 1998.

[5] "STAR 21: Strategic Technologies for the Army of the Twenty-First Century," N. R. C. Board on Army Science and Technology, ed., National Academy of Sciences, 1992.

[6] M. O'Hair, B. Purvis, and J. Brown, "Aided Versus Automatic Target

Recognition," *SPIE,* vol. 3069, pp. 332-341, 1997.

[7] D. Jamieson, "Why Predators need people," *C4ISR Journal,* vol. 8, no. 9, pp. 42-44, October, 2009.

[8] M. Hoffman, "Unblinking Eye: Tracking Software Could Help Monitor UAV Feeds.," *DefenseNews,* 64, 2008.

[9] A. Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons,* Surrey, England: Ashgate, 2009.

[10] P. W. Singer, *Wired for War: The Robotics Revolution and Conflict in the Twenty-first Century*, New York: The Penguin Press, 2009.

[11] P. Lin, G. Bekey, and K. Abney, *Autonomous Military Robotics: Risk, Ethics, and Design*, California State Polytechnic University, 2008.

[12] N. Sharkey, "Grounds for Discrimination: Autonomous Robot Weapons," *RUSI Defence Systems*, pp. 86-89, October, 2008.

[13] *Macquarie Dictionary*, 2009.

[14] S. J. Russell, and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd Edition ed., Upper Saddle River, NJ: Prentice Hall, 2003.

[15] M. Waxman, *International Law and the Politics of Urban Air Operations,* RAND Monograph Report MR-1175-AF, RAND, 2000.

[16] P. M. Asaro, "Modeling the Moral User," *IEEE Technology and Society,* vol. 28, no. 1, pp. 20-24, Spring, 2009.

[17] T. B. Sheridan, *Telerobotics, Automation, and Human Supervisory Control* Cambridge: MIT Press, 1992.

[18] "*Unmanned Aircraft Systems Flight Plan 2009-2047*," United States Air Force, 2009.

[19] S. Beer, *Brain of the Firm,* London: Allen Lane, The Penguin Press, 1972.

[20] T. D. Wolf, and T. Holvoet, "Emergence and Self-Organisation: a statement of similarities and differences," in 2nd International Workshop on Engineering Self-Organising Applications, 2004.

[21] T. D. Wolf, and T. Holvoet, "Towards a Methodology for Engineering Self-Organising Emergent Systems," *Self-Organization and Autonomic Informatics,* H. Czap and R. Unland, eds., pp. 18-34: IOS Press, 2005.

[22] G. D. M. Serugendo, M.-P. Gleizes, and A. Karageorgos, "Self-Organisation and Emergence in MAS: An Overview," *Informatica,* vol. 30, pp. 45-54, 2006.

[23] A. Paramythis, “Towards Self-Regulating Adaptive Systems,” in Proceedings of the Annual Workshop of the SIG Adaptivity and User Modeling in Interactive Systems of the German Informatics Society (ABIS04), Berlin, 2004, pp. 57-63.

[24] R. Trevellyan, and D. P. Browne, “A self-regulating adaptive system,” *ACM SIGCHI Bulletin,* vol. 18, no. 4, pp. 103-107, April, 1987.

[25] A. Paramythis, "Can Adaptive Systems Participate in Their Design? Meta-adaptivity and the Evolution of Adaptive Behavior," *Adaptive Hypermedia and Adaptive Web-Based Systems,* Lecture Notes in Computer Science, Berlin / Heidelberg: Springer, 2006.

[26] B. Copelin, “Defending The Indefensible: The Defence Of Superior Orders For War Crimes,” *Australian Army Journal,* vol. 6, no. 1, pp. 37-56, Autumn, 2009.

[27] D. Hofstadter, *I am a Strange Loop*, New York: Basic Books, 2007.

[28] R. Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, New York: Penguin Books, 1999.

[29] R. Pigeau, and C. McCann, “Re-conceptualizing Command and Control,” *Canadian Military Journal,* vol. 3, no. 1, pp. 53-64, Spring, 2002.

[30] R. C. Arkin, *Governing Lethal Behaviour: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture,* Technical Report GIT-GVU-07-11, Georgia Institute of Technology, 2007.

[31] P. Asaro, "What Should We Want from a Robot Ethic?," *Ethics and Robotics*, R. Capurro and M. Nagenborg, eds., Amsterdam: IOS Press, 2009.

[32] M. L. Cummings, “Automation and Accountability in Decision Support System Interface Design,” *Journal of Technology Studies,* vol. 32, no. 1, pp. 23-31, 2006.

[33] S. D. Bass, and R. O. Baldwin, “A Model for Managing Decision-Making Information in the GIG-Enabled Battlespace,” *Air & Space Power Journal*, pp. 100-108, Summer, 2007.

[34] M. L. Cummings, “Automation Bias in Intelligent Time Critical Decision Support Systems,” *American Institute of Aeronautics and Astronautics,* 2004.

[35] J. E. Mullin III, “The JFA: Redefining the Kill Box,” *Fires Bulletin*, pp. 38-41, March-April, 2008.

[36] D. S. Alberts, and R. E. Hayes, *Power to the Edge: Command and Control in the Information Age*: Command and Control Research Program, 2003.

[37] J. S. Canning, ““You’ve Just Been Disarmed. Have a Nice Day!”,” *IEEE Technology and Society,* vol. 28, no. 1, pp. 12-15, Spring, 2009.

[38] M. L. Cummings, S.Bruni, S. Mercier *et al.*, "Automation Architecture for Single Operator, Multiple UAV Command and Control," *The International C2 Journal*, vol. 1, no. 2, pp. 1-24, 2007.

[39] K. Button, "The MAC attack," *C4ISR Journal*, vol. 8, no. 9, pp. 30-32, October, 2009.

[40] A. Kalloniatis, "From Kuramoto to Boyd: applying networked dynamical systems to military Command & Control," in International Workshop on Complex Systems & Networks 08, Canberra, 2008.