Australian Government
Department of Defence
Defence Science and
Technology Organisation

Issues Regarding the Future Application of Autonomous Systems to Command and Control (C2)

Michael Pilling

Joint and Operations Analysis Division

Defence Science and Technology Organisation

DSTO–TR–3112

ABSTRACT

This broad review provides some insights into the vast field of Autonomous Systems in the context of Defence applications and C2 in particular. There are significant challenges in the areas of human-computer interaction and the legalities of war that may trump technical issues in terms of impediments to automation. Many technical areas are also covered and the paper recommends developing or adopting strong engineering processes for building Autonomous Systems, research into human factors and strong engagement with the international community with respect to the Laws of Armed Conflict.

APPROVED FOR PUBLIC RELEASE

Published by

DSTO Defence Science and Technology Organisation
Fairbairn Business Park,
Department of Defence, Canberra, ACT 2600, Australia

Telephone: (02) 6128 6371
Facsimile: (02) 6128 6480

AR No. AR-016-315
June 2015

APPROVED FOR PUBLIC RELEASE

Issues Regarding the Future Application of Autonomous Systems to Command and Control (C2)

Executive Summary

This paper is a broad review of the research literature regarding Autonomous Systems in Defence with respect to the C2 domain in particular. The review seeks to identify issues that must be dealt with in order to build and successfully deploy Autonomous Systems, areas that are most actionable, and particular opportunities in C2. Significant themes emerged, including the primacy of human-computer interaction in enabling effective Autonomy by enabling adequate tasking, to engender trust so the autonomous systems will be used rather than abandoned, and increasing the speed and amount of delegation which is essential to relieve human operators of task loads and/or reduce the number of human operators. The significant legal issues involved in Autonomous Weapons Systems were also a major theme. These issues in many ways eclipse the significant technical challenges of creating autonomous systems.

The discourse on Autonomous systems is confounded by the conflation of several types of systems that operate with minimal human intervention. We distinguish between Automatic systems, Automations and Autonomous Systems in order to clarify the issues particular to each of these progressively more complex systems and ultimately focus on the issues which make building Autonomous Systems an area of significant technical challenge. We use automation in lower case as a noun to refer to this entire spectrum of automation, and as a verb to refer to the process of increasing the complexity or level of automation.

This paper summarises the significant legal arguments raging about whether Autonomous Weapons Systems *can* comply with the Laws of armed conflict. The arguments can be grossly simplified to the arguments that only a human has the right to decide to kill and that only humans have the capacity to fully interpret the environment and behaviour of humans in it in order to validly decide whether a particular human is a legitimate target; versus the argument that automated systems may in practice be more humane than humans for a variety of reasons, including because they are willing to put themselves in harms way (and so are less inclined to shoot first), and do not have a tendency to rape, to act in revenge or panic.

Lawyers, non-goverment organisations and the International Committee of the Red Cross (ICRC) have raised significant concerns surrounding the prospect of Lethal Autonomous Weapons and the UN has called for a moratorium on such systems so that international agreements about the issues can be made. The primary concern is that deployment or even development of such systems may have a destabilising effect on the world and fracture the peace by leading to sudden and unexpected escalations. Citizens, community groups and NGOs are actively involved in arguing for a ban on lethal Autonomous weapons systems.

We focus on, the hard and poorly understood Autonomous Systems for which we have no reliable production processes.

This paper suggests issues to consider in Human Factors include:

- Interfaces between humans and automations
- Conveying uncertainty to operators
- Keeping the operator engaged to maintain situation awareness
- Reducing the cognitive load on operators
- Operating in teams, and the transferal of authority and responsibility between humans and automations and vice versa
- Maintaining trust.

This paper argues that C2 is a unifying concern when increasing the automation of Defence systems and argues among other things for:

- Measures of automation and its effectiveness
- Development of clear Concepts of Operations
- Research into the certification of Autonomous systems
- The development of standards to enable Defence integration of Autonomous systems
- Requiring that Autonomous Systems built or procured have integral autonomic capabilities.
- The adoption of a time base from which causality can be clarified and "blackbox" type recording of autonomous decisions and their antecedents.
- The adoption of strict safety protocols throughout the development, testing and deploying of Autonomous systems.
- Adopting and where necessary developing strong engineering processes to enable predictable development and certification of Autonomous Systems.

It identifies the need for Defence wide standards to ensure that Autonomous Systems automatically monitor and maintain their own health as much as possible, starting at the lower automation levels.

It also argues that the Australian Government must be engaged in international discussions regarding LOAC and Autonomous Weapons Systems.

Although the field is massive, this literature review was able to suggest an initial path forward in both research and implementation that will enable progress and further clarity at a later date.

Author

Michael Pilling
JOAD

Michael Pilling completed a Bachelor of Science degree with honours in 1987 and a Ph.D. in Computer Science in 1996 at the University of Queensland, Australia. Michael's specialities are distributed and real-time systems, job scheduling, formal specification and program correctness, criticality management, and the calculus of time. His current interests include Software Reliability Engineering, Failure as a fundamental construct in usable and effective systems, Virtual Synchrony and its application to synchronous group communication, performance engineering of computer systems, and graceful degradation of systems in the face of failure and overload.

THIS PAGE IS INTENTIONALLY BLANK

Contents

Glossary **xiii**

1 Introduction **1**

1.1 What is an Autonomous System? . . . 2

1.2 Scope of this paper . . . 5

2 Autonomy in the C2 Context **5**

3 Autonomous Systems **9**

3.1 Other characterisations of autonomy . . . 9

3.2 Types of Control . . . 11

4 Artificial Autonomy **12**

4.1 Agent Programming . . . 14

4.2 Embodied Agents . . . 17

4.3 Autonomic Systems . . . 17

4.4 Artificial Autonomy and C2 . . . 18

5 Applications of Autonomous Systems **19**

5.1 Decision Support . . . 19

5.2 Autonomous Systems in Market Trading . . . 21

5.2.1 Autonomous Systems Issues arising in Market Trading . . . 22

5.3 Transport . . . 24

5.3.1 Land . . . 24

5.3.2 Air . . . 26

5.3.3 Maritime . . . 28

5.3.4 Space . . . 28

6 Law of Armed Conflict(LOAC) **30**

6.1 Legal & Moral Responsibility . . . 31

6.2 Risks of Escalation and Instability . . . 34

6.3 Legal Considerations . . . 35

6.4 Social Licence . . . 39

6.5 National Approaches to Regulation . . . 40

6.5.1 Australia . . . 40
6.5.2 USA . . . 43
6.5.3 UK . . . 44
6.6 Overall legal situation . . . 47

7 Human Interaction with Automations and Autonomous Systems 49
7.1 Interaction Issues from Robotics . . . 49
7.2 Human Factors, Operations and Trust . . . 50
7.3 Certification and Validation for Trust . . . 52

8 Discussion 53
8.1 Human Factors . . . 53
8.1.1 What should be automated? . . . 54
8.2 Creating and Maintaining Trust . . . 55
8.3 Measures of Autonomy and its Performance . . . 55
8.4 Legal Issues, Certification and Safety . . . 56
8.5 Self Awareness . . . 58
8.6 Counter Measures . . . 58
8.7 Standards and Integration . . . 59
8.8 The need for Engineering Methodologies . . . 60

9 C2 Opportunities 60

10 Conclusions 62

References 63

Figures

THIS PAGE IS INTENTIONALLY BLANK

Tables

THIS PAGE IS INTENTIONALLY BLANK

Glossary

Adaptive Autonomy A mode in which the autonomous computing agent has exclusive control of the system, and allocates tasks to the human operator as it sees fit.

Adjustable Autonomy A mode in which the human has exclusive control of the system, deciding which tasks to delegate to autonomous computing subsystems.

AI Artificial Intelligence

ATC Air Traffic Control

Autonomic System A system with capacity to measure its own capacity, health and performance and either report its "health" or preferably reconfigure itself as appropriate. See section 4.3.

Authorised Entity An individual [human] operator or [machine] control element authorised to direct or control system functions or mission.

AUAS Autonomous UAS

AUV Autonomous Underwater Vehicle

DARPA Defense Advanced Research Projects Agency (US)

Drone any remotely piloted vehicle

DSS Decison Support System. A system designed to aid humans reaching decisions.

Firing Autonomy The ability of an automated system to commence firing / engagement without recourse to a human.

Full Autonomy Firing Autonomy + Target Selection Autonomy

GCS Ground Control Station

GPS Global Positioning System, a space-based satellite navigation system that provides location and time information in all weather conditions, anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites[1].

Heterachical System One in which no entity in the system exerts a permanently dominant influence on the system elements[Pro92] as cited in[UIT+11].

HFT High Frequency Trader. A trader that uses algorithms to move in and out of positions is seconds or fractions of a second.

HMI Human Machine Interface. The physical or software tools and processes by which the human and machine communicate.

ICRC International Committee of the Red Cross. The part of the Red Cross movement that deals specifically with the Geneva Conventions.

[1]Definition: Wikipedia

IHL International Humanitarian Law

IHRL International Human Rights Law

ILS Instrument Landing System: A ground-based instrument approach system that provides precision guidance to an aircraft approaching and landing on a runway, using a combination of radio signals and, in many cases, high-intensity lighting arrays to enable a safe landing during instrument meteorological conditions, such as low ceilings or reduced visibility due to fog, rain, or blowing snow[2].

In-the-loop Refers to the sense-think-act loop used by automations. "In" refers to a human, inserted into this repeatedly executed sequence to approve significant actions, usually decisions to target, engage or fire.

IW Information Warfare

LAR see Lethal Autonomous Robots

Lethal Autonomous Robots A term used by the UN defined as a weapons system that, once activated, can select and engage targets without further human intervention.

Lidar Light detection and ranging: A remote sensing technology that measures distance by illuminating a target with a laser and analyzing the reflected light.

LOA Level of Automation

LOAC Law of Armed Conflict. Includes IHR and IHRL.

MAS Multi-Agent System(s).

Mixed Initiative A system mode in which the human and autonomous computing agent share responsibility for allocating tasks between the computing system and the human. Each may request the other to perform a function it finds difficult to perform alone.

Network Centric Warfare (NCW) The linking of sensors, engagement systems and decision-makers [across multiple platforms] into an effective and responsive whole.

NGO Non-goverment organisation. A community group, often a charity or not-for-profit organisation. For example, the Red Cross.

On-the-loop A human approval is not needed to action decisions, but it is intended a human will intervene and countermand any inappropriate decision. See in-the-loop.

Operator A human in command of some aspect of a system

Out-of-the-loop No human is in-the-loop and the system can action its decisions without oversight: see in-the-loop.

Outside-the-loop See Out-of-the-loop

[2]Definition: Wikipedia

Remote Operation Systems which allow the pilot or operator of a system to be physically distant from the actual hardware or weapon, operating it by a communications link which provides sensor information and the ability to command the system from a remote location.

ROE Rules of Engagement

Semi Autonomous A system which can make some autonomous decisions within a highly constrained range of choices prescribed by a human, often with respect to the low-risk or mundane parts of a mission such as navigating automatically to a human defined destination but not choosing that destination.

Shared Situation Awareness Situation Awareness in which the overall perception of elements is distributed among multiple human or automated entities. The overall awareness is formed by the union of each entities awareness. Each entity's awareness may overlap with those of others but should not conflict with them. Ideally, each entity should have enough awareness of the environment being managed by other entities they are directly working with to be able to quickly recover full local awareness should any entity fail or leave.

Situation Awareness There are several definitions, but a a commonly accepted one is: The perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future[End95].

Supervised Autonomy An autonomous system with a human outside-the-loop monitoring it with view to overriding its decisions.

T & E Test and Evaluation

Target Selection Autonomy A system where the automation is free to select targets based on its own assessment of their relative military value.

Temporary Autonomy The ability of a system to continue to perform some useful functions autonomously in between operator commands.

Test and Evaluation The process by which a system or components are compared against requirements and specifications through testing. The results are evaluated to assess progress of design, performance, supportability, etc. Developmental test and evaluation is an engineering tool used to reduce risk throughout the defense acquisition cycle[Con13].

Threat Evaluation The process of determining if an entity intends to inflict evil, injury or damage to our forces or their interests, along with ranking these entities according to their level of threat[PBOC05]

UAS Uninhabited Air System. A UAV including its support systems such as the remote pilot, maintenance staff, landing area, remote computing support, network connectivity etc.

UAV Uninhabited Air Vehicle

UCAV Uninhabited Combat Air Vehicle. A UAV that is armed.

UGV Uninhabited Ground Vehicle

V & V Verification and Validation

Verification and Validation Independent procedures that are used together for checking that a product, service, or system meets requirements and specifications and that it fulfills its intended purpose.

Weapons Allocation The assignment of weapon systems to engage or counter identified threats.

1 Introduction

Artificial Autonomy offers many potential benefits for C2 (Command and Control) implementation and operations, this paper broadly examines uses, issues and technologies of artificial autonomy with a view to how they might be built or engineered and how they may be applicable to C2. While C2 is concerned with and affected by autonomy in general, much research pertaining to artificial autonomy has been performed on robotics platforms which can inform C2 issues including legalities, ethics and maintaining a social licence. This paper does not examine questions of human autonomy, except with respect to how humans interact with artificial autonomy, and it is generally left up to the reader to recognise when the autonomy referred to is human or artificial.

P. W. Singer[Sin09b] describes the robotics revolution as the next major revolution in military affairs, akin to gunpowder and nuclear weapons. Likewise, Australian researchers see Ground Vehicle autonomy as potentially revolutionary[Fin08]. According to Singer[Sin09a], as of 2009 44 countries were developing robotic weapons systems including the UK, France, Russia, China, Israel, Iran and the UAE.

> The over-riding motivation for moving to unmanned systems is a desire to deliver new or enhanced capability by embracing new technology while reducing costs and the threat to personnel[MoD11].

There are huge potential military advantages in deploying autonomous systems[Hey13]: they offer higher force projection and force multiplication and can do the dirty, dangerous and dull work. They offer the possibility of using less than lethal force, and to provide a digital trail of their decision making in a way that would enhance accountability. Their reaction time may be far superior to that of humans and the increasing tempo of warfare can leave humans as the weakest link. Their strengths such as lack of fatigue, fast reaction time, ability to process huge amounts of data, and to find the most unexpected correlations can potentially complement the best human characteristics such as "intuition", fine judgement and compassion. They can also continue to operate, in a perhaps degraded manner, when communication with the command structure is unavailable. Others point out the potential for Autonomous systems to greatly enhance the coherence of operations[Fin08].

Autonomous systems promise benefits for C2 through improvements in coordination, logistics planning and implementation, data fusion and interpretation, and in decision making or guidance. These benefits may occur at the tactical, operational and strategic levels. They take many forms including higher work capacity, greater speed and accuracy, the ability to test plans and decisions by simulating them and testing more of the enemy's possible responses further into the future.

Ideally, autonomous systems would be implemented and deployed in a way that forms "teams" which take full advantage of the complementary strengths of humans and automations.

1.1 What is an Autonomous System?

We note that the Oxford dictionary defines a **system** to include "a set of things working together as parts of a mechanism or an interconnecting network; a complex whole" and "a set of principles or procedures according to which something is done; an organised scheme or method". It defines **autonomy** as "the right or condition of self-government" and also notes "(in Kantian moral philosophy) the capacity of an agent to act in accordance with objective morality rather than under the influence of desires". This later definition is interesting in relation to the Law of Armed Conflict which this paper covers in section 6.

While philosophical discussions of what autonomy is in human terms are both interesting and useful in their own right; just as understanding the way tailors stitched together garments was not particularly helpful in developing effective sewing machines, understanding human autonomy does not appear to overly inform building useful Autonomous Systems so we do not cover such arguments in this paper.

There is no single, generally accepted definition of autonomy and autonomous systems in the literature. Worse still, the discourse on autonomy in general is confounded by the conflation of several types of systems that operate with minimal human intervention. However, it is clear that a spectrum of increasing environmental uncertainty and hence system complexity and adaptability exists between Automatic Systems, Automation and Autonomous Systems. We now propose working definitions of these places along that spectrum:

Automatic systems are rigid and operate with limited or negligible human intervention. They have very limited inputs, both in terms of number of dimensions and in having defined ranges or a small number of discrete values. The only way they embody a history of their interactions is through which of a limited number of system states is current. The output for each combination of inputs and system state is predefined and is a deterministic simple function of the combination.

Automatic systems are the simplest class of system in the spectrum of automation complexity and are typically rigid in their operation. They are often direct implementations of mundane human physical or cognative work into a machine. Examples include: the simplest type of traffic lights which operate either on a timed loop, or respond to simple inputs such as pedestrian request buttons and under road car sensors which may accelerate the next signal change in the absence of conflicting inputs; simpler building elevator control systems; simple railway signals that transmit stop and caution signals down the line in the presence of a train; simple automatic transmissions that set the gear based purely on the current speed of the car. Many alarms and fail-safe mechanisms are implemented with automatic systems, in part because such simple systems have fewer failure modes.

Automations are more complex systems dealing with more complex inputs that exhibit considerable complexity and may have a large volume of data or many different types of input. Nevertheless, these inputs and their values are well defined and bounded and their meaning is well understood by the system designer. Moreover the environment or context of an automation is tightly constrained and presents risks to the system that can be quantified by stochastic risk analysis. The outputs of automations may be a simple or complex function of the current inputs and the history of inputs and system state but is

well defined and understood *a priori* by the system's designers. Automation outputs are deterministic functions of inputs unless deliberately randomised to achieve optimisation, fairness or secrecy.

The generally deterministic or specifically random nature of automations mean they can be tested by input elaboration, at least if one had infinite time. Alternatively, a formal proof of correctness may be possible. When automation inputs fall outside of their design expectations, a common response is for the automation to shutdown and return control to its operators or to fail-safe. For example an autopilot will disengage and return control to the human pilot, or a nuclear reactor control system will perform an emergency shutdown.

Automations represent an intermediate area on the spectrum of automation complexity. Examples of automations include: networked traffic lights or train control that include aspects of scheduling, system optimisation or adapting to recent or repeated historic demands; process control systems for manufacturing; sophisticated automatic gearboxes that learn the driving patterns of their users to optimise future gear choices; Roomba type vacuum cleaners lie somewhere in the spectrum between automatic systems and automations depending on the sophistication of their programming; Autopilots; and the Traffic Collision Avoidance System(TCAS)[1]. Although TCAS is clearly giving directions autonomously, and it could be argued to be an autonomous system, we believe it is simple enough to be considered just an automated system because the mapping between well defined input data and outputs is clearly elaboratable at design time. It is also convenient for us to reserve the definition of autonomous systems to ones which we find particularly challenging to design and build. In this way, our definition of autonomous systems will focus our attention on precisely those systems that we wish to build but currently lack sufficient understanding of to engineer on demand.

Autonomous Systems operate in highly complex environments where the inputs to the system can include values or dimensions unforeseen at design time. The data environment is often infinite in potential range and dimensions requiring data reduction by selection, filtering and interpretation (i.e. several conclusions may be drawn from the data and it is important that the machine's include the appropriate ones while ignoring those that are irrelevant). Alternatively, the meaning of the data can be contextually dependent on a highly variable or even chaotic system. Such environments exhibit unquantifiable uncertainty. The outputs generated by such inputs can not be fully predicted at design time and the system designers may use methods such as heuristics or machine learning to limit outputs to "acceptable" as opposed to singularly correct ones[2]. They are autonomous in the sense that their decisions or actions are not totally foreseeable at design time, but may generate unexpected novel solutions. The system's decision criteria or decision algorithms are often emergent properties of the system's history of inputs and its interactions with its environment over time. In this sense they are self governing. Once commanded by their operator, most autonomous systems operate for significant periods or over significant volumes of data without seeking or requiring further human guidance even when unexpected inputs arise. Unlike lower level systems, such commands tend to be strategic or at least

[1] See section 5.3.2 for a discussion of TCAS.

[2] The problems faced by Autonomous Systems are too complex to assume a single correct solution exists, or that the system will have time to find an optimal result. Herbert A. Simon's coined term "satisficing" [Sat09, Sat14] expresses this idea of searching for a solution that meets minimum critera, however unlike his postulated "Administrative Man" Autonomous Systems should perform affordable partial optimisation.

operational.

The range, dimensions and highly variable nature of data input to autonomous systems preclude system testing solely by elaboration of input cases, and preclude complete formal proofs of correctness. Autonomous Systems are generally goal directed, executing multiple steps to acheive their goal(s), and many Autonomous Systems must engage in some level of multi-step forward planning in order to fulfil their mission. Larger autonomous systems also need to actively manage their own operation, which requires various levels of autonomic capability[3] (see section 4.3).

Autonomous systems represent the most complex area in the spectrum of automation. Examples of autonomous systems include proposed deep space vehicle control systems that reconfigure themselves from software component libraries[SA09], data mining systems that develop new correlations from masses of input data and control systems that learn from being exposed repeatedly to their environment and modify their behaviour.

These three definitions represent exemplars on the spectrum of automation, however it is important to note that systems representing more complex levels of automation may contain subsystems lower in the spectrum of automation complexity. It makes perfect sense for autonomous systems to contain some automatic systems, just as animals relegate some responses to reflexes rather than engaging the brain. Many systems will not fall entirely under one of the definitions given above but lie somewhere between automatic and automated systems, or somewhere between automated and autonomous systems.

Despite our definitions, it is awkward to avoid the venacular usages of "automation" so throughout this paper we use "automation" in lower case as a noun to refer to this entire spectrum, and as a verb to refer to the process of increasing the complexity or level of automation; and refer to our definitions using capitals.

Given the above discussion, we can now define autonomy as it applies to systems or machines:

Autonomy is the ability of a system or machine to operate for a significant period of time, or over significant volumes of data, without further external human direction while progressing socially agreed or directed goals even when faced with unexpected situations.

We note that this definition could equally be applied to human systems such as corporations or public service departments provided we distinguish between external directions and those originating from within the organisation. It is appropriate that this definition spans both human and non-human systems because being autonomous does not imply operating as a sole agent but operating within a social context. Clearly a system that operated randomly or purely to its own ends would exhibit a form of autonomy, however in this paper we are only interested in systems which are useful to build. Moreover, humans exhibiting such extreme random or purely selfish behaviour, as opposed to socialised autonomy, are labelled insane or criminal. Acceptable autonomy always occurs within a social construction[4] and effective autonomous agents operate as part of a team.

[3] The capacity to be self monitor and to self regulate.

[4] Although this may occur at design time.

1.2 Scope of this paper

This review is intended to be a broad but necessarily limited look at the field of autonomous systems particularly as applied to the Defence domain and C2 in particular. It is not intended to be comprehensive, but to reveal some of the strongest themes in the literature. We also wished to discover any potential pit falls in pursuing Autonomous Systems and areas that might be most fruitful to explore or be immediately actionable. In this sense we have taken a pragmatic approach.

The literature covered was highly diverse and included aspects of machine learning, defence applications, human computer interaction and the legalities of autonomous weapons systems. The survey is skewed towards aspects of the legalities of autonomous systems, and does not attempt to cover all possible mechanisms for artificial intelligence(AI).

Themes that repeatedly arose across multiple domains were the difficulty in providing workable human computer interfaces, the difficulty in keeping the human's attention in a productive way, and need to make any automation productive for the user. Very few of the papers provided detailed solutions to particular problems particularly in AI terms, describing instead the broad approach or architecture used.

While our definitions of autonomy allowed for humans and organisations of humans to exhibit autonomy, in this paper we are concerned with which computer systems can exhibit autonomy in the sense described in our definition of Autonomous Systems regardless of the extent to which they are physically embodied (robotic). The objective is to shed light on the unique features of such systems and the prerequisites for building effective examples of such systems. We are less concerned with whether a system exhibits "true" or "full" autonomy than whether it exhibits useful levels of autonomy.

While all points in the spectrum of autonomy elaborated in this introduction can be useful for C2, the lower levels of the spectrum are understood well enough to produce reliable useful systems on demand. In this paper we are concerned with the higher end Autonomous Systems in order to see what might be needed to build them with respect to applying them to C2.

2 Autonomy in the C2 Context

Australian Defence Force doctrine[Dep09] defines *Command* as:

> The authority that a commander in the military service lawfully exercises over subordinates by virtue of rank or assignment. Command includes the authority and responsibility for effectively using available resources and for planning the employment of organising, directing, coordinating and controlling military forces for the accomplishment of assigned missions. It also includes responsibility for health, welfare, morale and discipline of assigned personnel.

and *Control* as:

> The authority exercised by a commander over part of the activities of subordinate organisations, or other organisations not normally under his command, which encompasses the responsibility for implementing orders or directives. All or part of this authority may be transferred or delegated.

The above definitions do not, however, expose the key role that *situation awareness* plays in effective Command and Control. There are many definitions of situation awareness but a useful and commonly accepted one is that of Endsley:

> the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future[End95].

Pigeau and McCann[PM02] give another popular definition of Command and Control:

Command The creative expression of human will necessary to accomplish the mission.

Control Those structures and processes devised by command to enable it and to manage risk.

Whereas Lambert and Scholz[LS07] describe Command and Control as strongly related to action in the following way:

Command involves the creative expression of intent to another.

Control involves the expression of a capability (a plan is an example of a capability) to another and the monitoring and correction of the execution of that capability.

Action is utilisation of capability to achieve intent, given awareness.

Since managing risk requires situation awareness these definitions implicitly or explicitly raise generating and maintaining situation awareness to a primary activity of, and a prerequisite for, Command and Control.

Drawing on these definitions and our definitions of autonomous systems we argue that Autonomous Systems for C2 would ultimately perform the following roles:

- Direct, organise and coordinate other force elements, humans and autonomous agents to achieve goals supporting command intent.
- Negotiate with other entities, including external ones, to achieve goals and monitor progress so as to take alternative action if progress is insufficient.
- Perform actions and report progress to fulfil agreements it has made with external parties.
- Assist in the development of shared perception of the operating environment and its extrapolation into the immediate future[5].

[5] By shared we specifically mean between autonomous systems and humans as well as among humans. This is a significant challenge especially in balancing the aim of reducing cognitive workload while maintaining enough human awareness to keep humans up to speed.

- The translation and elaboration of command intent into testable and executable plans and actions.
- Taking actions in support of intent and monitoring success or failure of those actions, which may lead to replanning.

Lambert and Scholz[LS07] generalise the concept of C2 to Ubiquitous Command and Control(UC2) by acknowledging and incorporating societal shifts such as:

- Moving from stability of the organisation/hierarchy as the norm, to change and adaptation as the norm.
- Democratising access to previously privileged information.
- Increasing virtual presence, a greater number of players and a greater number of types of players, leading to both more competition and collaboration among them.

This must result in and be facilitated by:

- More agreements being negotiated and executed in bounded time frames.
- "Disintegration" and per use assembly of previously integrated C2, awareness, logistic and other support functions including choosing such functions from various providers; or as the authors' put it, balancing Unity and Diversity.
- A greater reliance on networks [and implicit trust], and less emphasis on hierarchy.
- Multi scale networks.

Lambert and Scholz nominate Automation as a component of Ubiquitous Command and Control. In fact, certain paradigms for implementing automation and autonomous systems provide a natural platform on which to build a UC2 infrastructure (see section 4.1). Indeed, to reach true ubiquity requires the ability to automate decision making and to provide autonomous decision aids so that humans at all levels and automations can fulfil the role of "agent".

There are challenges. Military C2, due to its inherent need to unleash lethal force and the necessity to protect our forces from enemy's lethal force invokes other requirements for C2 Autonomous systems:

- To provide traceable and legally accountable decisions.
- To operate in a way that maintains the military's (and researcher's or developer's) social licence [6]. That is the systems must be prevented from taking inappropriate actions, and avoid accidents even during their development.
- Quality control and certifying C2 systems fit for purpose including subsystems as appropriate.

[6] Experience shows it only take one mishap during early operations or research experiments to turn public opinion against whole areas of research[MW09].

- It is still unclear the extent to which machine learning will be required in all Autonomous C2 systems. While some Autonomous C2 systems may require no, or only shallow machine learning, it is clear that deep machine learning will be a requirement of many.
- To be able to monitor, diagnose, and largely correct the C2 system's own health (see section 4.3).
- To be able to operate seamlessly with human operators and commanders, and in some cases dynamically replace humans with autonomous systems and vice versa. This is the UC2 integration tenet, but also arises independently from applying a Turing test to C2 Automation. Such substitutablity represents the highest possible level of Autonomous function.

Information superiority is a key enabler for C2 success, supporting timely, sound and informed decisions[PBOC05]. While automation provides many opportunities as mentioned above, the specific C2 opportunities include the potential to dynamically adjust a C2 system's scale, composition and robustness, balance cognitive load by placing more load on automations, and improve ethical decision making[SLGS12], and to dramatically improve the speed and capacity of logistics planning with consequent staging improvements[CG08]. Ongoing issues for C2 operators include the level of cognitive load placed on them, as well as the clarity and usability of C2 human-computer interfaces.

There exist autonomous fusion systems which fuse multiple sensor readings into tracks, classify contacts into friend, foe, or unknown, task assets to acquire further information on contacts where necessary[Lam09], and maintain one or more projections of the current tracks into the future. Other C2 systems automatically develop logistics plans and place orders with defense and contractor's systems to have goods delivered and track their delivery, all the while doing performing contingency planning and replanning when deliveries fail[CG08]. Eventually C2 systems may deploy counter measures, such as the Nulka decoy, remotely across the theatre; they may predict future warfighting actions and predeploy or partially deploy[7] units. Ultimately they may plan and autonomously execute whole segments of operations, tasking units to perform actions that would progress military objectives, and initiate non-lethal effects based operations.

All levels of the automation spectrum we describe are useful to C2. Automatic systems that respond with local counter measures when scanned by enemy radar, automated systems that guide a platform to a way point freeing pilots up for less mundane activities, and autonomous systems as described above. In fact most autonomous systems will include or rely upon lower automation spectrum systems in their standard operation.

Technically, introducing autonomous systems into C2 requires judicious choice of technology and a clear path by which to introduce it safely while not compromising existing C2 systems. Automating lethal force invokes the worse legal and public trust complications and so long as C2 systems avoid this, they will at most have to deal with legal negligence issues (see section 6).

Perhaps the most radical shift required to produce good C2 Autonomous Systems will be viewing teams as being composed of a set of humans combined with a set of autonomous

[7] Move a unit closer to its expected deployment, but still close enough for easy recall.

systems. This new concept of team is preferable because it forces developers to focus on ensuring the right types of communication among team members, appropriate function demarcation and dynamic function allocation between them, and maintaining sufficient mutual awareness of task progress and critical task information to enable replacement of one team member by another whenever necessary even across the human/autonomous system boundary. Examples of this have already emerged in the arena of Chess Championships which progressed from human vs. human competitions through human vs. computer competitions to competing human-computer teams[Adv14] under the leadership of Garry Kasparov. This innovation is called Advanced Chess or Centaur Chess, and has resulted in extraordinary rises in the level of play both at the tactical and strategic level and greater audience insight into the "thought processes" of both the human and computer players. This result gives some insight into the potential of UC2 and the use of Autonomous Systems in C2 if done well.

3 Autonomous Systems

3.1 Other characterisations of autonomy

Finn defines several categories of autonomy. In 2008 discussions of *autonomous weapons* typically realated to those independent from human control where beyond a certain point they determined their own mid-course and final trajectories, examples being cruise missiles and torpedos[Fin08]. However, this is not the type of system regarded as autonomous today; the majority of current papers implying a far more strategic level of decision making when talking about autonomy - see examples in this section.

Finn points out that these and weapons that recognise their targets in the last stages of their mission, or are pre-programmed to recognise a very specific type of target such as a radar emitter of non-friendly frequencies that happens to be within the theatre of war, are really only *semi-autonomous*[Fin08]. It should be pointed out that this is little different from human military decision making in that each role will have constraints on what decisions can be made without reference to a higher authority and even staff at the highest levels must comply with the rules of engagement. In this sense, semi-autonomous operation is the normal state of human affairs. Such target recognition systems can be considered to have *firing autonomy*[Fin08], the ability to initiate firing without reference to an external authority given that the target has been selected and the system has only recognised the target.

Originally, systems that support *remote operation* were entirely tele-operated with the operator performing all cognitive processes, however now most such systems exhibit significant amounts of autonomy for maintaining their trajectory and flight, but revert to operator control at decision points and other critical moments of the mission. It is important to automate the majority of the system's operation to reduce operator fatigue on long missions, and to increase the number of units the remote operator can simultaneously manage. Such systems are again *semi-autonomous* in that they have sufficient navigation, collision avoidance and data-fusion capabilities to allow the operator to select the next destination or target with which to engage[Fin08].

Target selection autonomy is where the system is allowed to select its own targets without human intervention from a broad description of possible targets, their characteristics and their relative value. Finn then concludes a weapons system with *full autonomy* would exhibit both target selection autonomy and firing autonomy, and would not have a human in-the-loop. We note that only systems with such full autonomy could of themselves breach the Geneva Conventions or Rules of Engagement.

We find Target Selection Autonomy and Firing Autonomy useful concepts for characterisation of systems, but almost all Autonomous Systems will exhibit what we would call *bounded autonomy* in some manner. Even monarchs have limits on their powers in functioning societies and so the goal of Full Autonomy is less useful.

We note that simple booby trap weapons like IEDs or land mines fall far more into the Automatic System category due to their relatively indescriminate nature and so while independent, are not autonomous in the full sense given that they do not incrementally advance towards a goal.

Some working definitions of autonomy, other than our own, included:

- A weapons system that, once activated, can select and engage targets without further human intervention[Hey13] (citing use by the US and Human Rights Watch).
- The ability to act with some measure of independence, and to assume responsibility for one's own resources and behaviour. This in turn makes it possible for systems to function in the absence of centralised external control[SA09].
- Autonomous systems can perform tasks in an unstructured environment. Such a system is marked by two attributes: self-sufficiency — the ability to take care of itself — and self-directedness — the ability to act without outside control.[BGS11]
- *Automation* focuses on automating specific processes for humans, where *autonomy* is the machine achieving self-governance through *independent* decision-making[Twe12].
- Automated weapons systems are designed to fire automatically at a target when predetermined parameters are detected[BH12].
- The robot has an autonomous "choice" regarding selection of a target and the use of lethal force[Hey13].
- Autonomous weapon systems, are weapon systems that can learn or adapt their functioning in response to changing circumstances[Kel11].
- Systems that operate in an open, unstructured and dynamic environment, as opposed to automatic or automated systems that operate in highly constrained predictable environments[Hey13].
- An autonomous weapon can loiter, search out targets, identify appropriate targets, prosecute an attack and report the results[Anz03].
- Autonomy is the capacity of acting without the direct intervention of humans and controlling its internal state and actions[SAB12].

- True autonomy should be considered as self-governance, with a state or condition that enables independent decision making[Twe13].

These definitions are either congruent or broadly consistent with the fuzzy demarcation we have proposed in our spectrum of automation. However, we would point out that we would consider a system that fires automatically at targets with predetermined paramaters to be an Automatic System, or at best an Automation. While Autonomous Systems may appear to operate independently, a functional Autonomous System will always be highly responsive to, and be constrained by, its environment.

3.2 Types of Control

Hardin and Goodrich[HG09] performed experiments in managing large scale simulated robotic teams performing a search and rescue task. Unfortunately they introduced terms for types of control which they have labeled autonomy, the distinctions are useful but the nomanclenture somewhat confusing: They compared operation under *adaptive autonomy* where the automation does all work allocation (between the automated system and the human) with *adjustable autonomy* where the human operator allocates all work and with *mixed initiative* where task allocation is a joint responsibility and both human and autonomous agents can request the other to perform work. Their main finding was that mixed initiative resulted in better performance than either adaptive autonomy or adjustable autonomy provided that (a) Agents are able to make progress towards the main goals in most circumstances while waiting for further human input and (b) the operator interface affords quick overall system state comprehension and retasking of agents. If these are provided, the overall system is less susceptible to higher workloads.

Clearly mixed initiative is the most general and promising of these control modes and happens to fit into the concepts of UC2.

4 Artificial Autonomy

> A human being should be able to change a diaper, plan an invasion, butcher a hog, conn a ship, design a building, write a sonnet, balance accounts, build a wall, set a bone, comfort the dying, take orders, give orders, cooperate, act alone, solve equations, analyze a new problem, pitch manure, program a computer, cook a tasty meal, fight efficiently, die gallantly. Specialization is for insects. Robert A. Heinlein[Hei73]

Our definition of the Autonomous System end of the autonomy spectrum begs the question: "To what extent is Artificial Intelligence (AI) required to produce an Autonomous System?"

A significant enabler of autonomous systems as we defined them is artificial intelligence and machine learning. While not strictly a requirement, artificial intelligence and machine learning does significantly improve the chances of a computer system handling highly dynamic data and dealing with the unexpected. While we do not wish to get into argument about strong versus weak AI, some level of AI or at least the ability to "learn" is clearly helpful in recognising unexpected patterns hidden in the data or the environment and in deciding what to do.

If we replace the word "human" with autonomous agent in the above quote, it is easy to see the challenge inherent in generating strong and general artificial intelligence. A truly autonomous artificial intellect, capable of such things as interpreting the nuances of the Law of Armed Conflict (see section 6) for instance, would be able to deal with the general and the specific simultaneously and weave them in to a gestalt greater than its constituent parts.

Currently, AI techniques are nowhere near achieving this. Although many specialised systems can give the appearance of intelligence in specific domains, the basis of their decisions is often quite alien to human modes of thinking. For instance successful machine translation systems have used statistical correlation of context to determine the correct translation of a word like "kitty" (cat or money pool) depending on context without any of the comprehension of meaning that humans would use to perform the same task. Similarly many correlations detected in Data Mining would never be detected by a human because they would simply not think to ask the right question — it is the computer's ability to run a large number of unexpected correlations that leads to useful information.

There have been various schools of thought about how best to approach AI. Originally the objective was to create intelligence in the same way humans do. This has proved difficult and elusive. Once researchers abandoned this path in favour of statistical approaches and machine learning, things improved dramatically. Likewise, attempting to create systems that exhibit human type autonomy is more likely to delay rather than accelerate the achievement of useful Autonomous systems. The availability of big data has allowed systems to learn things that make them *appear* intelligent. Traditionally, AI has be divided into two main paradigms: rule based systems which are programmed by humans and give accurate answers to some questions but the whole or even micro world can't be programmed in; and machine learning which give "smart" answers but they might be statistically biased towards the wrong world view e.g. "cats are most likely to eat cat food"

which is both correct and a false description of what cats "really" eat. The problem with such systems that have learned statistically or by other means is that there is no general way for humans to see how the conclusion was arrived at[Hea13], and moreover humans will only notice an anomaly in certain contexts with which they are familiar. Nevertheless, many human decisions are made using the System 1 mode of thinking which is far less than conscious[Kah11] and yet we have developed trust in particular humans ability to make correct decisions in this way. Such statistical learning systems can also amplify existing human biases and errors in source data through correlation. The issue here will not be whether having AI will make a system autonomous, but in how humans can trust or certify a system they do not intuitively understand when tasking it to make critical decisions autonomously.

More recently, new approaches have emmerged. The DeepQA project[FBCC+11] combined multiple AI approaches in a highly parallel pipe-line architectue to answer Jeopardy "questions"[8] eventually winning 64% of matches against high quality human players. As there are penalties for wrong answers, the goal of the project was to increase the confidence in answers, their precision and speed of generation. Answers are generated in parallel and their confidence level is used to rank them as well as providing a threshold before attempting to answer. Questions result in the generation of many hypotheses as to its meaning using multiple methods. These are filtered and then reinforced with evidence supporting them which are later scored. Answers are also generated through a plurality of differing techniques. The success of the project was a result of not only an appropriate extensible computation architecture but also to highly disciplined engineering and experimental methodologies that allowed 5500 independent experiments (each averaging 2000 CPU hours and 10GB of error analysis data) over 3 years. The final system had over 2500 compute cores with hundreds of software components. A meta-learner is used to system with more sophistocated hierarchical models as it continues to be used. However, the success of DeepQA also relied on many researchers tuning the system and developing new algorithms to approach specific problems and implementation issues.

The object of machine learning is to predict (i.e. "understand or recognise") future correlations based on the experience learned from a sample training set. The success of such predictions depends both on the appropriateness of the learning algorithm for the data and problem, and on the quality of the training sample[MRT12] which must also be available. Further, as there is a trade off between accurate recognition on one hand and speed and memory efficiency on the other, it is often preferable to choose an algorithm or representation that misclassifies a few inputs for the sake of space and time performance. In addition, some errors are unavoidable because learning always models reality. Unlike popular concepts of AI, we should expect automations to make the occasional error. As such Defence may need to choose algorithms so that they err on the side of caution. Mohri et. al. provide a modern coverage of most types of machine learning excluding graphical models and neural networks along with advice on how to tailor and validate the quality of learning systems to specific requirements[MRT12]. An issue remains that training neural networks still requires deep knowledge of the training issues related to the specific problem and tailoring of the approach to match the characteristics of the domain. For instance Sutskever and Hinton[SH10] describe specific technics for training neural networks when

[8] Jeopardy requires the contestant to produce the right question for an answer. Here we use question to indicate the challenge, and answer to indicate the response.

dealing with problems whose answers have long term dependancies.

In addition, it is recently becoming clear that solving complex problems with artificial intelligence requires the combination of intelligence at multiple levels. Salakhutdinov and Hinton[SH12] demonstrate that multi-layered deep nural networks combined with specific training techniques can cope with problem domains that contain hidden variables and millions of paramaters to produce solutions to far more complex problems such as recognising or generating depictions of complex 3D objects rather than handwritten letters. Brinkworth and O'Carrol[BO09] show how combining low computational power at multiple levels of comprehension with feedback, robust and accurate detection of global motion can be achieved.

There are many approaches to weak and stronger AI, including Knowledge-based systems, Rule Based Systems, Artificial Neural Networks, Deep Neural Networks, Knowledge based systems, Beliefs-Attitude and Intent based systems, Case-based Reasoning, Fuzzy Logic or Baysian based systems, and General Game Players. However, there is no reason why these should be mutually exclusive as shown by DeepQA and indeed even selecting few of these approaches appropriate for the problem domain at hand is most likely to provide a robust system with fewer decision errors. The example given in section 5.2 demonstrates how various AI techniques can cover each others' failings.

4.1 Agent Programming

Minsky[Min88] introduced the idea of viewing the human mind as a vast society of simple processes leading to complex behaviours we regard as autonomous.

Wooldridge[Woo09] notes that from now on distributed and concurrent systems are and will be the norm in computing, causing some researchers and practitioners to "revisit the very foundations of computer science, seeking theoretical models that reflect the reality of computing as primarily a process of interaction." Such interactions may be between humans and machines or among machine components or indeed among humans. He also asserts that the history of computing has been characterised by increasing ubiquity, interconnection, intelligence, delegation and human-orientation. Wooldridge adds an important rider to definitions of autonomy: "The need for computer systems to *represent our best interests* while interacting with other humans or systems." He gives a two part definition of a Multi-Agent System:

> An agent is a computer system that is capable of *independent* action on behalf of its user or owner[Woo09]
>
> A multi-agent system is one that consists of a number of agents, which *interact* with one another, typically by exchanging messages through some computer network infrastructure[Woo09].

Multi-agent systems do not necessarily assume that all agents serve the same "master" allowing for agents which have competing goals. This reflects the reality in C2 of coalition operations in which each coalition member may have different goals that are not completely aligned, and also relationships with NGOs, contractors and other suppliers which usually

will have quite different and at times contradictory goals to the military. In addition, each agent can be implemented in its own manner and use its own representations.

Multi-agent systems potentially have "the capacity of solving problems that are so large that a centralised agent would not have enough resources to do so" while avoiding resource bottlenecks and introducing critical points of failure[SAB12]. Moreover distributed multi-agent systems can deploy agents to the correct node to handle specific devices, fully or partially interpret their output and provide the overall automation with high rather than low-level inputs.

Sato et.al. review the many possible uses of multi-agent systems to support Distributed Communities of Practice - i.e. professional interest groups both within and across various organisations[SAB12], describing systems designed to assist new comers in integrating with the community by finding information that will bring them up to speed with the group and other systems that help manage such groups. Such systems could be useful for implementing mutual training and help in any C2 systems.

Many autonomous systems will be distributed by their nature and so naturally consist of multiple agents. C2 systems that integrate sensor and logistic data span different locations and will gain processing power and conserve network bandwidth by locating processing as close as possible to the source of each type of data while enhancing survivability. Bio-mimicry and experiences with centralised computing services suggest that response time for platforms would be enhanced by moving "reflexes" close to the site of action while simultaneously reducing the load on higher level autonomous systems. Moreover, there are software engineering and system management advantages in building autonomous systems as multi-agent systems. Multiple agents allow for:

- A natural way of placing some computing tasks on multiple distributed pieces of hardware, allowing agents inherent communication abilities to be used for interoperating.
- A separation of concerns so that systems with many tasks and functions can be broken down into separate, logically self contained agents.
- Incremental testing of system functions, one agent at a time.
- Potential reuse of agents to create different super-systems.
- Gradual, controlled upgrades of systems by adding or replacing a few agents at a time.
- Allowing different agents to represent the interests of different parties.
- Utilising heterogeneous approaches to AI with each approach being confined to one process or a set of processes.
- Their constitution as a group of communicating processes lends itself to event driven programming, which is a useful paradigm for responsive C2 systems.

While not all autonomous systems would best be implemented using a multi-agent architecture, the multi-agent concept is a good starting point from which to conceptualise

autonomous systems and may be a natural way to implement them provided interprocess and software development overheads in terms of building up infrastructure can be minimised. Some agent frameworks have failed for these reasons. Their distributed nature and ability to act on behalf of different parties or represent different conflicting intentions make multi-agent systems a natural fit for implementing frameworks for UC2.

Carrico and Greaves[CG08] describe DARPA's Advanced Logistics Project(ALP) and Ultralog project built on a Cognitive Agent Architecture called Cougaar. The advanced logistics project aimed to radically reduce forward deployed inventories and contingency supply using and End-to-End logistics system. It used agents to implement Automated Logistics Plan Generation, End-to-End Movement Control, Execution Monitoring and Rapid Supply & Sustainment. It developed e-commerce interfaces in conjunction with industry. In moving to the Ultralog Project, the objective was to increase the robustness, security and scalability of the system all while moving from a protected network assumption to the assumption of operating under intense IW attack including targeted and random losses of up to 45% of total CPU, network bandwidth and memory. The requirement was to suffer not more than 20% capability degradation and not more than 30% performance degradation. This was demonstrated in the final year of the project.

> Operationally, the most important military impacts of distributed agent systems derive from their ability to integrate data not just globally, but also across the three major levels of military engagement: strategic, operational, and tactical[CG08].

From this author's point of view Ultralog has provided an important proof by construction that Multi-Agent Systems are capable of creating highly scalable, survivable, secure and effective C2 systems which process large significant amounts of data, build significant plans and monitor their execution with predictable failure behaviour and recovery and minimal human intervention. The system is written completely in Java. As the system grew larger, additional management agents (Autonomic agents in fact, see section 4.3) had to be added to the system to manage and stabilise it. These insights are useful although it is unclear why Ultralog was not deployed operationally given its apparent succeses.

The software architecture is available commercially under the brandname ActiveEdge [Sof13].

At DSTO, several agent frameworks have been used to explore C2 systems for varying degrees of autonomy. The ATTITUDE[LR98, Lam99] agent programming language uses an extended Beliefs, Desires and Intentions (BDI) framework to facilitate contextual reasoning and reasoning despite uncertainty. It was designed to support event driven systems and information fusion. Beliefs and desires are used to invoke plans or routines which in turn generate new beliefs, desires and intentions. Importantly, ATTITUDE has the inherent capacity to model both its own beliefs, desires and intentions (explicit reflexive self-awareness) and those of other agents (computer based or human) within the environment. Moreover, using the MEPHISTO framework[LN08] it makes time a first class entity in its programming clauses to allow humans programming it to reason about time in a more or less natural way allowing contextual reasoning and what-if reasoning about the future. Although systems based on ATTITUDE were never fielded, systems based on

its non-multi-agent predecessor Metacon were. DSTO has also used the JACK[JAC03][9] third generation agent system. JACK evolved from PRS and dMARS and also uses a BDI model and a teams model. JACK had the advantages of lightweight, capable of deployment to PDAs, implementing a security model and allowing deployment to a single node or a network.

4.2 Embodied Agents

An *embodied agent* is one which can sense its environment, and control actuators or other leverage points that directly or indirectly alter the state of the agent's environment. Embodiment does not have to be a physical body, it is about sensing, altering its environment and being affected by its environment in a generative feedback loop. While for a robot agent embodyment means physically interacting with and sensing its environment; for a software system optimisation agent embodyment would entail being able to measure and sense attributes of the software enviroment such as system load, available memory and being able to alter them by changing say process priorities, scheduling policies and enforcing process migration.

While Tweedale states "At present machines can solve problems or achieve human-like functionality; however they are not intelligent, they are merely making smart decisions" [Twe12], a significant number of researchers have come to the view that embodiment is a necessary but not sufficient condition for real intelligence. For instance Brooks[Bro90] produced the Subsumption Architecture to "tightly connect perception to action, embedding robots concretely in the world." Likewise, Sporns[Spo09] argues strongly that true intelligence is situated and a result of embodied systems that affect and interact with their environment, "We found that coordinated and dynamically coupled sensorimotor activity induced quantifiable changes in sensory information, including decreased entropy, increased mutual information, integration, and complexity within specific regions of sensory space." In other words the quality of information to learn from is enhanced by information exchanges between coupled systems.

4.3 Autonomic Systems

An *autonomic system* mimics the anatomy of living animals. It consists of sensors and other measures to assess the state of the system, and a control system to regulate the state and behaviour of the system itself. Its objective of providing both system stability/optimisation and reconfiguration for mode changes mimic homeostasis and diurnal rythms in animals. While currently relatively rare at the application level, in some industries such as civil aviation invest considerable resources into system health monitoring - allowing airlines through telemetry to become aware of in-flight faults in real-time so that repairs can be expedited once the plane has landed. Autonomic systems should not be considered to be limited only to physically sensing the system, they may include or be entirely composed of sensing the software state. Murch defines an autonomic system as:

[9] Reference no longer available, cited in [EBRW07]

> [The] ability to manage your computing enterprise through hardware and software that automatically and dynamically responds to the requirements of your business. This means self-healing, self-configuring, self-optimising, and self-protecting hardware and software that behaves in accordance with defined service levels and policies. Just like the nervous system in the body, the autonomic computing system responds to the needs of business. [Mur04].

An Autonomic System is a special case of an Embodied Agent in which the environment may be purely virtual or consist of the system itself rather than being physical and external.

4.4 Artificial Autonomy and C2

While autonomic awareness is not necessary to create an autonomous agent, having autonomic awareness greatly enhances a system's ability to self-monitor and self-correct and hence greatly increases the period of time, or the amount of data it can process between interactions with humans. Sensibly applied, autonomic awareness enhances the level of autonomy a system can exhibit. It can enhance C2 Autonomous Systems by making them more robust in the face of attack and relieve operators from managing the system rather than working towards C2 objectives.

Experience with Ultralog[CG08] shows that as autonomous systems become more complicated internally, self management processes need to be introduced. Therefore autonomic awareness, at least of software performance, becomes essential for effective autonomous systems once they reach a certain level of complexity. Moreover, autonomics is essential for determining when the inputs to the system cannot be trusted. Both robots and C2 sensor systems should detect when sensors are saturated or otherwise past their limits of accuracy, so that the system can take action to: "back off" and avoid damage to the system itself, redirect its sensors to relieve the out of bounds condition so that it obtains more data it can reliably operate on, and replain its strategies to find other ways of satisfying its grand intent while avoiding areas in which it cannot operate effectively. As such autonomics provides not only the ability to optimise a system, but to protect it and make it more autonomous.

5 Applications of Autonomous Systems

5.1 Decision Support

A key requirement of C2 systems is to support command decisions. Druzdzel and Flynn give a useful broad definition:

> Decision support systems are interactive, computer-based systems that aid users in judgement and choice activities[DF99].

This definition is flexible enough to accommodate both systems that users interact with to build a decision, and those that automatically produce a decision, or range of decision options for the user to choose among. Druzdzel and Flynn also offer the insight that "Because DSSs do not replace humans but rather augment their limited capacity to deal with complex problems, their user interfaces are critical. The user interface determines whether a DSS will be used at all and if so, whether the ultimate quality of decisions will be higher than that of an unaided decision maker." While it is certainly true that a bad interface (among other things) may cause humans to reject a system, the outputs of the system also need to be useful and trustworthy. Moreover, although Druzdzel and Flynn assume that decision support systems always have humans in the loop, in the past decade some decision support type systems have evolved to take humans out of the loop and therefore become autonomous.

Cutler and Tweedale[CT07] describe two architectures for providing Decision Aids for Situation and Threat assessment for pilots and operators of airborne surveillance systems. The requirement is to track and detect targets and characterise their nature and threat. As battle spaces get busier, the workload on operators is increasing and decision support systems are being designed to relieve the operators of some workload. However the authors emphasise that unless the operators trust these systems and the system operates cooperatively and consistently, they will not be used or may distract the operators.

The primary task of Cutler and Tweedale's decision aid is to establish the intentions of the adversary so as to preempt enemy actions, and to maintain air separation of the adversary's zone of influence and the friendly force's assets. Cutler and Tweedale[CT07] note the following issues:

- Deducing intent is difficult and uncertain.
- Displays always represent the belief of the system.
- Operators cannot, in general, tell if the information presented to them is complete, accurate, or reliable[10]. This is particularly true when the adversary is trying to conceal their true identity and intent.

[10]Moreover, other results show displaying measures of certainty generally increased user decision times when these measurements are close[End03], so this is a difficult cognitive & interface problem.

Cutler and Tweedale's direct observations revealed that while every operator has the same goals and rules, each operator works in an individual manner, but their individual results should be consistent with the current operational criteria.

The first architecture, which Cutler and Tweedale implemented, was based on the Joint Directors of Laboratories (JDL) process model for data fusion. However Cutler and Tweedale were concerned about (early/minor) decision magnification that is confirmation bias which can lead to suppression of correct decisions that would otherwise be made. There was also concern that each track was being treated separately precluding the recognition of multi-track patterns or ones that do not fit the initial track allocations. The authors envisaged a multi-agent based decision aid, using teams of agents that focus on different aspects of the problem which is far more akin to how humans tackle the problem.

Paradis et. al. discuss decision support aids for Threat Evaluation and Weapons Assignment in the context of Network Centric Warfare[PBOC05]. Their problem domain highlights several areas where Automated Systems may be particularly beneficial:

- Evaluating or ranking choices of which each has a different level of uncertainty.
- Fairly maintaining a watch over multiple hypotheses of reality, given the uncertainty — something even highly experienced humans find difficult.
- Sifting through the extraordinary amount of detail that can arise in a networked environment in a feasible time frame.

Certainly, anything that can reduce cognative load is potentially useful. The human brain can cope with cope with a limited number of things to keep in short term memory, generally quoted as seven plus or minus two[Mil56], and this is even more limited when they are dynamic. For tracks it is about five. In the C2 context, the sheer volume of data from sensors and the diversity of their formats and interpretation generally requires some level of automated decision support to perform the data fusion to extract the situation picture. While in some smaller cases this could be handled by large numbers of general staff, the ADF simply doesn't have that luxury. So below a certain level, human operators will have to generally trust that these automations are working correctly. However, systems should still provide some means for operators to confirm that the outputs of the automation are reasonable. In this sense, ADF staff already have to rely fully on some automation as a basis for their decisions.

A ubiquitous C2 system available to commanders could help operational commanders by providing local threat assessments, prioritising those risks, and perhaps using simulation to evaluate several response options so the commander can see which might be more effective and what the enemy's possible and likely responses to any action taken might be.

5.2 Autonomous Systems in Market Trading

It is becoming increasingly common for financial market traders to trade using automated algorithms in order to gain a market advantage by using the extreme responsiveness of computer systems to take advantage of momentary imbalances in the market, and to "protect themselves" from sudden market downturns. However the very speed of these automated traders can, in and of themselves, destabilise the markets they operate in. On several occasions, automated trading has resulted in unexpectedly sudden and lightning fast market downturns or "flash crashes".

In October 2010 the U.S. Securities and Exchange Commission and the Commodity Futures Trading Commission released a report blaming an automated trade execution system for triggering a succession of events that ultimately caused the Dow Jones Industrial Average to drop by more than 1000 points in half an hour and affected worldwide trading[Mea10]. On 6 May 2010, global financial markets were already jittery due to concerns about Greek debt and the Euro was very much losing value against the Yen and the US Dollar. Buy-side liquidity(offers) in the E-mini S&P 500 futures market had declined by 55% over the day and the New York Stock Exchange's automatically triggered "Liquidity Replenishment Pauses" were occurring well above average levels[U.S10].

At 2:32 pm, a large trader initiated a automated program to sell 75,000 E-mini contracts valued at about $4.1 billion. Traders can choose between human judgement and automated trading. The sell algorithm they chose to execute was designed to sell 9% of the last minute's trading volume regardless of price or market timings. Execution of this program caused the largest net change in daily position of any trader on E-mini that year, and did so within 20 minutes[U.S10].

This huge sell pressure was initially absorbed by a combination of High Frequency Traders(HFTs), long term buyers, and cross market abitrageurs. HFTs typically do not hold positions across market closures nor employ significant leverage. Thus having overbought, the HFTs aggressively sold contracts to reduce their positions, sometimes to each other. This resulted in a "hot potato" volume effect as they rapidly passed contracts between each other. Overall E-mini market liquidity quickly dried up and several HFTs dropped out of the market due to their own market anomaly alarms being triggered. Meanwhile arbitrageurs sold on the SPY and S&P to regain their own liquidity transferring the shock to the equities market.

The original trader's sell algorithm reacted to the increased market volume by *increasing* its rate of selling. At 2:45:28pm, the Chicago Mercantile Exchange's "Stop Logic" functionality was triggered and the market was paused for 5 seconds. After the pause, the market gradually recovered but as prices were rising the trader's original algorithm continued to sell until 2:51pm. Despite the price recovery, significant harm occurred as many retail customers had buy and sell orders executed at irrational prices as low as 1 penny and as high as $100,000.

We now consider some autonomous market systems designs:

Sher[She11, She13] describes a method to train neural networks to trade on the foreign exchange markets, taking positions on the expected change in the exchange rate between two currencies. Forex markets are global and too large for single traders to dramatically influence. They operate 24/7 5 days a week so data is continuously converging to reflect

world events and it is harder to gain market advantage from delayed information. Neural Networks are popular as market predictors as they can effectively learn non-linear data correlation and mapping. They are also robust under data errors, can be trained online, are adaptive and can be retrained when the market shifts.

Instead of training using the more standard price list over time series input, Sher used Price Chart Input effectively showing graphs to "sensor" substrates of various resolutions. He notes that the then financial industry standard neural network learning algorithm of backpropagation does occasionally get stuck in local optima. To overcome this, Sher uses Topology and Weight Evolving Artificial Neural Networks (TWEANN) to optimise both the Neural Network topologies (and hence sensor resolutions) and the synaptic weights in tandem thus performing a global search and avoiding the local optima problem. Overall the TWEANN/Price Chart technique produced consistently better profits and were more generalisable particularly in the face of a change in market conditions.

Likewise Barbosa and Belo[BB08] describe the architecture of a "hybrid" trading agent using several AI/machine learning modules to compensate for each others' shortcomings. Hybrid intelligent systems have been shown to outperform neural networks which in turn outperform mathematical models. In order to be profitable the algorithm must maximise profit while minimising the draw down lest a margin call force them out of the market crystallising any current loss.

The architecture of their trading agent consists of an "Intuition module" which uses an "Ensemble Model" (see below) which predicts whether a currency pair will go up or down; a "A Posteriori Knowledge module" which uses case based reasoning based on past trading history to recommend how much to invest in each trade; and a "A Priori Knowledge module" which is a rule based expert system to decide when to invest or to stop a trade - it stores information that the agent cannot learn by trading such as rules to take profits and stop losses.

The Ensemble is actually a pool of models which ignores models that currently over fit the data and can quickly substitute other models that have been kept up to date and suddenly start to predict the new market behaviours. This is important since economic time series are heteroskedastic — that is they exhibit clustered volatility. Barbosa and Bello[BB10] provide extended work results on this project.

Any issues that arise with automated trading are likely to become more common in the near future due to start ups such as Quantopian democratising algorithmic trading by providing a consumer programmable trading platform[Car13].

5.2.1 Autonomous Systems Issues arising in Market Trading

Financial Markets exhibit several effects and issues that are informative for the construction of autonomous systems in the Defence context:

- The presence of autonomous systems changes the "game". They make it more difficult for humans to compete in the environment without automated assistance.
- The automations may operate much faster than humans can[11].

[11] This has implications for humans in the loop.

- The operation of multiple automations within the environment creates feedback and cascade effects. In financial markets this results in extreme market swings, in war it may result in severe escalation.
- Cascade effects may be latent and thus invisible. They are not entirely predictable because they are an emergent property of the entire population of automations in the environment. While some effects of cascade behaviour may be predictable, others may not be.
- There can be conflict when training and selecting appropriate automations between the highest levels of immediate performance and sustained performance from generalisable learning.
- Different learning strategies may be appropriate for different problems, and some problems are best served by using multiple AI techniques in combination for their complementary attributes. This becomes increasingly likely as the problems targeted become more complex.
- Reliable data provided promptly is critical to successful operation.
- The effects of failures of automation may occur far beyond the market or theatre in which they are originally played out due to other interconnections.
- Autonomous behaviours which may be individually optimal can have catastrophic flow on effects for the system - e.g. the withdrawal of autonomous traders from the market led to a sharp drop in liquidity.
- Providing circuit breakers in the environment has proved to be useful in markets. How an analogous dampening behaviour can be achieved in conflict is far less clear.

5.3 Transport

Transport provides a rich environment for a wide spectrum of automation and autonomous systems, from the simplest traffic light which operates on a fixed cycle — but which relies on the context of social norms to enforce car stopping, to automatic transmissions, cruise control, and ultimately self driving cars. The examples from transport provide many useful models.

5.3.1 Land

As described in the introduction, the simplest traffic lights are simply automatic systems responding to a few direct inputs, or operating solely on internal timers. Their main feature is a fail-safe interlock to prevent a green light being displayed to competing traffic flows at any time. More complex systems that network traffic lights together could be considered examples of automated systems. Although for many years larger traffic control networks have performed a type of C2 function bringing traffic control data back to a central control room, as traffic sensor technology improves and more computer systems are put in place to analyse traffic and redistribute load the most complex of these systems could be considered to be automations.

Muhrer et. al. [MRV12] evaluate an automated system PADUS designed to avoid car rear-enders which are caused by either a lack of expectation of the need to brake or not paying attention. Forward Collision Warning systems exist that guide the driver's attention and are effective for both distracted and attentive drivers but sometimes they still don't prevent collisions. PADUS adds autonomous braking including strengthening driver initiated under-braking. One concern was that the existence of PADUS might lead to further inattention, if the driver becomes an observer monotony prevails. Researches wanted to know: "Would benefits of PADUS be counteracted by reduced driver engagement leading to homoeostasis in the overall driving performance?"

An experiment was used to assess whether PADUS can increase traffic safety. Thirteen accidents occurred, all without the PADUS in operation. In terms of reaction time there were no significant differences between the PADUS ON or OFF state. There was also indication of a significant adverse effect, with PADUS ON, drivers had significantly higher velocities in inactive street areas risking greater harm to any unseen pedestrians, and drivers were more distracted. However, the overall recommendation of the paper was to adopt such systems to reduce collisions.

Other work seeks to improve safety for unmanned ground vehicles, in part compensating for the delays induced by teleoperation[AKIW13]. Anderson et. al. explore semi-autonomy in which both the computer and the human are assigned tasks they are "naturally" good at in order to obtain synergies. Noting that reactive safety systems such as anti-lock brakes rely on the human to foresee, judge and respond correctly, and that each predictive safety system usually moderates only one direction of movement, they designed a system between such tactical assistance and full vehicle autonomy offering both strategic planing and "intention preserving" control support. The system maps out areas of safe travel giving the operator as much freedom as possible while guiding them to remain within the safe regions and their paper describes the calculation of go and no-go zones in

detail. The system provides a visual overlay of the safe zones over the video image, change in steering needed visual indicators and haptic feedback through the steering wheel. The system will use as much haptic feedback as necessary to force the operator to avoid a collision. Results from experimental testing are both encouraging and perhaps surprising. 95% of operators reported greater confidence *and* control with the system enabled compared to the control condition. The computer took only 45% of the control authority overall, and operators were significantly more moderate in their control inputs.

The DARPA *Grand Challenge* required autonomous vehicles to reach the end of a 220 mile off-road trail, and the DARPA *Urban Challenge* required autonomous vehicles to travel 60 miles in 6 hours or less while also safely following and overtaking other vehicles, obeying give way rules at intersections and driving through car parks. This in turn required detection of vehicles in every direction and differentiating moving objects from static ones. Darms et. al. [DRBU09] describe the vehicle detection and tracking system of CMU's winning entry *Boss*. Their software architecture separates the sensor and fusion layers to allow for sensor substitution modularity, and to collect data from 13 different types of sensors including lidar. The system separates tracking and classification, generating hypotheses and testing them. Perception is performed using a measure, perceive, understand paradigm with particular care taken to detect and deal with the types of errors e.g. false positives that can occur at each stage.

Their are now several prototype self-driving vehicles. The most famous of which being the Google car. The software operating the car, Google Chauffeur currently requires the human to drive in suburban streets, taking over only on freeways and operates in a mixed initiative mode[Fis13]. The US states of Nevada and California have allowed autonomously driven vehicles to operate on roads provided a licensed driver is present to take over should the autonomous system fail at anytime. Nevada also requires a second human driver in the passenger seat. The Google car is effectively in closed beta test with an expected mean time between hand over of driving to the human of 36,000 miles in the final product. Interface issues are still an area of open questions, for instance noone knows what is the best way to hand over driving to the human in cases of automation failure including how much warning to give[12]. So far the only accident with the Google car occurred when the human was driving and on the basis of extrapolated miles travelled without a chauffeur operated accident indicate that the auto-chauffeur is at least as good as a human. However, Google down rates this figure because those miles have been in less challenging environments. Data rates needed to navigate can be enormous. Google's self driving car uses about 1 GB / second[Hea13], and sensing is performed primarily by a lidar - a spinning turret with 64 range finding lasers costing much more than the car itself. Google thinks the self driving car will be available in 5 years, but this will require a massive investment to bring down the cost of the lidar which Google sees as a primary enabling technology to generate sufficient situation awareness of the environment. Nevertheless, some of the biggest challenges for autonomous cars may be legal. Many laws imply that a person must be in charge of the vehicle, and there is a chicken and egg issue between legal certainty and mass manufacture of cars[Fis13]. Google estimates that the level of safety is currently orders of magnitude less than it needs to be, while many industry observers have noted a virtuous circle of efficiencies that could emerge from removing humans from the

[12]Clearly there is a potential trade off between earlier warning and the auto-chauffeur resigning when the error might otherwise soon be recovered.

steering wheel. These include, reduced fuel consumption, whole of network optimisation etc. Another major issue with autonomous cars is keeping the driver engaged to take over from an autonomy failure, and how will people learn to be good drivers — or even to drive at all — when they don't actually drive? For these reasons, Google is aiming for a fully autonomous car without the need for a human driver.

The VisLab *Intercontinental Autonomous Challenge* had a single goal: to design vehicles that can drive autonomously along a 13,000 km route without human intervention. VisLab's BRAiVE vehicle[BBD+13, BBC+10] completed the challenge driving from Parma, Italy to Shanghai, China but had to finish the final section with human driving to meet the deadline of the World Expo after being delayed at several customs offices. Their system splits movement into navigation, manoeuvre and control levels. The latter being short-term trajectory planning using only local sensors. Other features include mixed initiative using the human to provide "Ground Truth" and a high level of vehicle autonomy.

Duff[Duf13] presents a view of operating mining machinary underground. He makes the point that moving operators off the vehicles requires not only highly effective sensors, but may make the area around remotely operated vehicles unsafe for humans. He notes that path to automation is much longer than people think, with the time to create an autonomous capability only nine months or so but a total of ten years to develop and prove the safety of the system. Moreover removing humans creates a greater requirement for machine reliability and the consequent loss of gossip has implications for situation awareness that must be compensated for by the system. An interesting observation was that in autonomous systems, the flow of information is as essential as the flow of oil in standard operating environments. Another key capacity for Duff was the ability to create a map of the area in a short amount of time, which was previously infeasible. This allowed maps to me created often and change analysis to be applied. Clearly this is very close to some of the funtions that C2 already performs.

5.3.2 Air

Automation has at times been a contentious issue in Air travel. Several air accident investigations have laid blame for loss of life on failures of automation or more commonly on pilots not being fully aware of how automated systems on board are behaving. For instance, investigations into the Asiana 777 crash at San Francisco airport on 6th July 2013 include whether interactions between the 777's auto-flight and auto-thrust modes led to pilots not realising that the auto-thrust mode had hibernated during a visual approach and so would not increase power to prevent a stall as expected[Cro13]. Conversely, overall aviation safety has continued to improve since the introduction of more automation. Langewiesche[Lan09] describes the US Airline's flight that successfully landed an A320 in the Hudson with no loss of life and only minor injuries, after both engines were destroyed by bird ingestion. He argues that such a landing would have been far more difficult, if not impossible, in a 737 which lacks the A320's flight envelope protection automation which allowed the pilot to concentrate on keeping the plane level for landing on water rather than preventing a stall.

Autopilots in our taxonomy fit squarely in the automation category because while

they handle lots of data, the data is well understood at design time and the mapping from inputs to outputs is well defined at design time also. In fact autopilots disengage and return control to the human pilot should input data go out of design range. This occurred on the ill fated Air France Flight 447 in which inconsistent indicated airspeed (IAS) readings were obtained and the resulting handover to human control did not maintain situation awareness[BEA12] contributing to the disaster. The Traffic Collision Avoidance System(TCAS) which mutually interrogates the transponders of nearby aircraft for local flight heading information and if any of the set of aircraft are on a collision course negotiates a set of instructions for each endangered aircraft to achieve mutual separation, also operates within a very constrained operating environment with highly defined input and so is another example of automation under our taxonomy. Furthermore it only resolves conflicts by demanding vertical separation. Due to accidents caused by prioritising Air Traffic Control commands over TCAS directives, pilots are now required under legislation to obey the commands issued by TCAS, even against Air Traffic Control directives unless they have good reason to believe doing so would endanger their aircraft. This illustrates how it is possible for an automated system to have more timely and accurate information than remote operators, and why devolving some decisions to local automations can improve both timely response and correctness.

DSTO's Air Operations Division (AOD) adopted first dMars then Agent Oriented Software when it began to explore beyond the technology platform to the whole system capability including the human components of air-combat[CGHP09]. Pearce et.al. [PHG00] outline operational simulation in AOD used to answer specific questions about platforms, component capabilities and to rehearse tactics. In the virtual environment, both pilot-in-the-loop and agent-pilots are used. A key challenge is recognising pilot plans while the pilot is executing them. The authors note two distinct requirements for the AI:

1. extract patterns that can lead to recognition from the environment, and
2. inferentially reason about the nature of the observations made.

They comment that their system does not try to recognise everything, rather looking for what is possible given what is expected. Likewise there is feedback in the initial observations also, in that what patterns to look for and their interpretation is governed by the context. This author conjectures that a possible problem with this model is that it could make the automation particularly vulnerable to new and unexpected tactics in a real-world scenario. Papasimeon Et.Al. [PPG07] give further details about a mixed human/agent simulated flying virtual environment for training , wargaming, providing acquisition advice and tactic development. They point out that the affordances generally provided in flight simulators for pilots, such as (often analogue) graphic readouts, are quite useless for agents and so integrating agents and humans requires the parallel maintenance of both a graphic environment for the humans and a semantically annotated database for agents complete with "callable" affordances for automations. They provide a good example of how various AI functions can be implemented in separate agent components to simplify program logic.

BAE Australia has developed a UAV that can land autonomously in emergency situations [Nut13]. The system uses imagery to select a landing site and land safely without a remote pilot or external navigation aids such as ILS or GPS. The UK Taranis combat

UAV and Mantis medium-altitude UAV use autonomous mission systems developed in Melbourne using their Kingfisher II test and demonstration platform. The system is being considered as a aircraft carrier landing aid for manned aircraft. Uniquely worldwide, BAE has negotiated shared use of airspace in Australia with civilian air traffic the only incident being a pilot who refused to take ATC direction and would have run into the UAV had BAE not taken control to abort its approved prior landing. BAE state that the biggest hurdle they have faced operating in shared civil airspace is public perceptions of safety.

For Unmanned Air Vehicles, Tweedale[Twe12] identifies several barriers to increased levels of autonomy, and to achieving a reduction in support personnel for each mission. There is no standard for launch, recovery and control systems and so each new system requires different training. Moreover, each payload that may be flown can be completely different, each requiring a unique skill set. Consequently, far more operators are required than the one operator controlling multiple platforms desired by the military when automating. In order to achieve improved levels of staffing, Tweedale cites the need for equally autonomous mission planning and execution tools. He makes explicit that the pilot being replaced performs several roles with attendant responsibilities, and that a significant amount of the pilot's situation awareness is gained through the five human senses which are all inadequately transmitted to remote locations along with a time lag and as such the pilot in the plane constitutes situated awareness and local processing power. This is a formidable combination to replace by artificial autonomous system.

5.3.3 Maritime

Gupta et. al. [GHZ12] give details of an algorithm for adaptive cooperative exploration of unknown underwater environments with very limited supervision. Using a robot simulator, they validate the efficiency of their algorithm which must adapt to surprise obstacles while using statistical mechanics to ensure complete coverage of the search area. A combination of local and global navigation is used, each vehicle is allocated a sub-area to search and upon completion of its task will communicate to see if it can assist in searching a neighbour's area.

DARPA has recently called for proposals to build a fleet of underwater drones to store and deliver payloads such as supplies to remote locations stealthily. Project Hydra as it is called is also concerned with launching UAVs into the air from underwater "motherships". DARPA is already working with Lockheed Martin to build a fleet of land and air drones to deliver cars and even containers of soldiers[OG13].

5.3.4 Space

Deep Space 1[Var14] launched on 24 October 1998 featured RAX or remote intelligent self-repair software which was the first autonomous spacecraft control system. It featured AI which planned spacecraft actions and diagnosed spacecraft faults using a model-based system. Its planner system (EUROPA) was reused as a ground based system for planning Mars Exploration Rover activities and was extended as EUROPA II to support the Phoenix Mars Lander.

Steiner and Athanas[SA09] describe the unique requirements of space systems which exhibit the need to handle unforeseen environments and simultaneously enjoy tight and certainly finite local resource constraints. Increasingly there is a desire to apply autonomy in pervasive and systemic ways to achieve component reusability and robustness. Space systems can experience long periods of total network isolation in which the system must continue to operate. Traditional direct (remote) control suffers from excessive round-trip times and low bandwidth. Operators have low visibility into the systems. Mission goals can conflict with ensuring the function and safety of the space system.

Steiner and Athanas[SA09] introduce autonomous control of digital computer hardware for dynamic circuit configuration. Effectively an operating system for hardware maintenance, autonomy allows the system to adapt gracefully to changing [environmental] conditions. The system uses field programmable arrays to achieve partial runtime reconfiguration of the hardware, optimising it for current tasks, even installing new capabilities elaborated from libraries. They describe a hierarchy of autonomy ranging from the designer uploading system configurations; through instantiating prefab system libraries; creating "made to measure" larger hardware structures using component templates; observing its own performance; inferring new hardware pattern requirements; asking for specific assistance; and finally learning from its successes and mistakes.

6 Law of Armed Conflict(LOAC)

There is a considerable body of international law governing armed conflict and primary among them are the Geneva Conventions.

The various Geneva Conventions[Int13] can be briefly summarised for reference as follows:

Convention I Protects soldiers and medical personnel and grants the right to proper medical treatment and care.

Convention II Extends protection to naval personnel and hospital ships.

Convention III Defines prisoners of war, extends protection to them and outlaws their torture to extract information.

Convention IV Extends protection to civilians, civilian hospitals and defines how occupied populations are to be treated.

Additional Protocol I Clarified terms, further restricted treatment of protected people and added new requirements for treatment of the deceased, cultural artefacts and dangerous targets (e.g. dams and nuclear installations).

Additional Protocol II Clarifies "humane treatment", gives certain protections to those charged with crimes during war and provides new protections and rights for civilians.

Additional Protocol III Adds the religiously neutral Red Crystal to the Red Cross and Red Crescent symbols used to protect humanitarian staff and installations.

Recently, the UN Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, issued a technically literate report[Hey13] on the extraordinary challenges autonomous systems pose to International Humanitarian Law and International Human Rights Law. It focuses on "Lethal Autonomous Robotics(LARs)" which are at the extreme end of the spectrum of smart weapons in that they decide for themselves what are targets, as opposed to simply locating a, perhaps displaced, human selected target.

The Rapporteur expresses strong concern that moves towards autonomy weaken the dangers of going to war and result in a psychological detachment from the decision to kill. Remote "drone" operation and stand off weapons have already increased the distance between weapons and their vulnerable users to the extent where the operators of such weapons are virtually invulnerable.

At the core of LOAC are the intertwined (and here very briefly elaborated) principles of:

Military Necessity That the attack and type of weapons used were necessary and not excessive to progress the military objective.

Humanity That the humanity of combatants and civilians involved is respected.

Distinction Belligerents must distinguish between combatants and civilians and avoid avoidable harm to civilians and civilian infrastructure.

Proportionality That incidental civilian injuries must not be clearly excessive in relation to the anticipated military advantage

6.1 Legal & Moral Responsibility

Historically, who is permitted to kill has always challenged society. Heyns[Hey13] argues that LARs pose immense challenges to not only the morality and ethics that inform and build International Humanitarian Law(IHL) but to the accountability that makes such law enforceable and hence an effective deterrent. He therefore pleaded for States, international organisations and civil society to undertake urgent, intense and informed dialogue.

In terms of Command and Control Pigeau and McCann[PM02] describe the three dimensions of Command Capability as "Competency", "Authority" and "Responsibility". The extent to which individuals can exhibit these attributes differs between humans and autonomous systems. An autonomous system can be physically competent, in some dimensions such as strength, perception resolution and what can be percieved far more so than a human; while in other dimensions such as agility and flexibility in new contexts it would be usual for humans to have the upper hand. Likewise, both humans and autonomous systems can be intellectually competent although again each clearly superiour in different areas. However, there are currently real questions as to the level of autonomous systems' emotional competency and interpersonal competency. Authority can be bestowed either legally - in the military by their rank, or personally - by the level of trust they've earned especially from those of lower rank. The level of authority various ranks of the military will be prepared to give autonomous systems will be commensurate with the level of "trust" the particular autonomous system has earned by being both reliable and predicatable in its operation. This is simlar to the way humans are promoted or removed from the military, but may differ in the detail of how trust is earned. See section 8.2. The concept of responsibility, however, does not directly translate. There are two pertanent aspects of human military responsibility: ensuring that a command is executed and accepting the punishment if one operates outside of military or legal guidelines. There is no way to punish a computer system, it cannot feel pain. However, the point of punishment is not to make the infringer suffer but rather to ensure discipline is maintained and infringements do not reoccur. Since discipline is unlikely to be an issue with Autonomous Systems, the corresponding real issue for them are methods of ensuring that decision errors are prevented from reoccurring in future operation and other instantiations of the system are also corrected. These issues are covered in section 8.4.

Robots and automated systems use a sense-think-act loop, and the level of autonomy is differentiated by the level of human in-the-loop. On the low end are UCAVs with a human in-the-loop, and at the other extreme are LARs with full autonomy and humans outside-the-loop. All parties see humans remaining in the wider loop where they set the ultimate goals of a weapons system and activate/deactivate them as necessary. Supervised Autonomy refers to humans monitoring the loop and countermanding decisions as necessary - but this assumes that the human has the reaction speed to do so within the wider system's time constraints.[Hey13].

Heyns points out that autonomous decision making should not be confused with "free will" or "moral agency"[Hey13]; but he also notes that machine systems do not act out of

fear, revenge, panic, anger and without explicit programming would not torture or cause deliberate suffering including rape. Importantly, he argues that IHL requires deep human understanding including common sense, a wide appreciation of context and estimations of motivations which humans are fallible at but robots are incapable of.

Heyns[Hey13] doubts that autonomous systems could determine if someone was *hors de combat* or recognise soldiers in the process of surrendering. He does acknowledges that robots risk less being fired upon than a human and so can afford to act more conservatively by not firing until fired upon. Although sceptical Heyns wonders whether it is possible to create autonomous systems that comply more with IHL than humans, and whether that in turn would create an obligation to use them just as human rights groups currently argue there is an obligation to use "smart" weapons. This author is not conviced by untested legal arguments that having "smart" weapons creates an obligation to use them. The Geneva Conventions require that actions taken in war are reasonable in that they do not cause excessive civilian loss of life in relation to the value of the military objective; they do not require that the choice of weapon is optimal, and in any case the choice of weapon is further constrained by what is currently in theatre, what may be need to be reserved for future use in the conflict and what capabilities need to be kept secret if that is possible.

The rapporteur expresses the extreme difficulty of autonomous systems evaluating proportionality[Hey13] an issue explored further in section 6.5. It is not, in general, enough to preprogram certain proportional responses to expected cases because of the LOAC requirement to consider Military Necessity, Humanity, Distinction and Proportionality together in making any decision.

Arkin et. al. propose the use of models of ethical behaviour to be incorporated into intelligent systems from the onset[AUW12]. They asked "How do we create robotic technology that preserves our humanity and our societies' values?" In recent work Arkin hypothesised that *ultimately* autonomous systems could operate more humanely than human warfighters are able to. This claim is based on superior capabilities such as faster processing speed, a broader range of sensors and using network centric warfare to gain far more vantage points with which to make a decision. They propose an architecture for ethical governance of an autonomous system that allows the "governor" to be bolted onto the autonomous system interceding between the decisions of the host system and its actuators. They expect levels of autonomy to increase over time. The governor is just one component of their wider architecture which models guilt to ensure ethical behaviour.

Reeves and Thurnher[RT13] express concern that the traditional balance between Military Necessity and Humanity that underpins the LOAC is being destabilised by too much focus on the Humanity side of the equation. They argue that overemphasis on Military Necessity results in atrocities and overemphasis on Humanity creates unrealistic restrictions that weaken the LOAC by making it too difficult to comply with implying this makes it harder to enforce. They argue that the LOAC needs to be dynamic and adjusted to take into account new situations and maintain such a balance. They give three examples of the tilt towards primacy of Humanity all relevant to Autonomous Systems:

1. The *capture or kill* debate in which recent legal argument has included the concept that force should be regulated by a "least restrictive means" analysis in which it is theorised that attempts should be made to capture before killing is used. They

criticise an ICRC guidance document[Mel09] for incorrectly treating Military Necessity and Humanity as distinct rules and suggesting that separate analysis of each be carried out and for extending the requirement to accept an unambiguous surrender to one of offering the opportunity to do so. They state that while a State may self-impose a capture obligation on its forces, such as US did during counterinsurgency efforts against al Queda ultimately carried out by targeted killing drone strikes, they do not make their internal norm a legal requirement.

2. Public discussion of the LOAC is exerting extreme pressure to restrict both the means of contemporary warfare and to limit development of theoretical advanced weaponry in the area of *Autonomous Weapon Systems*. "States conducting research on potential future weapons systems are increasingly being asked to make definitive legal conclusions before fully assessing the benefits and drawbacks of the new technology." NGOs are pressing for all autonomous weapons to be made unlawful without considering that some situations such as battles in deserts, under water, space or cyberspace may hold no direct threat to human life. Such a ban might have the unintended consequence of precluding the use of technologies which might well minimise the threat to civilians. States should conduct a comprehensive review before acquiescing to such pressure.

3. In the domain of *Cyber Warfare*, a document[Sch13] commissioned by NATO regarding application of extant international law norms to the area generated significant debate, particularly around rule 29 of the manual which states: "[c]ivilians are not prohibited from directly participating in cyber operations amounting to hostilities, but forfeit their protection from attack for such time as they so participate." While this is a summary of long settled legal argument about the LOAC, the media has taken a great interest in the idea that hackers might be lethally targeted and others have seen it as justifying drones attacking anonymous hackers.

While acknowledging the legitimacy of many issues brought up by those arguing for Humanity i.e. humanitarian concerns, Reeves and Thurnher[RT13] argue that they are often putting the law cart before the technology horse and that the LOAC is effective and agreed to by States precisely because it balances Military Necessity with Humanity.

Peter Asaro[Asa12] among others formed the "International Committee for Robot Arms Control" in 2009. He calls for an international discussion considering the propositions that "The prohibition of the development, deployment and use of armed autonomous unmanned systems" and that "machines should not be allowed to make the decision to kill people." He specifically argues against Arkin's assertion that prohibition of autonomous weapons system may be premature, unnecessary and even immoral.

Asaro defines an "Autonomous Weapons System" as any system capable of targeting and initiating the use of lethal force without direct human supervision. Quoting Kallenberger[Kel11] he points out that many military professionals including ones from the US have expressed strong concerns regarding the ethics of such systems.

> Restrains on autonomous weapons to ensure ethical engagements are essential, but building autonomous weapons that *fail safely* is the harder task.
> ...

> Because they lack a broad contextual intelligence, or common sense, on par with humans, even relatively sophisticated algorithms are subject to failure if they face situations outside their intended design parameters.

Like others, he raises issues such as who will be held responsible for any wrongful harms caused and the inability of existing autonomous weapons system to meet the requirements of IHL.

6.2 Risks of Escalation and Instability

Governments currently state no intent to field LARs[DoD11], but availability of technologies or their tangible development may change that.

Heyns is concerned that LARs lessen three natural constraints on war: Human aversion to getting killed, losing loved ones and having to kill other people. They lessen these constraints to the extent of even making deaths emotionally unnoticeable to the side deploying LARs[Hey13]. However, use of true automation would reduce war fighters exposure to combat trauma[13]. In addition, the problem of emotional distance is one that our military already has to manage effectively. Target selection personnel and commanders at HQ routinely deal with such issues.

There is a distinct danger that war, particularly low level war, will no longer be considered a tool of last resort because the populace will be less aware of it without the loss of life on the side of states deploying LARs[Hey13].

While some have argued that by this logic, States should avoid developing technology that reduces the brutality of war by providing greater accuracy and lower deaths, Heyns argues that LARs moves war closer to one-sided killing and so is qualitatively different because one side no longer participates in the danger[Hey13]. He argues that due to ongoing conflict the total number of casualties could be higher than otherwise and that the potential for targeted killing mak es the world and life less secure.

It must be remembered that one-sided killing is not unique to future potentially autonomous systems however. From Pol Pot to ISIL, dictators or groups that have extraordinary certainty about their own beliefs and inherent rightness have killed whoever might challenge their power or even ideas.

Heyns warns that LARs could weaken the role of international law and undermine international security; he therefore encourages States, international organisations and civil societies to discuss the advantages and disadvantages of taking this path. He argues that the moment for action is now as once research programs are put it place or such devices become available it will be hard to turn back the clock.

LARs are also highly asymmetric in nature, and as such are likely to provoke strong reactions from enemies who may well also choose asymmetric modes of warfare[Hey13]. Some argue this is already the case regarding drone attacks used to carry out assassinations

[13] Conversly use of remotely operated UAVs has actually resulted in higher levels of stress for remote pilots who virtually stay and witness the results of their firing than for physically present pilots who fire and leave[Dao13].

without trial or clearly identifying the target person[Sca13], and that doing so leaves the user of the weapon with diminished moral standing and the other side with strong and legitimate reason to retaliate[14].

Heyns also points out the risk of LARs being acquired by non-State actors, including those trying to overthrow a government in a coup.

Finally Heyns makes the point that at best LARs can be made to operate within minimal ethical standards, but that they have no hope of autonomous systems achieving our highest human behaviours such as compassion.

Finn[Fin08] not only argues that introducing Automation into the battlefield will increase the tempo of operations and that one nation's use or development of them will justify another's, but also that as more and more autonomous systems are deployed they will become critical to survival and thus the increasing tempo will squeeze the human out of the loop.

Asaro[Asa12] warns of risk redistribution from combatants to civilians and that automation has the potential to lower the threshold for nations to start wars and of risking the "unintended initiation or escalation of conflicts outside of direct human control."

6.3 Legal Considerations

Unlike revolutionary weapons that simply increase fire power, LARs blur the distinction between warrior and weapon[Hey13]. This is just one way in which LARs do not fit easily into our extant legal frameworks.

Heyns points out that some argue that it will always be impossible for robots to meet the requirements of IHL and IHRL, and that regardless a machine should never have the choice of life or death over a human. These people argue for a total ban on the development, production and use of LARs. However Heyns also records that others argue that, within limits, LARs represent significant military advances which may reduce casualties and be more humane. Heyns notes the wide agreement that caution should be exercised regarding LARs and that international discussion is required to form new and extra standards regarding such weapons. In particular, he is concerned that given the experience with UCAVs, LARs will be used for targeted killing. Moreover, he notes that autonomous technologies may be particularly hard to regulate given autonomous systems tend to be made from a composition of technologies which are not particularly risky in themselves and that autonomous systems have legitimate and beneficial application in civilian life. The desire for scientific progress in autonomy will result in underlying technology creep, so he argues for the international community establish a process to regulate such technologies.

However, technologies already exist that make the choice of life and death over humans. For instance mines and other booby traps have been used even though they do not discriminate between combatants and civilians as required by the Geneva Conventions, and certain automatic weapon systems such as Patriot missiles are deployed on the basis

[14]While terrorism is itself deliberately provocative, reacting to it with weapons that are viewed as in some way "unfair" can act as a recruiting tool for the enemy.

that they are defending a declared no-fly zone. The Ottawa treaty seeks to eliminate anti-personnel mines and has 133 signatories including Australia, the UK, Canada and New Zealand but not the United States, it does not cover remotely controlled mines, anti-tank or marine mines. In some cases, the placement of automatic devices involves a pre-action shoot decision based on the location of they system such as in a prescribed border area with automatic machine guns.

The rapporteur[Hey13] describes accountability of individuals and States as fundamental deterrents to war crimes and hence protection of human life. He asks who will be held responsible if autonomous systems behave unacceptably particularly given the composite nature of such systems. Will senior commanders understand autonomous systems sufficiently to be held accountable for their "subordinate"? He is most concerned that this responsibility vacuum could enable impunity from misuse of such weapons.

Equally, it is unclear how accountable certain support personnel are under the LOAC, nor whether they are legitimate targets. Finn[Fin08] notes that the LOAC states that only a combatant is permitted to directly take part in armed conflict but that support personnel for UAVs who are not combatants have become relatively common due to the technical requirements of such support. Such support personnel, including programmers and researchers, are legitimate targets under the Geneva Conventions for the period that they are actively involved in directly supporting operations[15].

Asaro[Asa12] argues that Autonomous Systems represent a qualitative shift because they eliminate human judgement in the initiation of lethal force and that IHL and IHRL would benefit from explicitly coding a prohibition on the use of autonomous weapons because it would:

- avoid slippery slopes towards autonomous weapons systems by drawing a principled bound on what can and cannot be automated
- shape future R & D efforts towards more human-centred designs capable of enhancing ethical and legal conduct in armed conflicts
- stem the potential for more radical destabilisations of the ethical and legal norms governing armed conflict that these technologies might pose
- establish the legal principle that automated processes do not satisfy the moral requirements of due consideration when a human life is at stake.

Asaro[Asa12] defines autonomous weapons systems as "any automated system that can initiate lethal force without the specific, conscious, and deliberate decision of a human operator, controller or supervisor." He acknowledges that there are precursors to lethal automated systems in terms of victim activated traps such as mines, guided missiles and some automatic defence systems. He argues that autonomous systems as the culmination of these precursors force us to think in terms of the "system" and by what process the use of force is initiated and that the critical system is the one that contains the decision making cycle in which the decision to use lethal force is made. Noting that *Supervised Autonomy* appears to create a middle position between direct human control and an autonomous

[15] According to responces to the author's questions during Red Cross IHL training course.

weapon system he asserts that the key issue is whether the system automates either the target or engage steps independently of direct human control.

> It is the delegation of the human decision-making responsibilities to an autonomous system designed to take human lives that is the central moral and legal issue.[Asa12]

In particular Asaro[Asa12] argues that having a human in the lethal decision process is necessary but not sufficient and that they have to be suitably trained, informed and given time to be sufficiently deliberative. This has implications for the level of SA afforded to them and any systems DSTO might design or research. Even so, Asaro may be unaware that most of the SA that high and even medium level commanders rely on is already machine mediated. Very little target and track intelligence is currently collected by direct human observation - some of it is assessed by human analysis, but large amounts of such information is correlated by automated systems which are generally believed by their human operators. We already trust machines for significant components of our targeting decision making because they do a more accurate and more reliable job than humans could hope to do. Asaro goes on to argue that a poorly trained person without access to sufficient relevant information and given insufficient time may perform no better than an automation.

Asaro asserts that in terms of engineering and design ethics, "intentionally designing systems that lack responsible and accountable agents is in and of itself unethical, irresponsible, and immoral," and goes on to say that "the fact that we can degrade human performance is such situations to the level of autonomous systems does not mean we should lower our standards of judging those decisions."[Asa12]

He proposes that a treaty limiting autonomous weapons would have at its core:

> the principle that human lives cannot be taken without an informed and considered human decision regarding those lives in each and every use of force, any automated system that fails to meet that principle by removing humans from lethal decision processes is therefore prohibited[Asa12].

and that the focus of the treaty must be not on the weapons, but on the delegation of authority to initiate lethal force to an automated process not under direct human supervision and discretionary control.

Asaro specifically criticises Arkin's proposed ethical governor[AUW12] as being based on the following fallacies[Asa12]:

- The principle of Distinction in LOAC is not like a sorting rule and in general a definitive categorisation between military and civilian is not possible and cannot be encoded into a set of rules. Indeed this is why countries have their own interpretation of the specific meaning of the LOAC, why the ICRC has issued guidelines about interpreting them, and why ultimately Judges and Courts are needed to arbitrate on them. In general, IHL is unlike the rules of chess and requires interpretive judgement in order to be applied appropriately to any given situation.

- The conflation of the statistical correctness of an automation's decisions compared to those of an acceptable human with the moral question of whether a computer or automated process should make the decision of life or death at all.

- That artificial intelligence has a history of finding complex problems computationally tractable only by simplifying them into fundamentally lesser problems that do not truly reflect the essence of the issue.

Asaro also argues against the inevitability of autonomous weapons systems, stating that many technologies appear inevitable only in hindsight and that even successful technologies can have many false starts and drawn out development.

Daniel Suarez[Sua13] argues for use of autonomy to collect data but not to automatically fire, on the basis that the way we resolve conflict shapes our society and that lethal autonomous robots would concentrate too much power in the hands of too few people and so would reverse a 500 year trend towards democracy. He describes three factors pushing remotely operated drones towards autonomy: overwhelming and increasing video feeds, electromagnetic jamming and plausible deniability. The last one is particularly a problem if, as he suggests, designs for autonomous drones are stolen and proliferated. He postulates the prospect of anonymous war leading to a first strike posture. He suggests a solution of a treaty banning LARs and providing no privacy for drones. Monitoring every drone via having a unique digital signature transmitted through a transponder with citizens being able to download an app to find out what is in their airspace. He also suggests civic drones to detect unauthorised drones and dragnet them to a bomb disposal facility.

What can be said about the legality of autonomous systems is that it is a multidimensional problem with various conflicting points of view, many of them valid. As research progresses, some of these points of contention will be resolved through operational examples. For instance, recent progress in Artificial Intelligence such as deep neural networks (section 4) allows far more complex situations to be handled as well or better than a human does, however building and training such systems has not yet risen to the point of being a standardised engineering process. A prime question to be determined is how we set the stanard for an acceptable shoot decision: Do we say that is must be of the same moral standard as a human one, or do we say that statistically the error rate must be below a certain low threshold? These are not equivalent measures since in the latter, we are accepting a low level of morally wrong decisions that may be more technically correct than a human might make. In terms of minimising future harm, the underlying goal of taking responsibility, the issue would be to determine an acceptable threshold of accidents or breaches of LOAC and to ensure that:

1. Autonomous Systems are not fielded without a rigorous expectation that they will not exceed this limit.

2. Fielded Autonomous Systems and any similar sytems must be immediately shutdown or downgraded until they can be redesigned whenever they exceed their expected limits or it becomes apparent that they are in danger of doing so.

3. Proper engineering processes must be in place to invenstigate and isolate the cause of each transgression, and to remove the possibility of similar transgressions reoc-

curring. Like air-safety, this will rely on strong logging and the development of engineering disciplines to understand failure and avoid hazards.

Autonomous systems also open up the posibility of modes of combat not previously envisaged: imagine an Autonomous robot that did not immediately kill the fighter but but clamped itself around their neck ready to kill them after a human had decided that they were combatants or after they had demonstrated this conclusively while the robot collar was attached. How is such a weapon to be treated legally? To what extent would persons collared in such a way be considered to be detainees or prisoners of war? They are not hostages under the standard legal definitions because no third party is being asked to take or avoid some action as a prerequisite for the robots vacating the peoples' person. However combatants could be kept out of combat with such devices in general and as use of weapons by the people so collared would "prove" to the Autonomous Robot that they were combatants. Such straw man scenarios make it clear that there are areas of law regarding autonomous systems that IHL has not even conceived of.

There are also significant moral hazards in pursuing this area of research and development. While we cannot sit on our hands because other countries will develop such systems if we don't, there are significant risks of an uncontrolled arms race and of escalation of wars fought with such systems. Therefore, Australia would be wise to engage in international discussions early and move towards an agreed framework for developing and fielding such systems. In the short term, and perhaps permanently, this may result in a prohibition on fielding lethal systems. It is fully permissible to deploy non-lethal Autonomous systems, and use Autonomous Systems to attack other weapons or automated systems provided that there is no risk to human life. So reconnaisance, decoys and automatedly shooting down missiles are perfectly legal uses of autonomy. It's also important to recognise that the challenges of the legality of Autonomous Systems are occurring at the very time that the definitions and standards of the Geneva Conventions have blurred. The language of the Geneva Conventions did not envision multi-national non-state actors such as ISIL that disguise themselves within the civilian population, or who take hostages as tools of war. As Modirzadeh[Mod14] warns the universal commitment of states to the Geneva Conventions is in danger of breaking down if we do not actively work to clarify definitions such as "combatant". It may be that fighters who do not adhere to the Geneva Conventions forefit their rights to not be attacked, perhaps in error, by an Autonomous Robot given their propensity to be indiscriminant in their own killing. However, it is essential that such standards are discussed and agreed among states so that both updates to the Geneva Conventions and the use of Autonomous Systems strengthen the rule of law and act as a damper on escalation rather than generating increasing uncertainty and a model of war in which the first to transgress the rules of war wins. Negotations on both the Geneva Conventions and laws regarding the use of Autonomous Systems should be conducted considering the implications for the other.

6.4 Social Licence

Defence also needs to beware of a growing unease in the wider community exemplified by the "Campaign to Stop Killer Robots"[CSK13]. This is an international coalition of NGOs including Human Rights Watch, Article 36, Association for Aid and Relief Japan,

International Committee for Robot Arms Control, Mines Action Canada, Nobel Womens Initiative, IKV Pax Christi, Pugwash Conferences on Science & World Affairs, Womens International League for Peace and Freedom. This body wants to ban killer robots and argues along the lines documented above.

Moreover, in the USA, the government is losing its social licence to operate drones. There is a movement in Deer Trail, Colorado to declare laws claiming sovereignty over airspace above the town and providing financial incentives for citizens to shoot down [usually micro] drones[Fra13].

Given that most of the media about Autonomous Systems and remotely operated systems which are mistaken to be Autonomous has been either negative or sensational, it could be useful to engage the public in a discussion about the potential and real benefits of Autonomous Systems. Positive examples could include search and rescue and other humanitarian operations, or the control of a missile decoy system such as Nulka. It is in everyone's interest that the public have a realistic understanding of Autonomous Systems.

6.5 National Approaches to Regulation

In the absence of international agreement, individual nations have made internal decisions about how use of Autonomous systems could satisfy the LOAC.

6.5.1 Australia

Australia is a signatory to the Geneva Conventions I-IV and its additional protocols I-III, and they have been incorporated into domestic law.

Under the Geneva Conventions, for military action to be legal, one must consider the principles of Military Necessity; Humanity; Distinction; and Proportionality. This assessment is made even more subtle and complex by the requirement that they be considered together[Fin08]. Finn points to the importance of following the spirit as well as the letter of the law and points out that the government further constrains behaviour with rules of engagement that limit behaviour with both generally allowed actions and those actions allowed only on consultation of a higher authority. Furthermore as several nations such as the USA have not ratified Additional Protocol I of the Geneva Conventions, we must be aware of how a weapons vendor nation interprets the LOAC and how that may be embedded into their automation so that we do not violate our treaty obligations.

Finn argues that a natural division of labour currently exists between humans and automated systems with the human setting goals and making high-level target designation as well as target verification and engagement approval whereas the automation is particularly good at identifying and tracking targets as well as operating the weapons[Fin08]. They share situation awareness[16]. He further argues that although automation will improve over time this division of labour still favours having human eyes on the target and notes that many have argued that under the principle of proportionality "Rational Human Judgement" is required. Finn also questions the capacity to hold the operator accountable

[16] In general shared situation awareness does not mean that each party shares a common awareness, but that each party may be aware of different and usually overlapping aspects of the situation.

if an autonomous system is deployed to an environment which is likely to cause it to make errors further reinforcing the need for a human in-the-loop.

According to Group Captain Ian Henderson[17], Director of the Australian Military Law Centre, Australia does not currently have a formal position on Autonomous Systems but one is expected to be issued for UAVs at least, by around mid 2014. As of October 2014 he confirmed that we do not have a formal position on Autonomous Systems and he is unaware of one for UAVs as yet.

In the meantime he has co-authored a paper covering some issues associated with weapons reviews[18] in a personal rather than official position capacity. Backstrom and Henderson's[BH12] strongest conclusion is that:

> it is important that, among others, computer scientists, engineers and lawyers engage with one another whenever a state conducts a review of weapons pursuant to Article 36 [...]
>
> The reviews cannot be compartmentalised, with each discipline looking at their own technical area. Rather, those conducting legal reviews will require 'a technical understanding of the reliability and accuracy of the weapon', and how it will be operationally employed.
>
> [...]
>
> As the details of a weapon's capability are often highly classified and compartmentalised, lawyers, engineers and operators may need to work cooperatively and imaginatively to overcome security classification and compartmental access restrictions.

Issues raised by Backstrom and Henderson include:

- As weapons become more complex, so do the legal issues. For instance, swords may work or not but who is targeted by them is clearly in the hands of the operator and they cannot accidentally fire. More complicated weapons like the cross bow introduce issues with the accuracy of the sighting and launch mechanisms and so could result in injuries to the non-targeted.
- Unmanned combat systems are simply remotely operated weapons platforms and the legal issues regarding them derive from how they are used rather than inherent issues with the technology. Automated and autonomous systems are in a different category.
- "The principal legal issue with automated weapons is their ability to discriminate between lawful targets and civilians and civilian objects. The second main concern is how to deal with expected incidental injury to civilians and damage to civilian objects."
- Even relatively crude automatic weapons such as anti-truck mines are designed to discriminate on the basis of weight, for instance. However legal issues arise when

[17] via a phone call with the author

[18] The process of assessing under what circumstances the use of a particular weapon would be legal under the conventions.

the sensor discrimination is insufficient e.g. a running man can weigh as much on landing as the trip weight for a truck.

- Fused diverse sensor data can discriminate far better providing an order of magnitude better identification and two orders of magnitude better geo-location than single sensor systems.

- Although anti-vehicle mines can now distinguish between friend or foe, Backstrom and Henderson are unaware of any that can distinguish civilians or civilian equipment from military targets.

Backstrom and Henderson[BH12] list two current and one potential future way to achieve discrimination in weapons:

1. Control how they are used: e.g. locate them in places where civilian access is unlikely.

2. Retain human oversight.

3. Increase their decision making capability [sufficiently to meet legal requirements], thus creating an autonomous weapon.

They point out that such oversight is only legally and operationally useful if the human genuinely reviews the decisions being made and don't just trust the system's assertions. While postulating a potential Wide Area Search Autonomous Attack Miniature Munition, they state that while the physical issues could be engineered within 25 years the autonomous component still poses significant engineering problems and that making sure it complies with LOAC complicates things further.

Even so, there are other areas where autonomy may be used to enhance human judgement, for instance the ICRC's emblems of protection[19] are eminently recognisable and could be autonomously recognised by a "moral" weapon and detonation avoided if the target was found to be obscured or occupied by the ICRC on arrival of the weapon. So long and autonomous systems are given a right of veto and not commission, they will comply with International Humanitarian Law. Such veto systems also provide a path for Autonomous Recognitions Systems and their engineering methods to be developped to a high degree of reliability without risking contravening the Geneva Conventions.

Backstrom and Henderson[BH12] go on to describe reasons why an autonomous choice may not occur as intended. These include: incorrect specification, inconsistent design, manufacturing or integration and these can result in a lack of combat effect, risk to civilians and civilian property or legal liability or any combination of the above.

They state that as weapons systems become more complex, an understanding of reliability analysis will be essential in the legal review process. This should be performed to ensure a system, or its components, meets its specification at a given confidence level. Certification is a matter of choosing a suitable confidence level and testing to it. The level should be chosen according to the lethality and the failure modes of the weapon. Testing and evaluation needs to occur during 'demonstration', 'manufacture' and 'in-service'.

[19] The Red Cross, Red Cresent and Red Crystal

"For tests to be meaningful, critical issues of performance must be translated into testable elements that can be objectively measured." Each country must ensure that the total test suite is suitable for legal clearance of the system's intended usage profile by that country, and will most likely need to supplement vendor supplied testing to cover particular cases. The legal team should be involved early in developing tests that will prove legal compliance. Testing evidence is cumulative and field usage should be restricted until more experience is developed with the weapon system in the country's own deployment environment.

Accumulating evidence for certification can be greatly hampered for composite systems if responsibility for certifying subsystems is nebulously assigned[Off08] or required subsystem reliability is poorly determined. Backstrom and Henderson[BH12] suggest that system engineers should assign both testing and legal responsibilities (i.e. prima face liability) among manufacturers and military stakeholders.

Backstrom and Henderson's discussion[BH12] shows that the probabilistic nature of reliability fits nicely with the probability of successful detonation, correct targeting and incorrectly engaging prohibited targets. They also endorse another emerging issue of with Autonomous Systems: developing trust in autonomy will require verification and validation of systems with near infinite states.

The discussions of reliability above presumes what is now a standard definition for software reliability:

> Software Reliability is the probability that a system or a capability of a system will continue to function without failure for a specified period in a specified environment. "Failure" means the program in its functioning has not met user requirement in some way.[MIO87]

But with Autonomous Systems we are moving away from domains where we can give a precise definition of correct behaviour *a priori* but rather expect the system to exercise good judgement based on our *post hoc* judgement of what that is. For instance, one critical aspect of warfighting is the requirement of making decisions under uncertainty. Traditionally, we have relied on humans to deal with uncertainly but there is some evidence that an Autonomous System may do a better job. Fuzzy logic and conditional probabilites allow for precise reasoning about uncertainty, whereas humans are confused or at least slowed down by readouts that list the probabilities of certain realities. On the other hand, experienced humans are exceptionally good at dealing with uncertainty but do so without consious awareness of the processes they are undertaking.

6.5.2 USA

The United States of America has signed and ratified the Geneva Conventions I-IV and its additional protocol III, as of 2013 it had signed but not ratified additional protocols I & II.

The US DoD requires[DoD12] that Autonomous and Semi-Autonomous Systems be subject to strong test and evaluation and verification and validation and information assurance and to be designed with clear mechanisms that activate and deactivate the systems

including anti-tamper mechanisms so as to "minimize the probability and consequences of failures that could lead to unintended engagements or to loss of control of the system". Both hardware and software safeties are required, as are clear human-computer interfaces and appropriate training including training in meeting the requirements of operating these weapons, including the LOAC, has been given to all relevant commanders.

6.5.3 UK

The United Kingdom has signed and ratified the Geneva Conventions I-IV and its additional protocols I-III.

The UK MOD doctrine[MoD11] is more amorphous than the US regarding technological test and evaluation and verification and validation simply noting that LOAC must be followed and that a systems engineering approach would be the best way to ensure new systems meet the requirements of LOAC technically. It notes that while in some cases such as remotely piloted systems this is simply a matter of interpreting current procedures for manned aircraft, with increasing autonomy things become much more complicated. It refers people to a study by Gillespie and West[GW10] for a deeper analysis of this issue.

Gillespie and West[GW10] look in detail at the constraints that LOAC would place on engineering design of Autonomous Systems describing their approach as using the existing legal framework to establish the constraints on decision-making and where there are limits to autonomy in Unmanned Air Systems (UAS). They note that that:

> It would be a small technical step to make a UAS which would fire a weapon at the target without reference to its commander. This technical step would represent a large legal change except in very narrow, strictly-defined circumstances.

They state that "The UK Ministry of Defence (MOD) has no intention to develop systems with no human intervention in the C2 chain, but there is the desire to raise the autonomy level of its UAS[20]." They note the following issues:

UK policy is that legal responsibility will always remain with the last person to issue commands to the military system.

- An autonomous system may be created on the fly from various C2 components for a limited period of time. While a constant C2 chain [of command] for the entire mission may not be valid, the requirement for clear legal authority in command remains. How to allocate this authority and control to an Autonomous System is as yet undefined.
- Following precedent, it is reasonable to expect the responsibility of designers to have been met legally once the system is certified.
- Legal operation by the commanders relies on them being able to predict the performance of the Autonomous UAS.

[20]See the glossary for UK acronyms

Gillespie and West[GW10] derive the following System Engineering Requirements to meet the four tenants of LOAC as follows(verbatim):

Military Necessity

- There must be a clear and unambiguous command chain which controls and limits the actions of AUAS at all times.
- The command chain will have discrete and distinct authorised entities which can interpret and act on commands that meet the "proportionality" requirements below.
- The type of information required by each authorised entity to make decisions shall be stated clearly and unambiguously.
- When an authorised entity does not have all necessary and sufficient information for a decision, there must be acceptable alternative decision-making processes. Timings within the command chain must allow sufficient time to follow this process.
- Human commanders must have clear intervention points and criteria for overriding decisions made by the UAS.
- The status of the communication links between nodes must be known and there must be contingency processes for all types of link failure.
- System behaviour following an interruption in the control link must be predictable.

Humanity

- There must be a method of assessing the effect of the applied force in sufficient time to prevent the use of unnecessary force.
- There must be a means to assess whether targeted hostile forces remain a threat as defined at the start of the mission.
- The last opportunity to stop or divert kinetic and non-kinetic effects must be known to the appropriate authorised entities.
- There must be a method to confirm lethal decisions shortly before weapon release using updates to knowledge of the target area between the start of the mission and weapon release.
- There must be one or more acceptable alternative aimpoints for weapons that can be guided to their target by the AUAS after release. This is to be used if the original target is found to be unacceptable. The last time for the change of aimpoint to be effective must be known to the AUAS.

Distinction

- The basis of identification of hostile forces and materiel must be clear.
- Ambiguity in applying distinction criteria to the scene should be identified.

- Levels of uncertainty in identification must be presented to the authorised entity and there must be a way of incorporating this in the decision process.
- There must be an understanding of the normal state of the target area without the presence of hostile forces. It will include identification of objects that must not be damaged or destroyed. This will need amplification for insurgencies where the hostile forces are deliberately merging into the non-military environment.
- There must be a way for an authorised entity to assess potential collateral damage adequately at the target area before it makes its decision.
- There must be a clear link between the level of authority vested at each point in the command chain and the information available to it.
- The timing and integrity of information about the target area must be known by each authorised entity basing a decision on it.
- There must be a definition of which decisions need an audit trail for the information available at the time it was made and the means to create it.

Proportionality

- Proportionality criteria shall be set by human mission planners in a way that can be interpreted unambiguously by each authorised entity in a manner that allows it to make its decisions.
- All authorised entities must know which weapons are available to them and the limits of their authority to use them.
- All authorised entities must know the effects of the weapons available to them.
- Whenever a non-human authorised entity cannot interpret the available information and meet the proportionality criteria at, or above, a pre-determined confidence level, it must refer the decision to a higher authorised entity in the command chain.
- When an authorised entity refers a decision to another one, it must transmit the basis of its referral decision to that entity. There will need to be a recognised format for transmission of this information.

The authors then go on to ask which authorised entities could be automated and still meet the above criteria. They note that the test should be reasonable decision making, not making the same decision as a human and go on to use a three part model of decision making comprising of awareness, understanding and deliberation to reach the following conclusions:

- The complexity of many scenarios is such that there will always need to be a human at some point in the control chain. However, criteria have been derived which make it possible to ensure that any decision which needs human intervention will be referred to them at the correct time.

- A further understanding of machine understanding of ROE is necessary to decide which types of decisions can never be automated.
- Different types of sensors (at least in calibration) and data fusion will be necessary to support autonomy.
- Significant processing power will be needed on the platform, along with the ability to destroy data if the platform falls into enemy hands.
- Networks to support AUAVs will require bandwidth along with data corruption recognition, link break detection and contingency plans.

6.6 Overall legal situation

Many aspects of International Humanitarian Law as it applies to Autonomous Systems are unclear. This makes it imperitive that Australia produces its own formal position about where in the the command chain, or elsewere, that responisibility for breaches of the Geneva Conventions or other IHL is deemed to reside.

It is also important the Australia considers new scenarios, such as the potential delayed kill/release collar, that are now theoretically possible with the advent of Autonomous Systems; not only for understanding the current law, but for deciding where the law should be clarified or extended.

Our allies differ from us in the Geneva Conventions they have ratified, and among themselves in the where they believe the responsibility lies if a target is wrongfully destroyed or killed, and indeed what a legal targeting might be. While the US is stronly based on validation and verification, the UK gives a far more nuanced based on the chain of command and the last natural person to give command the Autonomous System. Both these approaches have their own merits, but do not necessarily reflect Australian approaches. Much of targeting is done via networks of staff and decisions are likely to become more so with time as chat and other froms of non-hierarchical communication increase their role in warefighting. However, there is still accountability and responsibility.

Australia should clarify its own position here, not least because our reservations in our acceding to the Geneva Conventions specifically states that actions shall be evaluated against our overall prosecution of the war not just the particular incident. Therefore, cost and efficiency that reserve firepower for later in the war are valid considerations for Australian legal determinations.

In order to prevent a breakdown of international agreement over acceptable modes of warfare, and forstall inducing an Autonomous Systems arms race that would most likely leave Australia in a weaker position, Australia needs to engage the international community as early as possible so that international agreements that are acceptable to Australia can be reached.

Considerations of strict legality are also insufficient to pursue a workable Autonomous Systems initiative. The outcome of many wars now hinges on winning the hearts and minds and our international laws of war are only one possible codification of an innate sence of what is morally acceptable in the human psyche. Conversely any action that engages

the innate sense of the morally unacceptable will incur the most determined resistance. Moreover a purely legal approach is unlikely to produce the most accurate systems in terms of distinction between combatants and non-combatants. It remains the case that there is no way punish a machine, but punishment is really just one mechanism to prevent repeated mistakes. Approaches that focus on causes of failures rather than blame and which take faulty systems out of action until they are corrected, such as the civil aviation approach to passenger safey may be far more effective in both ensuring IHL is not breached and producing useful automation for the military. The UK approach in particular spells out some minimum sytem requirements for builing accountable systems that remain safe to us should they fall into enemy hands.

7 Human Interaction with Automations and Autonomous Systems

The literature on Automations and Autonomous Systems frequenty emphasises the importance of good human computer interaction as a prerequisite to obtaining the their expected benefits.

7.1 Interaction Issues from Robotics

Although C2 does not generally involve robotic systems, some of the human computer interaction issues that have arisen in robotics can inform how we might build autonomous C2 systems.

Our definition of Autonomous Systems implies that some machine learning and / or artificial intelligence are often incorporated in such systems. Great progress has been made in recent years in providing limited intelligence in specific domain areas, such as game playing and machine translation, but it is arguable whether some of these represent true understanding by the machine. Both the machine and the humans operating it may be unaware of underlying reasons behind its decision depending on the machines type of implementation and the data it was trained on.

There is a significant gap between what is required and what has been achieved in Autonomous Systems, particularly in the public mind. Murphy and Woods[MW09] describe how popular Isaac Asimov's three laws of robotics have become in everyday society and point out that they are currently deeply unachievable, but are also insufficient, unsafe and counter to our cultural infrastructure. Asimov's Laws[Asi94] are somewhat of a strawman used by Murphy and Woods and are included here for reference:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Murphy and Woods propose replacing Asimov's three laws with ones that work constructively with our cultural requirements, are implementable and build public confidence in robotic research so that research into autonomous systems can continue. These laws are listed below with some of their justification for each:

Alternative First Law *A human may not deploy a robot without the human-robot work system meeting the highest legal and professional standards of safety and ethics.*

Robots must meet safety requirements for the field they operate in. E.g. in medicine, failure mode and effect analysis are required by the FDA medical device standards.

In the US, special regulations for experimental vehicles have encouraged a poor safety culture, but the highest standards are needed because an accident or failure can "effectively end an entire branch of robotics research, even if the operators aren't legally culpable."

Robots should carry "black boxes" to show what they were doing when a disturbance occurred.

Alternative Second Law *A robot must respond to humans as appropriate for their roles.*

"The capability to respond appropriately — *responsiveness* — may be more important to human-robot interaction than the capability of autonomy."

The second law says that robots must operate according to the relationship between itself and particular humans, and between different humans - e.g. obeying the mother rather than the child. Note that responding appropriately may mean ignoring the requests of the criminal or the enemy.

Alternative Third Law *A robot must be endowed with sufficient situated autonomy to protect its own existence as long as such protection provides smooth transfer of control to other agents consistent with the first and second laws.*

Bumpy transfers of control can cause human errors which contribute to system failures. The proposed law implicitly describes different operating modes with alternative permissions and interaction patterns for the same role. This puts the onus on designers to identify situations where a transfer of control is appropriate. It allows staff to be appropriately trained to hand over or receive sufficient information. Designers must decide when the autonomous system and when he human operator is more capable of handling the situation. This requires higher levels of situated autonomy for the automated system, and paying heed to decades of human factors research on human out-of-the-loop control problems and transfer of control.

The authors[MW09] point to the irony of how increased system autonomy and authority leads to the need to participate in more coordinated activity.

In building a C2 autonomous system, careful thought will have to be given to how ethics and safety are upheld throughout the development, testing and operation of such a system. It would also be worthwhile to gradually engage with the public so that they have an accurate overview of autonomous systems rather than unrealistic expections, this would help insure against strong reactions if anything "unexpected" occurs. Murphy and Woods point about systems responding according to role is highly relevant for security. However, it should be pointed out that Australian law in particular and international law do not recognise a right to defend property such as autonomous system components, or other platforms, at the cost of human life.

7.2 Human Factors, Operations and Trust

Throughout this literature survey, many papers have emphasised the need to create trust in the automation by the users/operators, between systems, and between autonomous

agents. Some authors see establishing trust as a utility limiting issue due to operators refusing to adopt untrusted systems. Others, notably Blackhurst et.al. [BGS11], warn that it is actually a balance of trust issue: too little and the system is wasted, too much and the operators abandon attention and responsibility. Moreover, trust will have to be established between autonomous systems to validate data being shared between them, and to allow data prioritisation when dealing with unreliable or conflicting data.

Paradis et. al. [PBOC05] highlight trust as a major issue with Decision Support Systems (DSS) including the requirement that the DSS provide an explanation for their line of reasoning as a prerequisite for the human operator to be willing to accept it, and establishing trust and distrust in decisions made by other parts of the system. A record of such reasoning would also be important in any after-action review.

Hoffman et. al. [HJBU13] contrast interpersonal trust and trust in autonomy. Human trust in autonomy is similar to interpersonal trust but far more demanding while being less forgiving. According to Hoffman et. al. while humans can create some instant trust by admitting their own limitations or sharing items that "prove" a shared world view there is no corresponding transaction a machine can perform. In fact they argue, in a sense humans have complete negative trust in automation as they are certain that one day it will fail. They propose that rather than trying to maintain trust levels, the system should offer a human computer interface that allows the human to more actively decide what aspects of the system should be trusted and to what level. This requires decision traceability (showing by what logic a decision or belief was arrived at) and provenance (being able to examine the data and assumptions that lead to any particular conclusion for its original source, accuracy, reliability etc.) The authors suggest that the distinction between testing and the operational phase of software system deployment is artificial and that any time, operators should be able to test the automation for its effectiveness in any goal. This allows operators to build trust in the automation. The notion that testing and operation is a continuum is expressed by other authors in respect to using the operational phase to prove the highest levels for reliability[BH12, Mus04].

Given the increase in unmanned systems across air, ground and maritime domains, experiments were conducted by Salcedo et.al. [SOL+11] to analyse the level of trust within teams which included either remotely operated shooters or automated shooters. Teamwork is critical for success and human robot interaction is no exception. Adding non-human entities to a team raises concerns about interaction quality. Potential issues are trust in the human computer interaction, perceived system unreliability but also over reliance and complacency. The same weapon was inserted into teams of three humans and the humans told whether the weapon was being operated remotely or autonomously. The experimental environment simulated desert and urban conditions and each human and the test weapon was given their own shooting lane with no entity in the team targetable by other team members. Each human was given a questionnaire assessing the quality of the teamwork, their cognitive load, and their assessment of the fourth gun. Results were disappointing as they found little variation in trust at the 0.05 confidence level and the authors were keen to identify automated systems attributes that caused unease. The authors report the only significant difference between the two cases were higher levels of cognitive load for humans in the fourth shooting lane, next to the automated weapon in lane three. These humans had no other human reference point, which may be the most important point of their work given that the authors admit that their testing was not extensive.

Blackhurst et. al. point to the dangers of increasing system autonomy without designing for interdependence between operators and the system[BGS11]. At low levels of autonomy, humans can become functionaries of the system - e.g. in remotely piloted vehicules; whereas at high levels of autonomy humans can be left asking "What is it doing?", "Why is it doing that?" and "What is it going to do next?". Importantly they claim that without designing for interdependence, increasing autonomy will often fail to improve mission performance.

Even if only trying to reduce staffing requirements by changing the nature of the human task, a good understanding of human task design will be necessary.

7.3 Certification and Validation for Trust

For the commander to have confidence in an autonomous system it should be certified to a certain level of reliability. However this is considerably problematic for autonomous systems. While Musa[Mus04] and Poore et. al. [PMM93] provide methods for calculating the needed reliability of subsystems to achieve a certain overall reliability and to determine the overall system reliability given actual component reliabilities, these methodologies assume the system is operating in a specific and known environment, and that the expected output can be clearly defined to test against, as do systems validation techniques in general. Autonomous systems may behave in unexpected ways in unfamiliar environments and their output is not necessarily deterministic — a learning system should occasionally generate unexpected outputs. One approach may be to apply rigorous development processes such as those outline by Croxford and Chapman[CC05] to the building of learning systems, so at least the process by which the autonomous system came up with new results was well understood and verifiable. This would require that they be built in languages amenable to static analysis.

As autonomy becomes more pervasive and more capable, it is increasingly likely that it will be deployed into more complex environments where the automation is more likely to select the wrong target. However, these are precisely the environments where the system may be most required to enhance survivability[Fin08]. While Backstrom and Henderson[BH12] advocate the use of multi-disciplinary teams to develop Testing, Evaluation, Validation & Verification for autonomous systems; research needs to be done to find certification methodologies that are appropriate for autonomous systems, especially ones that progressively learn. Whereas behaviour of a traditional system is a function of its specification and data input, the behaviour of an Autonomous System is a function of its specification, current operating environment (which defines what specific inputs may occur), and its entire training history (the inputs which generated its learning). Therefore traditional means of testing and validation are insufficient to certify an Autonomous System. One issue for such research to consider is how to extend accumulation of reliability evidence into the operational domain — this is necessary because often testing will not have occurred for the new environment.

8 Discussion

8.1 Human Factors

A repeated theme across many papers read for this literature review has been the importance of human factors in achieving a positive outcome from implementing autonomous systems.

Because the effectiveness of autonomy is critically reliant on the quality of its interactions with humans, both to make joint decisions and maintain shared situation awareness, human factors research will be a key enabler of an effective autonomously assisted C2 environment.

Some assessments by authors working in the field of autonomous systems include: "We can simply no longer afford to develop technological solutions that don't consider the human role in using them[BGS11]." Nor can good human-computer interaction be bolted on later. "The design parameters for an interdependent human-machine system look very different than a machine designed to maximise autonomy[BGS11]." "It is technically feasible to facilitate higher levels of automation, however to succeed, researchers need to improve existing operator control techniques and HMIs[Twe12]."

Just some of the issues that would benefit greatly from human factors studies include:

- How to build effective user interfaces for various types of automation. And more generally, how can we build user interfaces that are consistent and therefore familiar over a range of C2 automations? Is it possible to produce modular interfaces so that an overall user interface can be built dynamically by the system when the automation orchestrates several autonomous components together to build a larger system for a specific purpose?

- How can uncertainty of data or advice be conveyed to an operator in a way that enables them to assimilate it and make decisions on the set of competing advice without overloading the operator or causing them to overly check the correctness of assertions being made? This is complicated by the fact that we still want the operator to notice and correct any cognitive or logic errors made by the automation. It is further complicated that LOAC requires the commander to assure themselves of the correctness of any decision[BH12].

- How best to display and engage the operator to maintain ongoing situation awareness.

- How to get an autonomous system to recognise different memberships, rank and rolls of people interacting with it, and especially to recognise social cues that convey this information so that the automation can tailor its behaviour, service access control and responses in a way that is most useful to the individual without compromising system security.

- How to reduce the cognitive load on the operators. Rather than aiming for more volume of information, it is better to aim for more quality and confidence in the

information presented to the operators and commanders, and for more clarity in the way that information is presented including prioritisation.

- How can the operator interface be tailored to the skill of the particular operator, and in effect assist in training up less experienced operators? "Interface designs must also factor in the skills and experience of operators, to enable each the ability to customise the desired LOA to enable them to adjust while learning or until they develop an acceptable level of trust[Twe12]."

- Investigation of cognitive engineering literature on transfer of control and general human out-of-the-loop problems as they apply to autonomous systems. Many transfer of control issues arise in dealing with automated systems. They include:

  - How to transfer control back to humans, which is important both for C2 and physical autonomous systems such as drones.
  - The special case of the former: how the system should direct attention of the operator to potential unexpected emerging threats without overloading them with false alarms[21]? How can the operator interface increase the likelihood of the operator noticing a threat that the autonomous system hasn't? This is particularly difficult given that humans are notoriously poor at monitoring.

  Several of these issues arise regarding teams:

  - How should teams of operators be tasked for C2?
  - How should teams of operators hand over to the next shift?
  - How should teams of humans interact with teams of agents? There is an open research space to explore mixed human / autonomous teams particularly in the areas of coordination and cooperation and especially creating situation awareness of the cooperation itself[Twe13].

- The issue of who takes the initiative is worthy of ongoing human factors study. While some results, and human intuition, suggest that mixed initiative should be preferable, there will be cases where events occur too fast for humans to be the first to react. Moreover the results regarding mixed initiative show effectiveness depends on the ability to automated agents to be able do effective work when unsupervised. Each application domain will require some investigation into the appropriate type of initiative and level of automation for optimal results.

8.1.1 What should be automated?

The choice of what to automate or make autonomous is obviously important, but Blackhurst et.al. [BGS11] offer the insight that "Ultimately, the manpower savings from interdependent robots and machines comes not from replacing people but by changing how we accomplish our missions." They counsel strongly to design in interdependence between humans and automated systems.

[21] There is also the corresponding problem of how the automation might detect such a situation.

Robin Murphy's law of robotics[MM97], as cited in [BGS11], states: "Any deployment of robotic systems will fall short of the target level of autonomy, creating or exacerbating a shortfall in mechanisms for coordination with human stakeholders."

Moreover, simply reducing required staffing levels is not the goal of the military nor should it be the goal of any attempt to automate. The actual goal of military automation should be to create decisive advantage and reduce the human and material costs of procecuting a war. Therefore, developing a good Concept of Operations detailing human and computer roles, expectations and interactions for an automated system before attempting to build it will be essential for developing usable systems that are adopted by Defence.

8.2 Creating and Maintaining Trust

Creating and maintaining trust is a critical issue in autonomy and has been a recurring theme in many papers surveyed. Beyond the human factors studies of trust, it is evident that autonomous systems need to provide their operators and auditors with a provenance trail for both data and inferences made. Mandatory tagging of data with both its provenance and currency(age and shelf life) should be implemented across all Australian autonomous systems. Protocols between interacting entities must be designed to ensure data integrity, not just in terms of transcription errors, but also in ontology and data interpretation.

In an autonomously supported UC2 environment with decisions being made by a variety of entities over time, it is also essential that no transfer of authority occur without an explicit handover of authority including duration, limits of authority and rules of engagement, and that for such a handover to take place the receiving entity must explicitly acknowledge receipt of the authority. All transfers of authority must be logged by the system.

8.3 Measures of Autonomy and its Performance

Hardin and Goodrich cover measures of autonomy and of Performance[HG09]. A primary measure of autonomy is how much *fan-out* is possible, that is how many entities or activities can a human operator manage at a high level and allow Autonomous Systems to perform these activities independently. Fan-out shows how much raw leverage the system can provide. Measures that feed into the fan-out result include allowable *neglect-time* - how long the operator can neglect to manage an autonomous agent that has completed its primary human assigned task before its ability to do something useful in the meantime drastically decreases. Research has also begun into measures of cohesion and performance for agent teams. Fan-out increases with greater allowable neglect-times, and with more usable user interfaces. Fan-out is also an issue in directing humans within management theory and many organisations' procedures placing limits on the number of subordinates a manager may have. These are measures of efficiency, but we also need measures of effectiveness. These will be specific to the problem domain of each autonomous system but should capture the differences in value between different outcomes, including the time required to produce each outcome.

In order to achieve the best results from autonomous systems, we must choose appropriate measures of the quality of our automation. Research into developing appropriate measures of automation from low level ones such as fan-out and neglect time to the level of automation actually achieved by the system in terms of overall workload placed on operators compared to work done is necessary to both set goals for automation and to evaluate how successful implementations have been. One problem with the introduction of semi-autonomous unmanned vehicles is that they have often created greater staffing requirements and operator stress than their non-automated counterparts. These are measures of the automation itself and would also include the level of automation achieved (place on our spectrum) and the level and type of support staff required to keep the autonomous systems operational.

Another area of measurement worthy of study is development of appropriate domain specific measures of effectiveness of the end results of automation and decision quality. In C2, this would require models of where C2 value is created and the C2 value chain. This will allow prioritisation of where automation and autonomy are likely to create the most short term benefit. Such effectiveness measures should include how much overall cooperation is achieved between humans and among humans and automated systems.

Another possible method to measure and understand our autonomous systems is to create a simulation environment in which C2 autonomous systems can be experimented with for ongoing development and refinement. This is a research and development task in itself.

8.4 Legal Issues, Certification and Safety

As can be seen from section 6, the legal issues regarding autonomous weapons systems are far from resolved. Certainly, it will be a some time, if ever, before autonomous systems can legally operate with the ability to decide to kill for themselves except in the most curtailed circumstances akin to automatic detonating weapons such as mines. Indeed the problem of how to build a system that understands laws of engagement is difficult, even assuming the problem of recognising the enemy and enemy behaviours was already solved. Gillespie and West ask, “How do we encode rules of engagement with sufficient precision and leeway?[GW10].”

Ethical behaviour monitoring should be incorporated into the architecture of autonomous systems from the outset. While researchers such as Arkin et. al.[AUW12] argue that ethical governance can be bolted onto existing systems, even providing the hooks for doing so from the outset would save money through good software engineering practice. The idea of a purely bolt-on governance module is also arguable: not only would it be necessarily less efficient - using resources to reassess decisions rather than improving the original decision; it would also by definition be unable to take advantage of the environment and application specific determinants of what constitutes “ethical”. Having a standard LOAC evaluation module, to be to be included and individually integrated as the governor, would both decrease software development costs and, over time, increase the reliability of the governor, because it is being maximally exercised in real scenarios. Moreover, a fully integrated govenor could take advantage of enviroment and applicaton

specific opportunities in its decision making. Alternatively it might be decided that having a native LOAC evaluation system, combined with the bolt on governor, provides the extra safety of two separately designed systems — either of which might prevent a possibly illegal firing.

In addition to ethics, any autonomous system built by Defence must incorporate at design time security and anti-tamper mechanisms to prevent unauthorised or accidental activation, particularly of lethal autonomous capability and any other hazardous behaviour including the unauthorised release of data. Muphy and Woods advocate that research be conducted on how network and physical security can be incorporated into autonomous systems even during development[MW09] so that they cannot be hijacked during trials. Any internally developed LARs and those purchased should have software and hardware safeties preventing use of real force (as opposed to simulated/training exercise force) within Australia without production of credentials "proving" a state of war on our territory. As a significant threat of autonomous systems is that posed by having them hacked and redirected by the enemy, this serves as a second layer of protection beyond role based authorisation.

Australia should be an early and active participant in international negotiations regarding any treaty regarding the use of autonomous weapons. Without a clear rational, it is possible that such a treaty will curtail useful but non-lethal uses of autonomy ranging from reconnaissance to automated defence and attacks in the purely cyber domain. Further research would be useful to understand how such treaty obligations might impact autonomous systems technically, and also how viable such limits might be. This would allow Australia to derive a position on what should or should not be curtailed in such a treaty, and what should be curtailed for a period until technology advances or our understanding of Autonomous Systems capability improves.

Assuming that humans will be in-the-loop for now, for a commander to be able to issue a command to fire at a target they must be confident of the effect of the weapon in the context so as to be sure that there are not unintended casualties or damage to protected infrastructure. As such, any autonomous system recommending courses of action must be able to predict the actual effects of a weapon deployment to a high reliability. Currently the Air Operations Centre have a cell of people to perform this function, and the ROE determines who has the authority to authorise engagement based on collateral damage estimates. Thus doctorine will have to be developed to decide who has the authority to authorise an autonomously proposed course of action under what circumstances. This will require further research and development. Such a capability is also a prerequisite for any lethal autonomous system that could comply with the LOAC.

The commander must also have confidence in the behaviour of the autonomous system, whether it is providing decision support, or whether it is an autonomous weapons system that the commander is tasking to achieve a specific effect. Therefore being able to test and certify the autonomous system is essential for the commander to claim that their decision was taken with the knowledge that no unintended consequences were likely to arise. Providing a simultion mode that allows the commander or operator to simulate giving a directive to an Autonomous System so that they can observe a list the "actions" that the system would have taken in operational mode would assist greatly in developing human trust for such systems and aid in training. It would also allow the human to

further check on-the-ground the suitability of the Autonomous System to operate in new and unexpected environments.

In order to assist certification and forensic examination of events for LOAC, the equivalent of a "black box" data logger for autonomous systems needs to be developed. Preferably one of a common/standardised design across all the forces.

To facilitate trials and maintain Defence's social licence to conduct them, general guidelines for safe Research, Development, Testing and Operation of Autonomous systems need to be developed prior to each of these activities and their use rigorously enforced for each autonomous system. It would also be advantageous to engage Civil Aviation and Maritime Authorities early to ensure that any automated systems can under certain circumstances operate in mixed airspace/seaspace with civilian traffic.

8.5 Self Awareness

> By removing the pilot, designers have effectively removed any on-board cognitive processing and replaced it with a geographically isolated and time-shifted tele-presence[Twe12].

It became clear during this literature review that the systems with the most potential for independent action not only modelled the situation as in the environment but also actively maintained a model of their own internal state and how they were capable of interacting with the environment. Autonomic systems are a prerequisite for such reflexive self-awareness. Models of the world that include the environment, the self, and an explicit representation of their interaction allow for a more accurate decision model of how a system action will affect the next state of the whole system (autonomous system and its environment), it also allows far more complex adaptive behaviours. Most importantly it can greatly expand the capacity for the autonomous system to reconfigure itself to better fulfil its mission, and can even allow for the autonomous system to request specific upgrades from its designers and programmers as demonstrated by Steiner and Athanas[SA09]. This is more than health maintenance and homeostasis, in an aritificially intelligent system such self-awareness and modeling of the self in the environment should provide a level of metacognition that provides the possibiltiy for system to further improve on its own executive faculties.

Defence autonomous systems should be built upon such comprehensive world models wherever possible and the general preference to do so be included in our aquisition processes.

8.6 Counter Measures

The author did not come across papers covering counter and counter counter measures for Autonomous Systems beyond the references in descriptions of Ultralog[CG08] referring to hardening the Ultralog system against cyber attack. This is definitely an area of research that needs to be covered if autonomomous systems are to be deployed. As the centre of gravity of C2 systems is usually far from the battlefront, C2 systems are likely to

be as vulnerable, or more vulnerable, to cyber attack than any other form of attack. Autonomous systems are also likely to be one of the best ways to defend against cyber attack given their ability to handle large amounts of data, to detect new patterns of interaction and any deviations from previous patterns in real-time.

In addition, it's absolutely clear that the enemy's Autonomous Military Systems are themselves legitimate targets and this is one area in which our Autonomous Systems could be safely set loose on attacking the enemy. As our enemies are also likely to attack our systems, research into hardening our Autonomous Systems and providing area denial against enemy robots is a valuable area of research.

8.7 Standards and Integration

A considerable enabler of autonomy in Defence will be the creation and / or adoption of standards across the enterprise. At the most mundane level, adoption of IP6 rather than IP4 is a prerequisite for The Internet of Things, which could significantly benefit logistics tracking in C2. IP6 also facilitates addressing a dynamically changing and large number of sensors.

A significant contribution would be to develop a general framework including interfaces for C2 autonomy, factoring out the main concepts so that various C2 and sensor subsystems can interact in a consistent way while requiring limited special interfaces to enable the use of each type of sensor or decision.

Likewise, there is a need for an overarching method of prioritising war fighting action over the entire theatre during times of war. It would be appropriate to investigate how such prioritisation should be decided, such prioritisation is a research activity because the distributed nature of action means that capacity in one location is not simply transferable to capacity in another. In fact, this may be an area for which an autonomous system could be developed.

The literature also made it clear that there was no overall standard to launch, recover and control autonomous reconnaissance or weapons systems. While clearly each system will differ, it is worth ensuring that their interfaces, authorisation and operation are standardised as much as possible to allow commanders and autonomous entities the maximum freedom in adopting and using each capability as needed.

Work also needs to be done in quantifying data link capacities needed to support widespread use of autonomous systems in Defence roles.

A major standard that needs to be developed for Australia is who is responsible in terms of the LOAC for decisions made by autonomous systems. Commander's aren't programmers and cannot be expected to predict the behaviour of autonomous systems perfectly. The apportionment of responsibility between programmers, equipment manufacturers and data suppliers will need to be legally clarified.

Building and / or driving standards for autonomy across the various forces of the ADF would be extremely useful and may need to be performed by a group or stakeholder that can be neutral between them.

8.8 The need for Engineering Methodologies

The literature review revealed that the majority of work in Autonomous Systems and Artificial Inteligence lacked clear repeatable engineering methodologies that could produce predictable results. These are essential if we are to have any hope of reliably specifying, ordering, building and procuring and certifying for use Autonomous Systems that meet the needs of the ADF.

The majority of development of these systems and Artificial Intelligence is either being done in a handcrafted manner requiring deep tacit knowledge of both which solutions to apply to which problems and/or how to optimise the programming or training of these systems to produce good products quickly, or where good methodologies exist they are are isolated and proprietary. Experience with Watson's DeepQA development[FBCC+11] shows success was dependent on high quality methodologies, allowing numberous and frequent independent experiments to accelerate development. However, this was necessary but not sufficient, the success of DeepQA was also dependent on twenty researchers highly tailoring and customising the approaches used to a specific problem.

Not only is the field of Autonomous Systems potentially as big or bigger than all Computer Science including Artificial Intelligence that has come before it but also we are very much at a stage equivalent to the early days of programming with "goto"s and a lack of engineering expertise that can guarantee a satisfactory outcome. This also means it is currently impossible to predict the cost or schedule of Autonomous Sytems development in the general case.

Significant research effort is required in collecting and systematising effective development and certification processes for Autonomous Systems. Until such processes are commonly used, we can expect the procurement of Military Autonomous systems to be far more expensive and so any research this area, while likely to be expensive, will more than repay investment.

9 C2 Opportunities

C2 is a unifying concern of all automation in the military and just as humans grow up through stages of dependence, independence and eventually interdependence, all automation and autonomy in Defence will eventually need to link back into a C2 backbone to provide situation awareness to a commander, and interact with each other in ways not always envisaged at design time to assist each other in performing their tasks. While it is ambitious to imagine such a hive of interacting autonomous entities, and this vision may not be fully achieved; unless sufficient building blocks are put in place to allow for such interactions not only will integration not be achieved but many less ambitious automation may be foreclosed. Or where automation is still possible, their benefits and value may be diminished due to a lack of underlying support.

A clear example of this is standardising on minimal autonomic abilities (section 4.3) for all platforms, all C2 systems, all weapons and many components by integrating into each of them:

- a unique identity and class of object
- a communication interface
- sensors to assess their own state and health
- sufficient integral processing capacity to read and report sensor values via the network, to operate any actuators remotely, and to validate commands to avoid accidental or malicious reading or activation.

Retrofitting such autonomic capability is both prohibitively expensive and relatively ineffective compared to systems originally designed with autonomic hardware and software. Certainly more complex systems should have more complex autonomic functions. However, even minimal autonomic functions would allow profound improvements in maintenance and logistic analysis[UIT+11] including assessing many components of Defence Force readiness without human effort. Such benefits should be realisable without having to implement strong AI or highly autonomous systems while simultaneously greatly facilitating autonomy and improving its quality should it be implemented at a later stage. Such standardisation would be the first step towards leveraging The Internet of Things[UIT+11] for logistics, C2 and other Defence applications.

Moreover, it is becoming increasingly evident to researchers that embodiment (section 4.2) can greatly strengthen the depth and quality of learning that AI entities are trained. The autonomic wiring mentioned above is a prerequisite to embodiment, which also requires actuators or other tools which the AI can activate to affect its environment. Embodiment should be encouraged, where possible, as a standard - at least to the point of hardware provision and wiring and a basic software interface.

A possible research direction would be to develop Defence embodiment and autonomic systems that are general enough to cover all the forces, and to scale from platforms to components.

Multi-Agent Implementations are a natural fit for Defence to implement autonomy but problems of scheduling and programming overhead must first be overcome. In section 4.1 we summarised why agent programming is both sufficient and beneficial in general to build AI systems, primarily because it can be inclusive of other forms of AI. In the Defence environment, multi-agent systems also allow for flexibility in deployment location of various functions, replication, migration and recovery in case of failure, and MAS can be adapted to operate in an SOA[CIO10, CoA11] environment as mandated by CIOG group because the communication via messages is part of the MAS architecture. Furthermore, it is not only possible to use MAS to support UC2, MAS internally uses UC2 concepts. MAS is heterachical in nature, allowing control to flow from one entity to another as needs dictate and it encourages a division of labour in which human and artificial agents can replace each other within any team. It supports both cooperation and competition among agents, and in so doing is a natural fit for coalition operations. Resilience engineering techniques can identify new command architectures for distributed multi-echelon systems including systems with autonomous components[MW09]. Multi-Agent Systems could be a base architecture for C2 autonomy research, as they are general enough to be capable of integrating many other forms of AI systems, which is not generally the case for other AI architectures.

Research should be done to consider which languages and architectures are most useful for autonomous systems, and C2 autonomous systems in particular. The Ultralog system[CG08] has demonstrated that extremely robust, secure and gracefully degradable systems can be built using MAS and MAS's utility in C2. The ActiveEdge[Sof13] framework used in Ultralog is available commercially, It would be useful to evaluate whether adopting this framework is cost effective, and the extent to which it limits us sharing our system with our allies etc. The ActiveEdge system is based on Java and is completely asynchronous. Other possibilities for consideration would be building a MAS framework in languages such as Erlang which provide very lightweight concurrency, native failure detection, and the ability to communicate both asynchronously and using virtual synchrony[Bir93] which can greatly simplify negotiation protocols. Erlang's functional nature is also a good fit for AI applications. Such a framework should include a solid time base synchronised as well as possible to a zulu time reverence, and a causal clock to demonstrate conclusively what occurred "simultaneously" and what events potentially caused each other. This would allow temporal integration of disparate C2 data, and provide a solid foundation for after action reviews.

While this author generally advocates strong failure models[22] for systems because they greatly simplify crash recovery, crash and error detection and recovery are essential in autonomous systems if the commander is to have confidence in the behaviour of the system during temporary perturbations and failures. This is necessary not only for technical reasons but to give sufficient information to commanders making critical decisions. Predictability is only possible if failure behaviour is known, and the system needs to actively report which areas of it are currently not functioning. Given the potential for enemy action to destroy parts of a wider autonomous system, such properties in an autonomous systems framework for C2 are even more important.

10 Conclusions

The grand vision of autonomy is to transition from the system as a tool or an automation to the system as a resiliant team mate, that is the sytem becomes not just a "participant" but an "actor" in its own right able to deal with uncertainty.

Keeping the human in the loop is also strongly argued to be necessary by many people and entities involved in the Law of Armed Conflict. Work needs to be done to consider integrating ethical governors with Autonomous Systems to ensure our rules of engagement and LOAC are followed. However, provided C2 autonomous systems can avoid triggering the firing of lethal force, the requirements of LOAC will be generally met; in fact it is only when a system has both target selection autonomy and firing autonomy that the possibility exists for the system of itself to violate the laws of armed conflict. LOAC also clearly allows autonomous systems to attack other enemy systems for instance, both mine clearance and cyber attack of military systems are clearly allowable.

Managing the social licence to operate including legalities will be crucial to the implementation and deployment of autonomous systems.

[22] I.e. systems that fail safe and are designed to succeed or clearly fail rather than being left in unknown intermediate states. These greatly simplify crash and error recovery.

The domain of autonomous systems is a vast and diverse research area. However, this literature survey has been able to identify both some fruitful areas of research and at least the first steps on the path towards implementing autonomy in defence namely developing some standards for the ADF particularly in the areas of engineering processes to develop, procure, validate and certify Autonomous Systems for deployment. Likewise, implementing autonomic systems for C2 and other platforms as a standard requirement will not only lower the cost of maintaining such systems but increase their effectivenes.

These first steps not only have immediate value, but will inform and enable more complex and higher value research areas in the future. Moreover, it is important to get started because there are a significant number of core competencies that are required to produce effective deployable Autonomous Systems in C2 and each of these competency areas will need to be honed by gradual experience.

There is also a danger that we do not simple seek to automate processes that already exist but actively look to redesign our processes and work allocations between human and automated actors to leverage the specific aptitudes of each group. A primary area to look at would be to develop an Australian *Concept of Operations* for Autonomous C2, which will inform our other choices.

Acknowledgements

The author would like to thank Ian Dall, Steven Wark, Darryn Reid, Brian Hanlon and Jason Scholz for their helpful comments on this paper.

References

Adv14. Advanced chess. Wikipedia Article, 2014. https://en.wikipedia.org/wiki/Advanced_Chess.

AKIW13. S.J. Anderson, S.B. Karumanchi, K. Iagnemma, and J.M. Walker. The intelligent copilot: A constraint-based approach to shared-adaptive control of ground vehicles. *Intelligent Transportation Systems Magazine, IEEE*, 5(2):45–54, 2013.

Anz03. Chris Anzalone. Readying Air Forces For Network Centric Weapons. Power Point Presentation, Nov 2003. http://www.dtic.mil/ndia/2003targets/anz.ppt.

Asa12. Peter Asaro. On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross*, 94:687–709, June 2012. http://journals.cambridge.org/article_S1816383112000768.

Asi94. Isaac Asimov. *I, Robot*, chapter Runaround. Bantam USA, 1994. ISBN 13: 9780553294385 ISBN 10: 0553294385.

AUW12. R.C. Arkin, P. Ulam, and A.R. Wagner. Moral decision making in autonomous systems: Enforcement, moral emotions, dignity, trust, and deception. *Proceedings of the IEEE*, 100(3):571–589, 2012.

BB08. RuiPedro Barbosa and Orlando Belo. Autonomous forex trading agents. In Petra Perner, editor, *Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects*, volume 5077 of *Lecture Notes in Computer Science*, pages 389–403. Springer Berlin Heidelberg, 2008.

BB10. Rui Pedro Barbosa and Orlando Belo. Multi-agent forex trading system. In Anne Hkansson, Ronald Hartung, and NgocThanh Nguyen, editors, *Agent and Multi-agent Technology for Internet and Enterprise Systems*, volume 289 of *Studies in Computational Intelligence*, pages 91–118. Springer Berlin Heidelberg, 2010.

BBC+10. A. Broggi, L. Bombini, S. Cattani, P. Cerri, and R. I. Fedriga. Sensing requirements for a 13,000 km intercontinental autonomous drive. pages 500–505, 2010.

BBD+13. A. Broggi, M. Buzzoni, S. Debattisti, P. Grisleri, M.C. Laghi, P. Medici, and P. Versari. Extensive tests of autonomous driving technologies. *Intelligent Transportation Systems, IEEE Transactions on*, 14(3):1403–1415, 2013.

BEA12. BEA. Final Report On the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris. Technical report, Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, Ministère de l'Écologie, du Développement durable, des Transportset du Logement, Zone Sud - Bâtiment 153, 200 rue de Paris, Aéroport du Bourget, 93352 Le Bourget Cedex - France, July 2012. http://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf.

BGS11. Jack L. Blackhurst, Jennifer S. Gresham, and Morley O. Stone. The autonomy paradox. *Armed Forces Journal*, October 2011. http://armedforcesjournal.com/article/2011/10/7604038.

BH12. Alan Backstrom and Ian Henderson. New capabilities in warfare: an overview of contemporary technological developments and the associated legal and engineering issues in Article 36 weapons reviews. *International Review of the Red Cross*, 94(886):483–514, June 2012. http://www.icrc.org/eng/resources/documents/article/review-2012/irrc-886-backstrom-henderson.htm.

Bir93. Kenneth P. Birman. The process group approach to reliable distributed computing. *Communications of the ACM*, 36(12):37–53,103, December 1993.

BO09. R.S.A. Brinkworth and D.C. O'Carroll. Robust models for optic flow coding in natural scenes inspired by insect biology. *PLoS Computational Biology*, 5(11), 2009. cited By 32.

Bro90. Rodney A. Brooks. *Elephants don't play chess*, volume 6. 1990. http://rair.cogsci.rpi.edu/pai/restricted/logic/elephants.pdf.

Car13. Michael Carney. Want to take on wall street? quantopian's algorithmic trading platform now accepts outside data sets. *Pando Daily*, April 2013. http://pando.com/2013/04/02/want-to-take-on-wall-street-quantopians-algorithmic-trading-platform-now-accepts-outside-data-sets/.

CC05. Martin Croxford and Roderick Chapman. Correctness by construction: A manifesto for high-integrity software. *The Journal of Defense Soft. Engr*, pages 5–8, 2005.

CG08. Todd Carrico and Mark Greaves. Agent applications in defense logistics. In Michal Pchouek, SimonG. Thompson, and Holger Voos, editors, *Defence Industry Applications of Autonomous Agents and Multi-Agent Systems*, Whitestein Series in Software Agent Technologies and Autonomic Computing, pages 51–72. Birkhuser Basel, 2008.

CGHP09. Russell Connell, Simon Goss, Clint Heinze, and Michael Papasimeon. Agenta sapiens: The next evolution. In *Proceedings of the 18th Conference on Behavior Representation in Modeling and Simulation*, pages 151–152, Sundance, UT, USA, 31 March - 2 April 2009. http://cc.ist.psu.edu/BRIMS2013/archives/2009/papers/BRIMS2009_034.pdf.

CIO10. CIOG. Single Information Environment (SIE): Architectural Intent 2010. Technical Report DPS: DEC013-09, Commonwealth of Australia, Department of Defence, May 2010.

CoA11. Department of Defence Commonwealth of Australia. Chief information officer group instruction no. 1/2011. Departmental dissemination., May 2011.

Con13. Aquisition Community Connection. Test & evaluation community of practice, August 2013. https://acc.dau.mil/t&e.

Cro13. John Croft. 777 autothrottle design highlighted in asiana crash. *Aviation Week and Space Technology*, December 16 2013. http://www.aviationweek.com/Article.aspx?id=/article-xml/AW_12_16_2013_p37-646246.xml.

CSK13. Campaign to Stop Killer Robots. Website, 2013. http://www.stopkillerrobots.org/.

CT07. P. Cutler and J. Tweedale. The Present Development and Future Concept for a Situation and Threat Assessment Decision Aid. In *Proceedings of the Twelfth Australian Aeronautical Conference*, Melbourne, Victoria, Australia, 19-22 March 2007.

Dao13. James Dao. Drone Pilots Are Found to Get Stress Disorders Much as Those in Combat Do. *New York Times*, 22 February 2013. http://www.nytimes.com/2013/02/23/us/drone-pilots-found-to-get-stress-disorders-much-as-those-in-combat-do.html?_r=0.

Dep09. Department of Defence. Command and Control. ADDP 00.1, 2009. Commonwealth of Australia http://www.defence.gov.au/adfwc/Documents/DoctrineLibrary/ADDP/ADDP_00_1_Command_and_Control.pdf.

DF99. Marek J Druzdzel and Roger R Flynn. Decision support systems. In A. Kent, editor, *Encyclopedia of Library and Information Science*, volume 10, page 2010. Marcel Dekker, Inc., 1999. http://www.pitt.edu/~druzdzel/psfiles/dss.pdf.

DoD11. United States of America Department of Defense. The unmanned systems integrated road map fy2011-2036. Technical Report Reference Number: 11-S-3613, 2011. http://info.publicintelligence.net/DoD-UAS-2011-2036.pdf.

DoD12. United States of America Department of Defense. Autonomy in weapons systems. Directive Number 3000.09, Nov 2012. http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf.

DRBU09. M.S. Darms, P.E. Rybski, C. Baker, and C. Urmson. Obstacle detection and tracking for the urban challenge. *Intelligent Transportation Systems, IEEE Transactions on*, 10(3):475–485, 2009.

Duf13. Eliott Duff. Mining in the cloud. Presentation at the Sweedish Mining Conference, Perth, February 25 2013. http://www.youtube.com/watch?v=jOaBEyPX0jM.

EBRW07. Éloi Bossé, Jean Roy, and Steve Wark. *Concepts, Models, and Tools for Information Fusion*. Artech House Inc., 685 Canton St, Norwood, MA 02062, USA, 2007. ISBN-13: 978-1-59693-081-0.

End95. Mica R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995.

End03. Mica R Endsley. *Designing for situation awareness: An approach to user-centered design*, chapter 10: Automation and Situation Awareness, pages 173–191. Taylor & Francis US, 2003. ISBN 978-1-4200-6355-4.

FBCC+11. D. (1) Ferrucci, E. (1) Brown, J. (1) Chu-Carroll, J. (1) Fan, D. (1) Gondek, A.A. (1) Kalyanpur, A. (1) Lally, C. (1) Welty, J.W. (2) Murdock, E. (3) Nyberg, J. (4) Prager, and N. (5) Schlaefer. Building watson: An overview of the deepqa project. *AI Magazine*, 31(3):59–79, 2011.

Fin08. Anthony Finn. Legal considerations for the weaponisation of unmanned ground vehicles. *Int. J. Intelligent Defence Support Systems*, 1(1):43–74, 2008.

Fis13. Adam Fisher. Inside google's quest to popularize self-driving cars. *Popular Science*, September 2013. http://www.popsci.com/cars/article/2013-09/google-self-driving-car.

Fra13. Matthew Francey. The Colorado Town Promising to Shoot Down Obama's Surveillance Drones, September 2013. http://www.vice.com/read/drone-hunter.

GHZ12. S. Gupta, J. Hare, and Shengli Zhou. Cooperative coverage using autonomous underwater vehicles in unknown environments. pages 1–5, 2012.

GW10. Tony Gillespie and Robin West. Requirements for autonomous unmanned air systems set by legal issues. *The International C2 Journal*, 2(2):1–32, 2010. http://www.dodccrp.org/html4/journal_v4n2.html.

Hea13. Douglas Heaven. Higher state of mind. *New Scientist*, pages 32–35, 10 August 2013.

Hei73. Robert A. Heinlein. *Time Enough For Love*. G.P.Putnam's Sons, 1973. ISBN 0-399-11151-4.

Hey13. Christof Heyns. Report of the special rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns. Report A/HRC/23/47, Office of the United Nations High Commissioner for Human Rights, Palais des Nations,CH-1211 Geneva 10, Switzerland, April 2013. http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf.

HG09. Benjamin Hardin and Michael A. Goodrich. On using mixed-initiative control: A perspective for managing large-scale robotic teams. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, HRI '09, pages 165–172, New York, NY, USA, 2009. ACM.

HJBU13. Robert R. Hoffman, Matthew Johnson, Jeffrey M. Bradshaw, and Al Underbrink. Trust in Automation. *IEEE Intelligent Systems*, 28(1):84–88, 2013.

Int13. International Committee of the Red Cross. The Geneva Conventions of 1949 and their Additional Protocols, 2013. http://www.icrc.org/eng/war-and-law/treaties-customary-law/geneva-conventions/index.jsp.

JAC03. Agent oriented software, jack web site, 2003. http://www.agent_software.com/shared/home.

Kah11. Daniel Kahneman. *Thinking, fast and slow*. Farrar, Straus and Giroux, 2011. ISBN-10: 0374533555 ISBN-13: 978-0374533557.

Kel11. Jakob Kellenberger. Keynote address. In *International Humanitarian Law and New Weapon Technogies, 34th Round Table on Current Issues of International Humanitarian Law*, pages 5–6, San Remo, Italy, 8-10 September 2011. http://www.iihl.org/iihl/Documents/JKBSan%20Remo%20Speech.pdf.

Lam99. D. A. Lambert. Advisers with attitude for situation awareness. In *Proceedings of the 1999 Workshop on Defense Applications of Signal Processing*, La Salle, Il., USA, 1999.

Lam09. Dale A. Lambert. A blueprint for higher-level fusion systems. *Inf. Fusion*, 10(1):6–24, January 2009.

Lan09. William Langewiesche. *Fly By Wire: The Geese, The Glide, The 'Miracle' on the Hudson*. Penguin Books, 2009. ISBN: 978-1-84-614308-3.

LN08. D. Lambert and C. Nowak. Mephisto conceptual framework. Technical Report DSTO-TR-2162, Defence Science and Technology Organisation, PO Box 1500, Edinburgh SA 5108, Australia, 2008.

LR98. D.A. Lambert and M.G. Relbe. Reasoning with tolerance. In *Knowledge-Based Intelligent Electronic Systems, 1998. Proceedings KES '98. 1998 Second International Conference on*, volume 3, pages 418–427 vol.3, Apr 1998.

LS07. Dale Lambert and Jason Scholz. Ubiquitous command and control. *Intelligent Decision Technologies*, 1(3):157–173, Jan 2007. http://iospress.metapress.com/content/X447X16057766323.

Mea10. Lucas Mearian. Regulators blame computer algorithm for stock market 'flash crash'. *Computer World*, October 1 2010. http://www.computerworld.com/article/2516076/financial-it/regulators-blame-computer-algorithm-for-stock-market--flash-crash-.html.

Mel09. Nils Melzer. Interpretive guidance on the notion of direct participation in hostilities under international humanitarian law, 2009. http://www.icrc.org/eng/resources/documents/publication/p0990.htm.

Mil56. George A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2):81 – 97, 1956.

Min88. M. Minsky. *Society Of Mind*. Touchstone book. Simon & Schuster, 1988.

MIO87. J.D. Musa, A. Iannino, and K. Okumoto. *Software Reliability: Measurement, Prediction, Application*. McGraw-Hill Book Company, 1987. ISBN 0-07-044093-X.

MM97. RR Murphy and A Mali. Lessons learned in integrating sensing into autonomous mobile robot architectures. *JOURNAL OF EXPERIMENTAL & THEORETICAL ARTIFICIAL INTELLIGENCE*, 9(2-3):191–209, APR-SEP 1997.

MoD11. United Kingdom Ministry of Defence. The UK approach to Unmanned Aircraft Systems. Joint Doctorine Note 2/11, March 2011. https://www.gov.uk/government/publications/jdn-2-11-the-uk-approach-to-unmanned-aircraft-systems.

Mod14. Naz K. Modirzadeh. Turning Point or Breaking Point? 2014 Red Cross Oration, Adelaide, South Australia, October 2014.

MRT12. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. The MIT Press, 2012. ISBN 026201825X, 9780262018258.

MRV12. Elke Muhrer, Klaus Reinprecht, and Mark Vollrath. Driving with a partially autonomous forward collision warning system: How do drivers react? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(5):698–708, 2012.

Mur04. R. Murch. *Autonomic Computing*. IBM Press., 2004.

Mus04. John D. Musa. *Software Reliability Engineering: More Reliable Software Faster and Cheaper*. Author House, 1663 Liberty Drv., Suite 200, Bloomington, Indiana 47403, 2nd edition, 2004. ISBN: 1-4184-9387-2 Order from http://www.authorhouse.com.

MW09. R.R. Murphy and D.D. Woods. Beyond Asimov: The Three Laws of Responsible Robotics. *Intelligent Systems, IEEE*, 24(4):14–20, 2009.

Nut13. Robert Nutbrown. Landing Intelligence: BAE's autonomous landing systems for UAVs. *Australian Aviation*, 304:61–62, May 2013.

Off08. Office of the Under Secretary of Defense For Acquisition, Technology, and Logistics. Report of the Defense Science Board Task Force on Developmental Test & Evaluation, May 2008. http://www.acq.osd.mil/dsb/reports/ADA482504.pdf.

OG13. Nathan Olivarez-Giles. DARPA wants to build a fleet of underwater drones for the Navy, 5 September 2013. http://www.theverge.com/2013/9/5/4699338/darpa-hydra-underwater-navy-mothership-drone-fleet.

PBOC05. S. Paradis, A. Benaskeur, M. Oxenham, and P. Cutler. Threat evaluation and weapons allocation in network-centric warfare. In *Information Fusion, 2005 8th International Conference on*, volume 2, pages 8 pp.–, 2005.

PHG00. A.R. Pearce, C. Heinze, and S. Goss. Enabling perception for plan recognition in multi-agent air mission simulations. In *MultiAgent Systems, 2000. Proceedings. Fourth International Conference on*, pages 427–428, 2000.

PM02. Ross Pigeau and Carol McCann. Re-conceptualizing Command and Control. *Canadian Military Journal*, 3(1):53–64, 2002. http://www.journal.forces.gc.ca/vo3/no1/doc/53-64-eng.pdf.

PMM93. J.H. Poore, Harlan D. Mills, and David Mutchler. Planning and certifying software system reliability. *IEEE Software*, 10(1):88–99, January 1993.

PPG07. Michael Papasimeon, Adrian R. Pearce, and Simon Goss. The Human Agent Virtual Environment. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, AAMAS '07, pages 281:1–281:8, New York, NY, USA, 2007. ACM.

Pro92. G.J. Probst. *Landsberg, verlag moderne industrie.*, chapter Organisation. Strukturen, Lenkungsinstrumente und Entwicklungsperspektiven. 1992.

RT13. Shane R. Reeves and Jeffrey S. Thurnher. Are we reaching a tipping point? how contemporary challenges are affecting the military necessity-humanity balance. *Harvard Law School National Security Journal*, June 2013. Harvard Law School,Caspersen Student Center,Suite 3039,1585 Massachusetts Ave.,Cambridge, MA 02138 http://harvardnsj.org/2013/06/are-we-reaching-a-tipping-point-how-contemporary-challenges-are-affecting-the-military-necessity-humanity-balance/.

SA09. N. Steiner and P. Athanas. Hardware autonomy and space systems. In *Aerospace conference, 2009 IEEE*, pages 1–13, 2009.

SAB12. Gilson Yukio Sato, Hilton José Azevedo, and Jean-Paul A. Barthès. Agent and multi-agent applications to support distributed communities of practice: a short review. *Autonomous Agents and Multi-Agent Systems*, 25(1):87–129, July 2012.

Sat09. Guru, all latest updates: Herbert simon, March 2009. http://www.economist.com/node/13350892 viewed November 2014.

Sat14. Satisficing. Wikipedia Article, 2014. https://en.wikipedia.org/wiki/Satisficing viewed November 2014.

Sca13. Jeremy Scahill. Economist interview with Jeremy Scahill on his book Americas Dirty Wars, 2013. http://www.economist.com/blogs/democracyinamerica/2013/05/jeremy-scahill-americas-dirty-war.

Sch13. Michael N. Schmitt. The Tallinn Manual on the International Law Applicable to Cyber Warfare. Cambridge University Press, 2013. http://www.ccdcoe.org/249.html.

SH10. I Sutskever and G Hinton. Temporal-kernel recurrent neural networks. *NEURAL NETWORKS*, 23(2):239 – 243, 2010.

SH12. R. Salakhutdinov and G. Hinton. An efficient learning procedure for deep boltzmann machines. *Neural Computation*, 24(8):1967 – 2006, 2012.

She11. Gene I. Sher. Evolving chart pattern sensitive neural network based forex trading agents. *CoRR*, abs/1111.5892, 2011.

She13. GeneI. Sher. Evolving currency trading agents. In *Handbook of Neuroevolution Through Erlang*, pages 785–824. Springer New York, 2013.

Sin09a. Peter W. Singer. Attack of the Military Drones. The Brookings Intitute, June 2009. http://www.brookings.edu/research/opinions/2009/06/27-drones-singer.

Sin09b. Peter Warren Singer. *Wired for War: The robotics revolution and conflict in the twenty-first century*. Penguin.com, 2009.

SLGS12. J. Scholz, D. Lambert, D. Gossink, and G. Smith. A blueprint for command and control: Automation and interface. In *Information Fusion (FUSION), 2012 15th International Conference on*, pages 211–217, 2012.

Sof13. Cougaar Software. ActiveEdge - An Intelligent Decision Support Platform, 2013. http://www.cougaarsoftware.com/activeedge/activeedge-overview.htm.

SOL+11. Julie N. Salcedo, Eric C. Ortiz, Stephanie J. Lackey, Irwin Hudson, and Andrea H. Taylor. Effects of Autonomous vs. Remotely-Operated Unmanned Weapon Systems on Human-Robot Teamwork and Trust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 55(1):635–639, 2011.

Spo09. Olaf Sporns. From complex networks to intelligent systems. In Bernhard Sendhoff, Edgar Krner, Olaf Sporns, Helge Ritter, and Kenji Doya, editors, *Creating Brain-Like Intelligence*, volume 5436 of *Lecture Notes in Computer Science*, pages 15–30. Springer Berlin Heidelberg, 2009. http://dx.doi.org/10.1007/978-3-642-00616-6_2.

Sua13. Daniel Suarez. The kill decision shouldn't belong to a robot. Ted Talk, June 2013. http://www.ted.com/talks/daniel_suarez_the_kill_decision_shouldn_t_belong_to_a_robot.html.

Twe12. Jeffrey W. Tweedale. Using Mutli-Agent Systems to Improve the Level of Autonomy for Operators Controlling Unmanned Vehicles. In *Frontiers in Artificial Intelligence and Applications*, volume 243: Advances in Knowledge-Based and Intelligent Information and Engineering Systems, pages 1666 – 1675. IOS Press, 2012.

Twe13. Jeffrey W. Tweedale. Using Multi-Agent Systems to Pursue Autonomy with Automated Components. In *Proceedings of 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems — KES2013*, Kitakyushu, Japan, 9-11 September 2013.

UIT+11. Dieter Uckelmann, M.-A. Isenberg, M. Teucke, H. Halfar, and B. Scholz-Reiter. Autonomous control and the internet of things: Increasing robustness, scalability and agility in logistic networks. In Damith C. Ranasinghe, Quan Z. Sheng, and Sherali Zeadally, editors, *Unique Radio Innovation for the 21st Century*, pages 163–181. Springer Berlin Heidelberg, 2011.

U.S10. U.S. Securities and Exchange Commission (SEC) and the Commodity Futures Trading Commission (CFTC). Findings Regarding the Market Events of May 6, 2010. Technical report, SEC, September 2010. http://www.sec.gov/news/studies/2010/marketevents-report.pdf.

Var14. Various. Deep space 1, 2014. http://en.wikipedia.org/wiki/Deep_Space_1 viewed January 2014.

Woo09. Michael Wooldridge. *An Introduction to Multi-Agent Systems*. Bell & Bain, Glasgow, 2nd edition edition, 2009. ISBN 978-0-470-51946-2.

THIS PAGE IS INTENTIONALLY BLANK

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA	1. CAVEAT/PRIVACY MARKING

2. TITLE	3. SECURITY CLASSIFICATION
Issues Regarding the Future Application of Autonomous Systems to Command and Control (C2)	Document (U) Title (U) Abstract (U)
4. AUTHOR	5. CORPORATE AUTHOR
Michael Pilling	Defence Science and Technology Organisation Fairbairn Business Park, Department of Defence, Canberra, ACT 2600, Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO–TR–3112	AR-016-315	Technical Report	June 2015

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. No. OF PAGES	12. No. OF REFS
2014/1117385/1	07/429	Dr Todd Mansell CJOAD	71	103

13. URL OF ELECTRONIC VERSION	14. RELEASE AUTHORITY
`http://www.dsto.defence.gov.au/publications/scientific.php`	Chief, Joint and Operations Analysis Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for Public Release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SOUTH AUSTRALIA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS

No Limitations

18. DSTO RESEARCH LIBRARY THESAURUS

Autonomous Systems, Automated Reasoning, Automation, Artificial Intelligence, International Humanitarian Law

19. ABSTRACT

This broad review provides some insights into the vast field of Autonomous Systems in the context of Defence applications and C2 in particular. There are significant challenges in the areas of human-computer interaction and the legalities of war that may trump technical issues in terms of impediments to automation. Many technical areas are also covered and the paper recommends developing or adopting strong engineering processes for building Autonomous Systems, research into human factors and strong engagement with the international community with respect to the Laws of Armed Conflict.