Chapter 18
Trusted Autonomous Command and Control

Noel Derwort

18.1 Scenario

> Unmanned systems and autonomous software offer significant potential advantages for meeting the challenges of a newly forming adversarial environment. Speed of light cyber-attacks, anti-access/area-denial (A2SD) actions that keep our forces operating at a distance, and potential attacks on our space-based assets all require innovative solutions for maintaining mission effective air, space and cyber operations in the face of these new challenges.
>
> Mica R. Endsley, Chief Scientist United States Air Force, 1 June 2015 [1]

As highlighted by Endlsey there are significant opportunities brought about by advances in technology, these opportunities can be equally exploited by allies and adversaries. Autonomous systems and enhanced human-cyber-machine interaction may well provide the essential linkages required in order for the 'human' to keep pace with the decisions and actions occurring around them. The implications surrounding the exploitation of these technologies and autonomous systems, such as 'Artificial Intelligence' (AI), will likely challenge many established rules and norms - this led to an open letter being announced at the July 2015, International Joint Conference on Artificial Intelligence held in Buenos Aires. The letter was subsequently signed by 'thousands' of scientists articulating their fears over the potential significant adverse effects of the militarisation of AI.

> The key question for humanity today is whether to start a global AI arms race or to prevent it from starting. If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow. Unlike nuclear weapons, they require no costly or hard-to-obtain raw materials, so they will become ubiquitous and cheap for all significant military powers to mass-produce. We therefore believe that a military AI arms race would not be beneficial for humanity. There are many ways in

N. Derwort (✉)
Department of Defence, Canberra, Australia
e-mail: noel.derwort@defence.gov.au

H. A. Abbass et al. (eds.), *Foundations of Trusted Autonomy*, Studies in Systems, Decision and Control 117, https://doi.org/10.1007/978-3-319-64816-3_18

> which AI can make battlefields safer for humans, especially civilians, without creating new tools for killing people [2].

There are clear parallels in the extract to July 1945, when Leo Szilard and 69 fellow workers on the Manhattan Project co-signed a petition, seeking to urge the President of the United States, Harry S Truman, to consider carefully the decision to employ the atomic bomb against Japan [3]. Their concerns regarding the burden of responsibility over the precedence and implications following the employment of the weapon never reached the President - although the petition served as a prescient warning, noting the ensuing 'Cold War' and era of the 'Mutually Assured Destruction' doctrine. If history were to be repeated, then looking forward AI and weaponised autonomous systems are a likely outcome.

Imagine in the turbulent August 2030 winter oceans off the Western Australian coast, far out in the Indian Ocean, 1175 km south-south west of Cocos-Keeling Islands, an autonomous - and to this point silent - sea glider chooses to alter its parameters to better trail a confirmed contact. Concurrently it sends a quantum encrypted picoburst transmission to one of its near neighbours. The neighbouring glider modifies its own profile, surfaces and shares awareness through the omnipresent stealthy UAV orbiting high overhead. The message is relayed through to the Command and Control centre where the implications are analysed and modifications to the remainder of the Theatre Anti-Submarine Network are calculated and relayed back. A further series of encrypted burst transmissions inform the changes - reshaping the net to guarantee contact is maintained whilst developing response options without compromising the strength of the overall surveillance network. Such movements are rare and although the modifications to the network will be achievable with-in tolerable levels, there will be very real implications for the logistical and technical supporting chain. These processes are modified and transshipment of the required equipment ordered. The unique identity of the contact also triggers the need for a potential kinetic outcome, building on the non-kinetic actions already in train - accordingly specific loads are apportioned for the pending logistics sustainment flight as well as initiating a line on the subsequent Air Task Order. The final message transmitted is to a Royal Australian Navy submarine about to commence a patrol. As all actions to date are within the requisite authorities, Commanders Intent; Rules of Engagement; and Commanders Intelligence, Indicators and Warnings Requirements; when the message is finally received by the submarine, it is for the first time read, and responded to, by a human.

The glider and its broad array of peers, including other autonomous platforms, form part of an overarching autonomous monitoring and response network. The Theatre Anti-Submarine Network shield had been conceived and then urgently developed in the early 2020 s to overcome the challenge of surveilling and monitoring the Australian Area of Responsibility/Interest. The rapid increase in regional submarine capabilities outstripped the ability for the Australian defence acquisition organisation or its allies to keep pace, and when combined with the vast area to be covered it simply precluded historical surveillance means being effective. The ability for autonomy combined with extended range, relatively low cost and wide area coverage (through

deployed networks) made the glider a critical enabler, and when coupled with the recently developed autonomous Command and Control System it proved to be an effective answer. Getting to the level of autonomy required for ultimate success was a series of stepping stones - each breaching further into the tide of technological change to ultimately reach the goal. Looking back the steps are easier to see than they were looking forward, with two distinct paths merging: overall technology and the ability to support/enable better decisions, particularly with 'on/outside the loop' frameworks; the second being cyber and social media triggering an evolution in thinking, exacerbated by an environment typified by dissonance in the global rules based order.

History is replete with examples of technological leaps born of necessity and opportunity. A little over a decade after man's first flight in December 1903 the British Army was employing aircraft such as the Avro 504 in 1914 for observation and re-connaissance, by the end of the First World War aircraft had developed to the point where reasonable range and relatively accurate bombing was a reality. Whilst growth during the interwar period was steady, development during 1939–1945 was explosive - with the first jet fighters operational in 1944. Other technology and weapons developed apace, such as the invention and operational employment of radar. Weapons with intercontinental strike capability became a reality when the German V1 flying bomb and V2 Rocket (Guided Ballistic Missile) became operational in 1944. Arguably the culmination came on August 6 1945 when the first atomic bomb was dropped on Hiroshima Japan. While the arguments surrounding the moral/ethical and military justification remain today, the reality remains that one of the most powerful weapons devised was used against cities with significant loss of life and in turn opening the pandora's box of the atomic age - bringing life to Szilard's scientists concerns. One constant amid all of the development was the tenacity amongst all belligerents to equal or better the opposition. The need for tactical, operational and strategic equity (if overmatch could not be achieved) was seen as essential in order to succeed, or even survive. Following the Second World War and Germany's collapse both the Allied forces and Soviet Union raced to gain access to German rocket technology in order to secure and develop the capabilities for themselves. Wernher von Braun and a significant number of personnel surrendered to the US forces whilst the Soviets gained the V2 manufacturing facilities and technology. This race was indicative of future events.

United States Air Force Colonel John Warden highlights the deliberate steps of 'Observe, Orient, Decide, Act' (OODA) in decision making. Continual growth in the capability of weapon systems through the 1950s, 60s and 70s required a commensurate development in defensive and employment systems, driven by a requirement to process ever increasing amounts of data and react 'quicker' than the opposition 'human' - to get inside the OODA loop. Through this need and evolution came the US Aegis combat system. Drawing much from its namesake's Greek etymology, the Aegis was intended to use automation to provide a god like shield over a battleship and the surrounding battlespace by the detection, tracking and subsequent engagement of multiple incoming missiles or aircraft in a prioritised engagement. Once fixed on a target, the system relayed the oncoming threat's position to the ship's

main computer enabling the crew to 'quickly and decisively' determine defensive countermeasures engagement. The system was first fielded in early 1983 on the US Navy Ticonderoga class cruiser [4]. USS Vincennes (CG 49) was the third ship of the class and on 03 July 1988 she shot down Iran Air Flight 655 during a hostile engagement involving up to seven Iranian Revolutionary Guard gunboats [5].

The events surrounding the shooting down of the Iran Air flight offer much to the study of the challenges of command and control during a military engagement, even one where there is an apparent force overmatch - it would take a stretch to argue equity of combat power between an Aegis cruiser and a relatively small number of gunboats. The mindset and approach of Vincennes' Captain, combined the performance of her crew have certainly been called into question, and arguably were at least contributory to the ultimate outcome. As reported by the New York Times in 1988, a particular note of the event was the digital recording of the incident by the Aegis computerised defence system. It was clear that in spite of the automation of the Aegis working correctly the Vincennes' crew was reporting erroneously [6], and in the time critical engagement this was directly contributory to the tragic outcome [5]. Despite acknowledging crew errors, the Investigating Officer made the statement 'The fact is the sensors gave no clear piece of information that it was not an F-14. However, if the F-14 identification had never been made, the contact would have remained designated "unidentified assumed hostile." In that event, it is unlikely that the CIC Team would have proceeded any differently or elicited additional information in the extraordinarily short time available. As long as it remained a possible "hostile," the Commanding Officer would be obligated to treat it In the same manner as he would an F-14' [5]. Contrarily the Aegis sensors were clear in displaying a civil flight, not an F-14. Such comments allow questions over the Commander's intent and more importantly the suitability of human vice 'machine' decisions to be raised, and they have been ever since. The Vincennes Commander, and the Investigating Officer, demonstrated the limitations of human frailty, and behaviours - regardless of the implications, or accountabilities they held. There is little doubt in the strategic consequences of the crew's actions. In this instance, had the automation been 'employed' fully, it is entirely possible this tragedy could have been averted - further this example provides fertile ground for understanding the challenges surrounding human trust and performance of automation.

Despite setbacks and errors, capability increases in radar, missile and Close In Weapon Systems (CIWS) technology systems such as Aegis demonstrated overwhelmingly the potential for 'area' defence weapons. Counter - Rocket Artillery and Mortar (C-RAM) systems became operational in the Iraq theatre in 2004. Israel sub-sequently operationally deployed its Iron Dome system in 2011 [7]. Such systems were specifically designed to operate in a defensive mode through the active engagement of incoming short range rockets, mortars and other projectiles. The time available for engagement preclude human intervention, making the Iron Dome one of the first man-on/outside-the-loop systems - albeit recognising its operations were still bound by relatively traditional rules based autonomy. Even in the early years of its employment some authors and commentators raised the challenges created when the relative comfort of living beneath the shield arguably decreases the motivation

to address the cause of the conflict [7]. A potential side effect of taking humans out of the decision, authority and responsibility chain may have been the de-humanising both of the weapon systems and the conflict itself. These intangible risks/challenges were seen as acceptable when weighed against the advantages borne of high success rates in de-fending against incoming strikes. Fear was replaced by autonomy enabled confidence.

By 2016 China's assertions for sovereignty and rights over the South China Sea had been the subject of an International Tribunal for the Law of the Sea (ITLOS) ruling. Arguing increasing legitimacy over its claims with the justification of the 1947 'nine dash line', China proceeded with 'unprecedented' reclamation of maritime features within the area. Tensions had escalated during 2008–2012, particularly with the Philippines [8]. Linkages to the Chinese A2/AD approach held sway with a number of academics and military planners like Aaron Friedman, with a reasonable case made for the both reactionary and considered nature of this strategy [9]. A key premise surrounded the defensive cordon offered by the reclaimed islands. Arguably more challenging for the Chinese Government was the concerns of regional governments over disputed claims. The situation came to a head when the ITLOS ruling was made in favour of the Philippines. The Tribunal rejected Chinese historical claims over the South China Sea and the 'nine dash line' as well as ruling that China had violated Philippine sovereign rights. The reaction to the ruling by Beijing was a swift rejection, with broader emotional and nationalistic sentiment strongly opposing the finding being voiced [10]. The Chinese reaction could have been anticipated noting their earlier rejection of the validity of the Tribunal and its very legitimacy to consider the claim [11, 12]. An interesting twist to the announcement was prophetic suggestions that the Ruling may ultimately push the recently sworn Philippine administration of President Duterte more toward Beijing [12].

In the middle of the 2018 Afghan fighting season the war had dragged relentlessly on. The move by ISILs Afghanistan offshoot, the Islamic State-Khorosan (IS-K), to increase its foothold in the war torn country effectively re-escalated a multiple front conflict for the allied forces. This was followed almost immediately by escalations in Libya and outbreaks in Africa which stretched the allied ability to counter the threat Despite the offensives, the IS group had been in overall decline. The exception to the decline was its continuing ability to draw Foreign Fighters, including a highly educated, tech savvy youth element and the group's inexplicable ability to draw willing martyrs. These two elements combined in what was to arguably become the first example of autonomous control over 'humans' and 'human weapon systems'.

Having depleted significant elements of their fighter force the IS group was keen to identify alternate means to overcome the significant technological and military overmatch brought to bear by the allied forces - the war was clearly at a significant turning point. The non-state actor had already managed to create a conflict during which they spurned the global pillars of the Westphalian order and Geneva Convention - the very notion of the Sykes-Picot agreement was an anathema to the group. Such disregard for global rules and norms should have been an indicator of things to come. The overwhelming asymmetry of the intelligence and power projection of

the combined allied forces drove an adversary, already known for its barbarity, to abnegate ethical boundaries constraining most global powers.

Already able to harness levels of fervour barely imaginable, and demonstrating a complete disrespect for the rules of war, the IS group exploited their strengths - blind commitment of their followers; combined with a technical (and psychological) mastery of a multitude of cyber networks, including social and communication media systems. In an activity that defied comprehension, every technology was exploited, and in the absence of any normal/moral constraints they re-wrote the rules. The IS group managed to develop a system employing active learning 'big data' analytics combined with social media communications to both effectively get inside the allied OODA loop as well as subverting established command and control structures. The system operated on two levels: first a simple automated warning system triggered an avoidance/withdrawal response for IS fighters exposed to a high likelihood of kinetic strike. This had been achieved by breaching several layers of security through the inadequacies of a lower-order allied partner which enabled the group to develop an 'operating picture' of the broader coalition air picture, this in turn provided the 'fight-er on the street' a 'last minute' warning to take evasive action. The second level of the system - crudely known as Martyr Net - prioritised the notification and messaging against a simple algorithm matching density of fighters against the tactical and strategic weight of the target relative to the potential threat. Martyr Net was considered entirely autonomous and its operational effect easily outweighed the complete absence of 'concern' over the human fighter - they were after all nothing more than a weapon whose effect needed to be maximised. The worth of the system became apparent within days of activation with clear and demonstrable effect. The darker side of Martyr Net was the offensive stream, continually enhanced through its active learning.

In essentially the same manner as predicting strikes, Martyr Net was able to predict likely targets of opportunity. Through active monitoring of social and communications media, Martyr Net identified potential opportunities for optimising attack effectiveness by increasing target fatality rates through employment of suicide bomber attacks in order to maximise fear, and seemingly perversely, retaliation. A series of bots generated automated messages to be sent, again for algorithmic proximity prioritised targets, however this time to effect a kinetic attack - with the bot engaging both the suicide bomber and intended victims. The suicide bomber had no real awareness the message was sent by a 'system', however, the tactical results were devastating with an upswing in the ratio of victims per attack. Logically, Martyr Net identified a relationship between the offensive and defensive streams and then determined where the two aspects could be made coincident, namely triggering an attack with sufficient warning in the right geographic area, resulting in allied weapons occasionally being dynamically diverted from their initial intended strike in order to respond to an emergent threat - following the 'diversion' the IS were able to 'withdraw' from the original target before allied re-engagement. Crude, inhuman and effective it proved capable of undermining the effectiveness of one of the most militarily powerful coalitions through the exploitation of human failings combined with ubiquitous social and communications media. The IS leaders paid no heed to

global rules and norms, enabling full control and authority for their operations in the capacities of Martyr Net - with the operations only bounded by the networks they were operating on.

The early 2019 allied forces identification of the behavioural and tactical change undertaken by the IS group was relatively swift, the understanding of the cause took some time longer - however the subsequent reaction was immediate and far reaching. In the Syrian area of the conflict the proximity to the Russian forces had already permitted observation of, and access to, allied capabilities on an unprecedented scale - particularly as the US were forced to employ 'accelerants' to address political pressures to end the conflict through the increasing application of high end systems; further exposing their capabilities and limitations. Whilst Martyr Net had been employed by the IS Group across its campaign, in Syria both the US and Russian intelligence hierarchies observed its effect and identified implications for the future. Much like the race for the German V2 rocket technology and the emergent atomic age of six decades past, the race was on to both develop and field systems to counter Martyr Net, as well as construct similar systems of their own to retain the advantage - reliving the predictions of Szilard in 1945 and the July 2015, International Joint Conference on Artificial Intelligence open letter, this time in the arena highlighted by Endsley [1].

In 2027 China has reasserted its claims for key resources and territory in the South China Sea. The relationship between US and Philippines had gradually declined through a combination of exploitation of soft power by China, buying votes and support on one level and absolute and vocal support for the Philippines' President in his populist harsh rule of law, targeting drug and criminal cartels on another. Repeat-ed humanitarian and human rights violations on the part of the Philippines regime pressured the distancing with successive US administrations, with the ultimate pulling out of US forces from the long time ally, reminiscent of the withdrawal in 1991/92. Concurrently China invested heavily in the Philippines and through its overt support to, and covert pressure on, the Philippines Government was able to execute land reclamation and occupation of first Scarborough Reef, and then Second Thomas Shoal with 'manageable' reactions from the global community. The evolutions within the South China Sea were starkly juxtaposed by the events to the north in the East China Sea.

The Sino-Japanese relationship continued to decline with escalatory negative and hostile rhetoric with commensurate behaviours surrounding their dispute over the Diaoyu/Senkaku Islands. Long held fears that increased militarisation within the region, including a series of naval encounters and reportedly unsafe air intercepts culminated in the 2025 sinking of the Chinese PLAN Frigate Jing Zhou (FFG 532) during an engagement with the JSDF JS Ashigara (DDG 178). Although historically engagements were tense, with both sides offering provocations including occasional warning shots, during this incident an 'overly zealous' junior Chinese Officer pre-emptively prepared a load/launch sequence for a YJ-83 Anti-Ship cruise missile. An error in process saw the missile launched - completely outside parameters for the relatively close range encounter. Well within the parameters of Ashigara's updated Aegis defensive systems, including its phalanx CIWS (this dealt with the errant YJ-

83) which should have been the kinetic end point, however a commensurate error on Ashigara saw her Mk 45, 5 in. gun, loaded with a full 20 round automatic load (in the unmanned gun mount) triggered independently of the Aegis system. In under a minute the full 20 round salvo was fired at Jing Zhou with several rounds clearing her own defensive systems, ultimately resulting in the loss of the ship. Despite the obvious differences, parallels were quickly drawn to the USS Vincennes incident. In 'independent' parallel investigations each unsurprisingly blamed the other party, however both were unanimous (and accusatory) in highlighting that if allowed to operate 'autonomously' both ships weapons and defensive systems would have prevented the incident. Indeed the investigations found human interference with both the defensive system and the original decision to strike, vice reliance on those same defensive systems, was erroneous, an overreaction, and disproportionate use of force. Despite calls for reparations the key outcome was to highlight again the frailties of man in the loop systems.

Through public condemnation and pressure, comparisons were made to the successful and widespread use of the CRAM and Iron Dome systems of the 2010s, leading to calls for increased autonomy in theatre based defensive systems. Amidst the evolutions China perceived/claimed a significant Japanese/US capability overmatch in the region allowing it to justify its announcement for the first time in 2028 of its own Dragon Dome - a fully autonomous integrated theatre defence system. The Chinese believed and argued that all of the US activities within the region were designed to overthrow the government and only an autonomous system would guarantee that humans could not make the same errors. A key argument within the Chinese commentary was accusations the US (and Russia) had already developed and deployed autonomous systems following their earlier exploitation of IS group material. China also declared a cyber-equivalent, designed to protect and respond to attack in the cyber domain.

During the intervening years, US and 'Western' militaries had suffered from successive dilettante political leadership. The shift in global order had seen the realisation of an era of three superpowers. The rise and rise of China, paralleled in Europe by the rise of Russia - exacerbated by a fragmentation and ultimate disintegration of the European Union - saw the US as a relatively weakened 'Super-power'. India had also continued its ascendancy and was on the cusp of 'joining the major-power club'. The US position in the global order had been significantly impacted by long periods of protectionism and anti-globalisation movements, each moving to erode global credibility and sway. This period also saw unprecedented and increased cost per unit growth for 6th generation capabilities, placing them beyond the reach of most nations. Concurrently cyber capabilities had experienced exponential growth in capability at an inversely proportional drop in price. Cyber enabled weapons became the only means to balance the perceived destabilising overmatch being held by the increasingly belligerent and aggressive super-powers. Indeed ongoing selective adherence to international rules and norms changed the shape of 'ethical, moral and human' debate leaving the door open for fully autonomous systems, particularly in the cyber realm.

It came as no surprise, to any hawkish observer, when in 2029 in response to the Chinese Dragon Dome announcement the US responded with public acknowledgement they too had developed and deployed an Autonomous Defence Command and Control system, which had also been made available to key and favoured allies: England, Scotland, France, Germany, Australia and re-unified Korea. Public statements were vague however key attributes of the system included trusted, reliable, 'deep learning', autonomous 'point of target termination', prioritised simultaneous non-kinetic and kinetic effect, Command by negation capability enabled, 'absolute' adherence to Area of Operations boundaries and of course 'absolute' adherence to rules of engagement. The era of the trusted autonomous Command and Control had arrived and questions over moral and ethical equities were either simply ignored, or likened to a 'Nash Equilibrium' we have too, they have....

By late August 2030 in the Indian Ocean off the Western Australian coast, approximately 1250 km south-south west of Cocos-Keeling Islands, the autonomous sea-glider continues on its silent course. Having identified its target as a hostile submarine the glider commanded its peers to move into a sheparding pattern, each sensing and adapting to the environment and changing tactical situation. Patience was a human virtue the gliders regularly demonstrated, routinely adapting and resetting as the target appeared to out maneuver them through a slight speed advantage. As the hunt progressed, despite the best efforts of the target submarine captain and crew, the glider force identified repetitions in the targets behaviours, enabling the force to intuitively adapt their actions. The gliders directed the Australian Submarine in support to transit to a uniquely identified location, exploiting the water column variations and environmental challenges to lie in ambush for the target submarine.

When the enemy finally enters the perfect location for a firing solution from the Australian submarine the optimum glider overtly broadcasts the enemy's position, defeating both the offensive and defensive strengths of the submarine. Much like the Lyre bird native to Australia the glider mimics an Australian submarine in an active track mode - concurrently broadcast openly, triggering and immediate defensive and de-escalatory departure by the enemy submarine. The mimic message is carefully choreographed through the slower gliders to maintain the ruse as the submarine de-parts the area.

When the autonomous glider fleet is questioned during the debrief as to why they acted in the manner chosen - to drive the enemy submarine from the area rather than executing a kinetic attack, the response was both a shock and a revelation. The glider reported it had exercised 'mission command' and operated outside its initial plan as it believe it had identified a course of action which would achieve improved operational and strategic effectiveness, with a minimal and acceptable level to the tactical scenario - and it was prepared to achieve the directed strategic endstate over its own progression. Trusted autonomous command and control systems were here to stay.

References

1. M. Endsley, *Autonomous Horizons: System Autonomy in the Air Force-a Path to the Future (Volume I: Human Autonomy Teaming)* (US Department of the Air Force, Washington, 2015)
2. http://futureoflife.org/open-letter-autonomous-weapons. Accessed 23 July 2016
3. https://www.theguardian.com/world/2016/jul/12/philippines-wins-south-china-sea-case-against-china. Accessed 27 July 2016
4. https://www.trumanlibrary.org/whistlestop/study_collections/bomb/large/documents/index.php?documentdate=1945-07-17&documentid=79&pagenumber=1. Accessed 30 July 2016
5. http://www.lockheedmartin.com.au/us/100years/stories/aegis.html. Accessed 22 July 2016
6. http://www.nytimes.com/1988/11/18/opinion/witness-to-iran-flight-655.html?pagewanted=all&src=pm. Accessed 22 July 2016
7. http://thediplomat.com/2015/07/south-china-sea-philippines-v-china/ (2015). Accessed 26 July 2016
8. Al Jazeera English. Israel deploys 'iron dome' rocket shield, http://www.aljazeera.com/news/middleeast/2011/03/201132718224159699.html (2011). Retrieved 26 July 2016
9. W.M. Fogarty, Formal investigation into the circumstances surrounding the downing of iran air flight 655 on July 3, 1988. DTIC AD-A203 577. Retrieved 26 July 2016
10. A.L. Friedberg, Beyond air-sea battle. The Debate Over US Military Strategy in Asia (2014)
11. The Lowy Institute, South China sea: Conflicting claims and tensions, http://www.lowyinstitute.org/issues/south-china-sea. Accessed 27 July 2016
12. E. Graham, The bolt from the hague, http://www.lowyinterpreter.org/post/2016/07/13/the-bolt-from-the-hague1.aspx. Accessed 26 July 2016