Chapter 6
Legal Issues for UVS

Many of the advances in computing and technology have increased the decision-making capabilities of UVS to the point where they may now make truly independent decisions regarding by which route they might travel, what constitutes a target, and whether and by what means these targets should be engaged, possibly with lethal force. As a result, it is now widely acknowledged that, if permitted, the use of such systems would mark a sea-change in the role of technology in warfare as the human could potentially be removed from the decision-making loop. This introduces a number of issues that the need to be assessed in relation to the Law of Armed Conflict. In this chapter we discuss these issues.

Additionally, as UVS will need to work in environments shared by people, property and other vehicles, we also discuss the legal status of UVS, their operation in a shared environment, how they fit into the existing legal frameworks, who might be responsible for any infringements perpetrated by them, whether or not they should (or could) hold some sort of legal personality, and what the implications of such concepts might be.

It is recognised that such considerations will not resolve the issues over whether military UVS will one day operate in complex adversarial environments in a more reliable manner or autonomously target and control weapons. Moreover, there are many level-headed people who will consider it futile to discuss such concepts when we clearly have trouble making relatively simple UVS work. Regardless, such a discussion contributes by providing a way of thinking about the issues from the perspective of legal responsibility. Additionally, such deliberations also help to avoid technology outpacing the regulatory regimes and allow us to develop an understanding of existing technologies in the light of certain hypothetical possibilities, as well as exploring established legal concepts. This in turn enables us to consider what the responsibilities of users, owners, developers, etc. might be given the autonomy of UVS.

Furthermore, without a full understanding of such issues developers may be forced to focus on making supervised systems rather than fully autonomous ones, or users may refuse to accept liability for the unsupervised actions of UVS, thereby creating capability gaps. Additionally, if we do not frame the responsibilities of the UVS and those responsible for their construction and use, they will likely develop unfettered, with some humans progressively absolving themselves of any liability. As a result, adversaries who may be less restrained in

A. Finn and S. Scheding: Develop., Challenge. for Autonom. Unmann. Vehicles, ISRL 3, pp. 155–207.
springerlink.com
© Springer-Verlag Berlin Heidelberg 2010

their interpretation of their international obligations may accord UVS unlimited responsibility for their actions. In the case of the Law of Armed Conflict this may have far-reaching consequences.

6.1 Legal Issues for UVS Platforms

Regardless of whether or not UVS carry weapons they are usually large objects that move at speed and must avoid people, property, and other vehicles. They are also constructed of sub-systems and materials that are inherently hazardous. At one level this means we can treat UVS as "unremarkable technological artefacts, similar in nature to toasters or cars" [24] and for which the law has a highly developed set of principles that apply to product liability. That is, we can apply the law to UVS as purely commercial products and as there are many examples of intelligent, safety-critical systems whose malfunction may result in death or injury (e.g. medical equipment, railway signals, air traffic control systems) it is believed that the majority of the concerns usually touted as being possible dangers for UVS or their weaponised counterparts will fall under this mundane interpretation.

At present, however, as a result of the novelty of the technology, the difficulties with allocating responsibility, and an unwillingness to burden the relevant agencies responsible for drafting such laws with additional work, the use of most UVS is inadequately regulated. Nevertheless, there is a great deal of interest and investment in building autonomous UVS of increasing complexity and it does not seem an unreasonable prediction that within a decade we shall see a fully autonomous unmanned combat aerial vehicle (UCAV) in service. Moreover, as occurred with cruise missiles and GPS, where within a decade of their first use during Operation Desert Storm, they will likely then become vital components of military considerations for national defence forces. Consequently, as UVS become more sophisticated, unlike tethered or remotely operated vehicles, which are effectively extensions of their human operators, issues of liability will become more complex. Sometimes accidents will happen, but if UVS do not demonstrate sufficient and predictable capacity to obey the *rules of the road*,[80]matters of liability will be raised in the courts. It is perhaps instructive therefore to quote from [239]:

> "In 1936, a Duke University law student published an article summarizing the path of automobile liability law. He observed that in 1905 all of American automobile case law could be contained within a four-page law review article, but three decades later, a "comprehensive, detailed treatment [of automobile law] would call for an encyclopaedia." That law student was Richard M. Nixon, who would later become President of the United States. His conclusion was that courts were mechanically extending 'horse and buggy law' to this new mode of transportation in most doctrinal areas. However, some judges were creatively crafting new doctrine in certain subfields of automobile accident law by stretching the legal formulas at their command in order to reach desired results."

[80] The phrase 'rules of the road' is used as an expedient way to describe the legal framework within which a vehicle must comply when traveling from point-to-point, regardless of whether it is on or off-road, in the air or at sea (on or under the surface of the water), or operating within the Law of Armed Conflict.

Richard Nixon's observation appears to apply equally well to the age of unmanned vehicles as, at present, the use of UVS falls within a regulatory gap. That is, the technology appears to be under a loose legal framework, is self-regulated and/or only allowed to operate in restricted areas. There are many legal precedents connected to the use and production of UVS as purely commercial products and the related principles of responsibility and product liability are discussed in the relevant sections on Accountability and Liability. Before we do this, however, let us consider the legal status of UVS and their operation in the presence of other users of their environments.

6.1.1 Maritime Vehicles

As we have seen from earlier sections, UMVs have existed for many years, although their development and usage in the last 20 years has increased significantly. There is a recommended Code of Practice for UUVs, which was produced by the Society for Underwater Technology [60]. Despite the comprehensive nature of this document, it is not legally binding and the safety and legal framework for UMV use has not yet been formally adopted. As a result, the growth of the technology's use appears to be evolving under a loose legal framework such that the UMV community has had to be self-regulating. In this section, we present a condensed version of that contained in [60] [126] and [253] and readers are referred to these documents for more information.

There is a long history of International Maritime Law (also known as Admiralty Law) governing maritime issues and offences and the relationships between entities operating vessels on the oceans. It is distinguished from the Law of the Sea (also a body of public international law) that deals with navigational and mineral rights, jurisdiction over coastal waters and international law governing relationships between nations [148].

Although not endorsed or ratified by all nations, the 1982 United Nations' Convention on the Law of the Sea (UNCLOS) [274], which proscribes rules of navigation for vessels at sea, and the International Regulations for Avoiding Collisions at Sea (COLREGS) are the most widely used legal references in this regard. These conventions are usually enforced through local Coast Guards and courts. However, they were specifically written for guiding human behaviour and are not suitable for direct input to UVS control systems.

6.1.2 The Legal Status of UMVs

First, it is not clear that UMVs would be classified as "vessels" as the majority would fail the test against most formal definitions[81], either because they are used

[81] The term "vessel" is undefined in UNCLOS [271]. COLREGS defines it as "every description of watercraft, including non-displacement craft and sea planes, used or capable of being used as a means of transportation on water." The International Maritime Dictionary [148] describes a vessel as "a general term for all craft capable of floating on water and larger than a rowboat. The term vessel includes every description of watercraft or other artificial contrivance used or capable of being used as a means of transportation on water".

to explore the ocean environment for scientific or military purposes or due to their size and design or because they could not be used as a means of transportation (although the question may be posed, "transportation of what" [126]). Unfortunately, the clearest definition ("a human-made device, including submersible vessels, capable of travelling at sea") is non-binding and therefore has no legal status[82] [40].

Despite this, from a purely practical point of view, it is likely that UMVs will be considered some form of vessel and will be required to comply with international maritime law. Following the "rowboat rule" [148], however, it may be that small UMVs fall under one set of laws and larger ones under another. That is, there are legal precedents to the effect that small watercraft are considered to be under control from where they were launched rather than nearby support vessels (i.e. from where their operators were receiving instructions). UMVs may therefore achieve some vicarious status on the basis of their size or launch platform.

Another possibility is that UMVs might be considered "warships" (which need not be armed). According to UNCLOS, however, a warship is a ship "belonging to the armed services of a State bearing external markings distinguishing such ships of its nationality under the command of an officer duly commissioned by the government of the State and whose name appears in the appropriate service list or its equivalent, and manned by a crew which is under the regular armed forces discipline" [274]. UMVs, which do not have a commander and crew, would not seem to qualify. Those deployed and/or remotely operated from a support vessel might be considered to qualify under extensions of that ship, but would presumably enjoy the same level of sovereignty as the support vessel (i.e. they would be immune from seizure, etc[83]).

If considered vessels in their own right, but not warships, UMVs might be considered Auxiliaries, which are "vessels other than warships owned or operated by or under the exclusive control of the armed forces" [126]. Unfortunately, as previously mentioned, UMVs may not qualify as vessels so this is still not clear. Regardless, auxiliaries also enjoy sovereign immunity, which means that UMVs might still be protected from seizure by a foreign state.

Finally, UMVs may be considered weapons, particularly if they are weapons delivery platforms (i.e. a torpedo that carries torpedos). This has particular relevance to where and how such systems may be used in foreign waters [60].

[82] This definition is provided by the American Branch of the International Law Association (ABILA) Law of the Sea Committee (www.ambranch.org), a non-governmental association with UN consultative status.

[83] It is a well-established principle of international law that warships are an extension of their respective states and cannot be seized, boarded or searched without the permission of their commanding officer. They are, however, legitimate targets in a war. Somewhat perversely, if UMVs do not qualify as warships or vessels they may also not qualify as legitimate military targets under the LOAC, even though an armed enemy merchant vessel, a merchant vessel acting as an auxiliary, one conducting intelligence operations or directly belligerent acts on behalf of its armed forces also constitutes a legitimate target [126].

6.1.3 Operations in Ocean Areas

There are different classifications of ocean areas in which maritime vessels operate, each of which determines the degree of control that a coastal nation is able to exercise over foreign vessels operating in those areas. They typically include: Internal Waters, Territorial Waters and High Seas, although more recently have come to include Exclusive Economic Zones (EEZ) and Archipelagic Waters [274].

Users of the High Seas enjoy complete freedom of navigation. An EEZ extends 200nm from the low water mark of a country. Navigation in an EEZ is not restricted as long as foreign vessels do not interfere with resources. Consequently, as long as UMVs can navigate with respect to other users and pose no territorial threat to the coastal nation, they would seem to be able to operate with freedom in these zones.

Sailing through International Lanes and Archipelagic Zones requires vessels to "proceed without delay" via "continuous and expeditious transit" [126]. This may prevent a support vessel stopping to launch or recover a UMV and/or a UMV performing sweep searches. Nevertheless, the protocol does allow submarines (and presumably, therefore, UUVs) to transit submerged. Unfortunately, research and survey activities are prohibited so REA and ISR-related activities may not be possible without consent. That said, the protocol allows self-defence against suspected threats, so a warship would be entitled to deploy a UMV ahead of its path for the purposes of force protection, say for a UUV to undertake mine countermeasures or a USV to guard against a RHIB[84] terrorist threat.

Vessels wishing to navigate through Territorial Waters, which extend 12nm from a coastal nation's low water mark, will be subject to even more stringent regulations. Like those passing through International Lanes and Archipelagic Zones a vessel's passage must be "continuous and expeditious" [126]. However, its transit must also be for a specific purpose and must not be "prejudicial to peace, good order or the security of the coastal state" [272]. This includes "any exercise or practice with weapons; any act aimed at collecting information to prejudice the defence of the coastal state; the carrying out of research or survey activities; the launching, landing or taking onboard of aircraft; the launching, landing or taking onboard of any military device; or any other activity not having a direct bearing on passage" [272]. Additionally, as submarines must both travel on the surface in these waters and show their flag the same must be assumed for UUVs.

From a practical standpoint, the use of UMVs in foreign territorial waters could face considerable legal challenges, particularly if they carry weapons even though it is the conduct of a vessel that defines the nature of its passage, not its class.

6.1.4 Navigation Environment

Safety requirements refer to vessels operating on the surface of the ocean being required to carry internationally agreed lighting and signal equipment (e.g. a white masthead light, visible for up to three miles for inconspicuous or partly submerged

[84] RHIB – Rigid Hulled Inflatable Boat (i.e. a rubber dinghy).

vessels or multiple lights for vessels longer than 12m) [253]. The vessels are also required to carry equipment, which varies in accordance with the size of the vessel, so they can be heard and can respond in a particular fashion to allow safe passage. Presumably UMVs (including UUVs navigating on the surface) would be required to carry the same equipment and to operate it according to these regulations.

Strict interpretation of these procedures also implies that a vessel must maintain an "able lookout" at all times [274], although for UMVs it is not clear what the term lookout might mean. While in their experimental stages of development lookouts might be considered to be on the support vessels, but it is an open question as to whether the separation between the support vessel and the UMV constitutes able, particularly as the separation distance between them grows.

In addition to the basic capacity to navigate and avoid other vessels and obstacles, UMVs will also require knowledge of signs, rights of way, etc. and an understanding of the guidelines consistent with standard maritime operating conventions. For instance, there is a hierarchy of vessels at sea so that mariners can operate safely and understand their responsibilities in regard to 'giving way' to one another. From highest to lowest, the list is: not under command; restricted in ability to manoeuvre; engaged in fishing; sailing and underway; power-driven and underway; and, sea-plane underway [41].

It is not clear where in this hierarchy a UMV should sit. In one sense, declaring the UMV to be "not under command" seems appropriate, but unless there was some technical malfunction[85] this would not be accurate as the classification describes a vessel without command. An alternative would be to declare the UMV "restricted in its ability to manoeuvre" or "an obstacle" although given the level of sophistication likely to be onboard and the fact that it is motorised and autonomous, it seems more sensible to at least attempt to get it to follow the relevant regulations. In the final analysis, if a vessel is incapable of avoiding a collision it is considered "unseaworthy" and therefore negligent; and liability will be imposed in any collision. A similar set of considerations will be needed in regard to UMVs becoming entangled with nets. In the final analysis, the specific definitions and conveyance of risk and liability are likely to be defined by civil law arising through normal operations: conclusive determination will need case law.

Benjamin [39] makes reference to an interesting case in which a US warship (which was blacked out) collided with an Australian ship in harbour at night. The Australian ship was not blacked out and was displaying the requisite navigation lights. Even though the warship maintained it had right of way due to the "starboard rule"[86], because it also had the ability to see other ships without being seen, it was held to be at fault as it had the last opportunity to avoid a collision.

[85] It is not clear how the UMV would recognise any such malfunction and, to comply with regulations, then declare itself not under command.

[86] The COLREGS make it clear that vessels should not assume a "right of way" – it is either the "stand on" or "give way" vessel. In other words for two ships on a collision course, the ship on the left (the *give way* vessel) must give way to the ship on its right (the *stand on* vessel).

This may have bearing on the safe operation of UMVs, which may be deemed not to be able to ‘see’ other ships.

Environmental and regulatory regimes also exist to protect marine mammals from noise and harassment. For instance, one may not “harass, hunt, capture or kill or attempt to harass, hunt, capture or kill any marine mammal” [253]. It is thought highly unlikely that a UMV will collide with a marine mammal as the UMV moves slowly, but its operation may constitute harassment. However, it is not clear what operations, beyond those of manned vehicles, UMVs might undertake in this regard except that they may well be used in shallower waters.

Finally, certain operations will require that some UMVs carry weapons (UCMVs). Just as with UCAVs, then, procedures will need to be developed and legal arrangements made regarding emergency operations for UCMVs. For instance, in the event of engine problems, loss of a command and control signal, weapon malfunction and so on, the UCMV will need to follow precise and fail-safe procedures. These may include pre-planned trajectories or self-destruct points, but at present are not universally agreed or enforceable.

6.1.5 Air Vehicles

The foundations of international civil aviation and the related legislation were laid during the Chicago Convention on International Civil Aviation held in 1944 (which was ratified in 1947). This convention charged the International Civil Aviation Organization (ICAO), an agency of the United Nations, with coordinating and regulating international air travel. It also established rules of airspace, rules for aircraft registration and safety, and codified the principle that each nation maintains sovereignty over its airspace [136].

Article 8 of this convention states “no aircraft capable of being flown without a pilot shall be flown without a pilot over the territory of a contracting State without authorisation by that State and in accordance with the terms of such authorisation. Each contracting State undertakes to ensure that the flight of such aircraft without a pilot in regions open to civil aircraft shall be controlled as to obviate danger to civil aircraft” [136]. Article 3 of this convention recognises that “state aircraft[87]” are exempt from civil regulations; although the article also stipulates that national regulations for state aircraft must “have due regard for the safety and navigation of civil aircraft” [136].

Military UAVs[88] are considered state aircraft [161] regardless of whether or not they carry weapons. Those that are not state aircraft must abide by ICAO and national regulations. Beyond the technology of the air vehicle, therefore, even though their status may not necessitate them to abide by civil regulations, the fundamental problem for most UAVs is that their practical and safe operation

[87] The Chicago Convention does not define the term “state aircraft” but suggests that the term is determined functionally by the use of such aircraft by the military, customs services, or the police.

[88] Some nations actually rent some of their military UAVs (e.g. Australia and the US) from civilian corporations. As civilian aircraft these are managed in accordance with standard airspace regulations.

requires them to interact with, and hence to have due regard for, civilian manned aircraft.

While this clearly makes good common sense, from a legal perspective it means that, if they are to operate in uncontrolled airspace, the UAVs must effectively comply with the regulations approved by the national and regional administrative bodies for airspace and operations. Unfortunately, the precise meaning of the term "have due regard for"– and the commensurate civil regulations and procedures – still needs to be codified by most of the regional bodies and ICAO[89].

At present, most authorities restrict the use of UAVs to special airspace where civil and commercial aircraft access is strictly controlled. They also prevent them from flying over populated areas. UAV operations within uncontrolled airspace usually require significant notice to national air traffic administrators and near-perfect weather conditions. Military operations also usually manage the use of airspace for UAV operations on a temporally or spatially segregated basis.

The key starting point in regard to operations in uncontrolled airspace is that the UAVs pose no greater risk to persons or property on the ground or in the air than that presented by a manned aircraft.[90] That is, the UAV must provide an equivalent level of systems performance to that of a manned aircraft and be as reliable. These issues of performance and reliability need to be addressed from the perspective of the structural integrity of the UAV, the systems and sub-systems performance, the stability and control of the UAV and the procedural and regulatory regime. Additionally, most regional authorities recognise that the air traffic management procedures for UAVs should mirror those applicable to manned aircraft and the provision of air traffic services to UAVs should be transparent to air traffic controllers (i.e. UAVs must fit in with other airspace users rather than requiring the existing users to adjust to accommodate UAVs).

Fundamentally, there are three aspects of safety legislation that must be addressed [103]:

- *Airworthiness*: The design of the aircraft must be approved; the aircraft must be manufactured in accordance with this design; and, the aircraft must be maintained in accordance with appropriate maintenance and configuration control procedures.

[89] There are a number of ongoing initiatives and are taking place at the national and regional level (e.g. JAA/EUROCONTROL, ASTRAEA, NATO FINAS WG) and several that have concluded (e.g. ERAST and ACCESS 5).

[90] At present, UAV's cannot reliably detect other aircraft and conflict situations so they are unable to share air or runway space with them. To enable this, they will need a fault-tolerant, multi-function 'pilot' capable of operating day or night and in all-weather conditions, with the capacity to at least replicate a human's ability to sense and avoid problems. In other words, the UAV must be able to sense other aircraft in its operating environment, monitor the health of its component systems (e.g. sense loss or corruption of communications, sense structural, systems, and other onboard failures), and take appropriate action on the basis of the situation. Since not all aircraft carry transponders the UAVs will need to use onboard sensors to detect other aircraft, and ideally coordinate this information with other available information [103].

- *Flight rules:* The responsibilities and authority of the 'pilot' must be defined, as must operating rules for different classes of airspace, weather conditions, etc and any equipment that may be required onboard the aircraft.
- *Operator qualifications:* The licensing and training regimes for any pilots or crew need to be defined, together with any periodic activities required to maintain the currency of these qualifications.

Within any 'responsibilities and authority of the pilot' section of the legislation we must also address the degree of autonomy that we consider acceptable for any command and control procedures that might involve a UAV. In other words, we must establish procedural approaches for autonomously, semi-autonomously or directly controlling a UAV in any given phase of its mission or section of airspace. At present, many national bodies require that UAVs operating outside special restricted zones have either certified pilots at the controls or that the UAVs be commanded and controlled in a particular way (e.g. under semi-autonomous or direct ground control for take-off).

Another component of the command and control procedures concerns the use of UCAVs. Like UCMVs, for at least half of each mission UCAVs will carry weapons, which significantly increases the potential threat to other airspace users as well as those on the ground. Consequently, specific procedures and legal arrangements will need to be developed regarding emergency and command and control procedures for UCAVs [161].

Small UAVs – comparable in size to model aircraft – seem to be regarded by most authorities as equating to model aircraft and unlikely to require integration with most civil airspace users. Indeed, similar constraints to those imposed on the operators of model aircraft are often applicable to small UAVs (e.g. height, line-of-sight, proximity to airports, etc). In effect, therefore, while there are many ways of categorising UAVs (e.g. weight, height, endurance, role, type, etc), and each is valid in context, from the perspective of operating in uncontrolled airspace the discriminator seems to be flight rules, as these are the most relevant to UAVs and manned aircraft. In other words, although it may be (say) kinetic energy, size, or endurance that defines whether a UAV is regulated in a particular way, from a legal perspective the categories will probably be defined on the basis of the applicable flight rules because they govern the regulations [103].

The implications of insisting on an equivalent level of safety for UAV airworthiness to that of manned aircraft is that civil certificates of airworthiness (or their military equivalents) may be required if the UAVs are to fly in uncontrolled airspace. Moreover, the UAV would need to be manufactured in appropriately certified facilities to agreed standards using approved materials and subject to rigorous inspection, maintenance and flight test processes [97]. Ground control, launch and recovery mechanisms, which are critical to UAV operations, would also need to be included in the above procedures.

Similarly, stringent security standards may need to be developed to protect the UAV command and control links so that they are resistant to jammers and spoofers. As many UAVs will fly beyond the line of sight of their ground stations, these communications will need to accommodate satellite or other 'multi-hop' technologies.

The result is that many UAV facilities may also need to invest in an appropriate level of physical security and the cost of many UAVs will increase and quite possibly force many of today's manufacturers out of the market; or make UAVs less cost-effective options to their manned counterparts.

6.1.6 Ground Vehicles

Even though the technology was immature, unmanned ground vehicles have been around since the late 1930's, 1940's, or possibly early 1970's, depending on your point of view. More recently, and in particular since the 1990's, there have been a series of successful UGV research programs and three DARPA Grand Challenges, which have fast-tracked the rate at which the technology has developed.

However, unlike maritime and air vehicles there are not a set of internationally agreed conventions or principles governing the interaction of ground vehicles with one another and/or other users of their environment. There are laws and conventions that transcend single states (e.g. Europe and the United States), but to all intents and purposes the principles by which ground vehicles are governed are developed and executed at a national or state level. As a result, Australian law is used here and readers can extrapolate for their own countries.

At present, under Australian law [120], the definition of a vehicle is very broad and not exhaustive. It includes, for example, motorised wheelchairs that can travel over 10km/hour, but does not include trains, wheeled recreational devices or wheeled toys. Riders of these are treated as pedestrians. Consequently, like UMVs, the legal status of UGVs must first be established. Moreover, the status and obligations of UGVs may vary according to their function and size, just as they vary for commercial and domestic vehicles, vehicles carrying passengers, vehicles of a certain size, and so on.

It is probably reasonable to assume that an autonomous UGV will be classified as a vehicle, although unlike tele-operated UGVs, which are really just an extension of their operator, autonomous vehicles may be considered separate legal entities. Moreover, most laws applying to the movement of vehicles apply to drivers or riders,[91] as they are assumed to be in control of the vehicle. Consequently, attribution of control, and hence responsibility for any breach of the relevant legislation, is harder to ascertain unless we ascribe these laws to the IDT.

To this end, it is probable that we will need to establish the specific role of either the functional components of the UGV that result in its mobility or at a more holistic or system level, both of which may have far-reaching consequences. We will discuss these later in the section on UVS and Tort Law. However, an alternative to ascribing responsibility for the vehicle's operation might be to declare that the vehicle is 'not under control' in the manner that we might classify a run-away vehicle. In some sense this seems appropriate because there is no human operator 'in charge' of the vehicle. However, from a practical perspective,

[91] The distinction being when the vehicle is horse-drawn or a motorcycle, etc.

the run-away status would seem to imply that the vehicle is temporarily out of control, which is clearly inaccurate.

For UGV's the 'rules of the road' will not be the only laws that require modification. For instance, what if an autonomous vehicle was to trespass and/or cause property damage? Additionally, as autonomous UVS must have sensors for navigation purposes there may well be issues of privacy raised, just as with Google Maps and Street View [125]. Clearly, the existing law of tort would probably be used, but against who; the owner, the user, the developer/integrator, the component manufacturer? Clearly, as is compulsory for commercial and domestic ground vehicles, insurance might be purchased or insisted upon so that liability may be assumed vicariously by underwriters. However, this really just shifts and quarantines the liability it does not introduce novelty into the problem with respect to solving it.

6.2 UVS and Tort Law

Regardless of whether or not UVS carry weapons they have many safety implications as they are typically large objects that travel at speed and are constructed of sub-systems and materials that are inherently hazardous. Furthermore, most of these cannot be avoided as they are needed for system operation. For a weaponised UVS the list includes:

- Inadvertent firing of weapons;
- Engagement of unintended targets;
- System fratricide from weapon firing;
- Unintended system operation injures personnel;
- Exposure of personnel to hazardous chemicals;
- Exposure of personnel to damaging levels of radiation;
- Exposure of personnel to fatal electrical voltage;
- Damage to the environment; and,
- Collision with other people, vehicles or property.

Clearly, if a UVS is involved in an accident, the issue of how it was used is of interest. However, it may be extremely difficult to say whether this was inappropriate as often the best that can be concluded is that the operator used the UVS or its HMI in a certain way and certain consequences followed. This may be particularly true for complex environments, where the only data gathered and usable in regard to identifying the proximate cause may come from the UVS itself, which may of course be purposefully designed not to allow information to be disclosed that may 'incriminate' the developers, programmers, etc,[92] or admit liability on behalf of any insurance underwriter.

[92] If, as is suggested in later sections, UVS are provided with separate legal identity then intentionally not providing "self-incriminating" evidence is considered legally sound practice in many jurisdictions.

The key tort[93] law concept is *negligence*, which implies that the developer, owner, or user failed to do something that was required. Legal culpability for negligence depends upon a *failure to warn* or a *failure to take proper care* [24]. A *failure to warn* occurs when (say) the manufacturer was knowingly aware of a risk or danger but failed to notify customers of this risk. *Failure to take proper care* is more difficult to prove because it usually involves cases where the risk or danger cannot be shown to have been known to the manufacturer but, it is argued, was in some sense obvious or easily foreseeable, even if the manufacturer failed to recognise it.

The usual legal defence against charges of *failure to warn* or *failure to take proper care* is that the plaintiff acted in accordance with the 'industry standard' [17]. In other words, (say) the manufacturer followed 'world's best practice', as measured against their peers, regardless of whether these standards are explicitly stated or are simply implied. Clearly, a first step for militaries around the world would be to establish a set of standards and principles for the development, operation and force-integration of their UVS, although these may need to be characterised in terms of parameters that are fiducially referenced. At present, however, most UVS are simply demonstrated in some representative military environment.

Another aspect of legal liability is that it can be differentially apportioned, even for a single event: that is one party might be 10% responsible, the other 90% responsible and analysis of causal links in product liability cases is common and will be an important consideration when considering UVS-related mishaps. For instance, a badly designed object recognition algorithm may be responsible for the prosecution of the wrong target, but so might a weak battery, a bad sensor, or a malfunctioning actuator [68]. Moreover, the environment in which the UVS is used or the training regime may also be the principal or contributing factors to the failure.

Fortunately, the law already has a highly developed set of principles and precedents that apply to product liability, so we can apply them to the use and production of UVS as if they were unexceptional technological objects. Moreover, it is believed that many of the concerns frequently touted as being possible dangers that UVS might pose will ultimately fall under this interpretation [24].

Furthermore, as indicated previously, insurance might be purchased for the UVS such that liability is assumed vicariously by an underwriter against a range of responsibilities and duties (e.g. the duty to exercise due care when operating in the presence of pedestrians or other vehicles). In fact, as is often the case, the technology may operate more correctly and provide greater reliability than its human counterpart and it may turn out that insurance for unmanned systems is less than that for the manned systems. The idea of insuring artificial agents is usually

[93] Tort (or civil) law deals primarily with property rights and infringements and seeks justice by compelling wrong-doers to compensate those who were harmed for their loss The criminal law, which in our case includes the Law Of Armed Conflict (LOAC) seeks justice by punishing the wrong-doers. We will consider the LOAC and criminal law implications later.

attributed to Solum [255] and one of the principal objections used against it is the lack of any novelty that insurers provide. In other words, in the final analysis, someone has to pay the insurance premiums and it will probably be the owners, users, manufacturers, or programmers, who are ultimately therefore bearing all the risk anyway, albeit amortised over time.

An important principle in compensation is that of causation analysis [145], which seeks to attempt to fix liability on those responsible for the legal cause of an injury (the drunk driver was the cause of the car accident, not the brewery that made the beer). This is a necessary element of any civil lawsuit, although a number of other elements are also necessary (e.g. duty to the injured party and damage caused by the failure to execute that duty). Moreover, it is frequently difficult to pick the single (or few) legally responsible causes from the variety of possible ones. For example, most accidents result from multiple causes and it can be difficult to untangle the various contributing factors (a UVS that crashes into a house may be the fault of the manufacturers, the developers of the IDT that suffered a ‘glitch’, the ‘pilot’, the mission planners who allowed it to pass by or over the house, the air traffic controller (in the case of a UAV) who failed to notify the pilot of the thunder clouds, and so on).

Consequently, before any compensation could be paid a court would have to decide out of the many possible factors that were necessary for the accident to have occurred, which one(s) had actually caused the damage. To do this it would first seek Causation in Fact, which connects the injury suffered, through a chain of circumstances, to the act of an accused, no matter how many other contributing factors may be present [24]. This enquiry is a fundamental and intractable element of proof of a tort and is roughly equivalent to establishing the ‘but for …’ causes of the accident. Next, in order to establish liability, Proximate Cause is determined, which selects from all of the causes in fact the entity (or entities) which will be held responsible for the injury.

Given the distributed, complex and polymorphic nature of the IDT likely to be responsible for controlling the UVS, to say nothing of possible interaction with its human supervisor, the courts and civil parties may find it extremely difficult to specify the proximate cause of the damage. In fact, they may not be able to even identify the UVS as the cause. Moreover, if we assume that automation is sophisticated (i.e. it is as competent at driving as a human) we are presented with a series of other questions. When should a human elect to use the IDT’s judgement and when his own? Should these decisions be based on a formal framework of metrics, the relative performances of users and the IDT, or less rational factors? What would the consequences be of inappropriately selecting one or other of these options if the same choice had previously resulted in a beneficial outcome? What might any mitigating circumstances be?

One solution might simply be to identify an errant behavioural trait or the presence of a fault within the UVS using a truncated form of proximate cause. That is, not take the detailed causal relationship into account, but accord culpability to the UVS at the system or possibly sub-system level. Although this may seem unreasonable (and legally questionable), economic pressures, driven through either consumer choice or insurance policies may well determine the cost and hence the viability of key UVS components, thus providing a ‘market forces’ solution to any ‘fairness’ test.[94] It is understood, however, that as this solution may hold ‘innocents’ liable for actions for which they were not responsible, albeit vicariously through an insurance premium and this may not be politically or socially acceptable.

However, given the typically adversarial nature of law; that no general rules exist on how to identify the causes of UVS accidents and infringements; that in the abstract the law does not provide good answers to questions; that accidents will not usually be a function of the human operator, but of the system inaccurately or ineffectively facilitating user understanding of system functions; and that human error is not removed by simply increasing the level of automation, it may be that ascribing liability in this way is preferable to the alternatives.

The most practical way for this to become a reality is for UVS or their IDT to assume some sort of separate legal personality. While it is recognised that some people will dismiss the notion of a machine holding a separate legal identity as fantastical and not something deserving of serious consideration, this concept does not imply that the UVS is sentient: they are real-world machines controlled by algorithmic processes and are tools, not people. Furthermore, the practical instantiation of UVS holding separate legal status is unlikely to differ substantially from the vicarious liability assumed by an underwriter.

Before discussing the prospect of a silicon-based architecture rather than a carbon-based one having a legal personality, however, issues of accountability and liability relative to the Laws of Armed Conflict (LOAC) are discussed. In this domain inappropriate actions may be criminal, any redress may need to be non-monetary, and insurance is more than likely unsuitable.

6.3 UVS and the Law of Armed Conflict

To date the IDT onboard a UVS has largely been used for navigation purposes. However, it can also be used for the autonomous identification and engagement of targets. This means we can now provide UVS with the potential to calculate which objects within a sensor’s field of view constitute targets, whether or not to engage them, and by which route or means they should be engaged. If the payload includes a weapon, then a UVS has the potential to make the final determination on whether to prosecute a target with lethal force, without operator intervention. If permitted, this would mark a sea-change in the role of technology in warfare as the human could be removed from the decision-making loop, with potentially lethal consequences. As soon as such weapons are justified by one nation, others will likely follow. The use, development or acquisition of such weapons may then

[94] In this regard, it is also assumed that there would be threshold tests relating to health and safety applied to avoid a technological ‘free-for-all:’.

trigger our obligations under Article 36 of Additional Protocol 1 to the Geneva Conventions.[95]

Even though a significant number of influential writers [47] [56] [92] [159] [195] have argued that future IDT will possess intelligence equal to – and probably in excess of – humans, particularly in stressful and high workload environments, at present most would agree that there are advantages of having real 'eyes on the target'. Moreover, many commentators have suggested that "Humans will always make the final decision" [246] [256] [297] as, under the Laws of Armed Conflict (LOAC), there is a requirement to assess ambiguous situations with rational human judgement such that the response is not excessive to the anticipated military advantage. For example, in the words of Air Chief Marshall, Sir Brian Burridge [62]:

> "When we go into combat, we have got to be sure what we are doing is legal and moral. I do not believe that, in future, even though technology will allow it, we will be allowed to indulge in robotic warfare. I simply do not see the international community regarding that as an appropriate way to fight. The notion of using UCAVs controlled ten time zones away to prosecute a battle is not something international law of the future will regard as acceptable. I think the notion of a person in the loop, the notion of positive ID, the notion of someone feeling the texture of what is going on in the battlespace, is going to be more and more prevalent. … Overall, I think robotic warfare drives you away from what I term as emotional connectivity with the battlespace. My view is that winning the hearts and minds battle with the indigenous population requires this emotional connectivity."

Unfortunately, retaining humans in this way is likely to provide its own problems as the issue is not so much "Will there always be a human in the loop?" but "Where will the human be in the loop?" Will it be at the level of each and every target engagement, or at the level "use of lethal force in this mission is granted?" Moreover, there is only modest evidence to suggest that autonomous or UVS-based weapons will be considered any differently from other weapons systems.

For example, IDT that permit lethal response options may be introduced as systems that act if human input/veto has not occurred within a certain period. As IDT improves and is introduced onto the battlefield in greater abundance, critical survival decisions will start to depend upon them. As a result, operational tempo

[95] Protocol Additional to the Geneva Conventions 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol 1) - adopted on 8 June 1977 by the Diplomatic Conference on the Reaffirmation and Development of International Humanitarian Law applicable in Armed Conflicts. **Article 36, New Weapons:** In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party [217].

may well be diminished if the systems rely upon human intervention, particularly for complex or networked environments. Similarly, as information is increasingly processed and manipulated by these systems prior to decisions being made or confirmed by humans the link between *real* decision-making and automated response will become blurred.

Furthermore, even though active vision and intelligent sensing techniques are not yet as sophisticated at interpreting a situation as humans, automation is almost universally welcomed by operators because, if introduced correctly, it allows reliability and the overall system effectiveness to be improved, allowing humans to concentrate on other tasks. For instance, an automated vision system has the advantage of being able to 'stare' equally effectively in all areas of the sensor's field of view. The human fovea is only about 0.5 deg, degrading to one tenth of the visual acuity at 2 deg [28]. This provides the human with a remarkably narrow angular focal view. Moreover, the human suffers from fatigue and is unable to focus on (or even detect) multiple targets within a scene. An automated system can alert the user to the presence of multiple potential targets and overlay them on the video – perhaps in a head up display to the operator. The user can then use his superior interpretative skills to verify or dismiss the targets.

Prior to the development of technologies that are able to meet the implied challenge of full autonomy, therefore, it is likely that users will retain responsibility for high-level tasks such as target designation, target verification, and the lethal application of force. Similarly, the intelligence onboard the UVS is likely to assist in the process by identifying and tracking potential targets and controlling and aiming the weapon system. Maintenance of broader situational awareness will probably be a shared and less controversial responsibility. Moreover, as the human reaction time is around 400ms, whereas a machine can respond in a few nanoseconds, the first use of fully autonomous UVS weapons will probably be in defensive roles, where critical decisions of survivability must be made quickly, or in the prosecution of fleeting targets, where the timeline is also necessarily short. Hence, it is very likely that the target verification processes will be compressed and highly dependent upon IDT, perhaps to the point where the only human input is oversight or veto.

Unfortunately, in order to achieve the appropriate level of human oversight, we will still be reliant upon technology for the users and officers superior to those engaged in combat operations to remotely watch (and potentially interfere with) a tactical engagement. This raises a number of issues:

- There is the question as to whether the senior commanders are legally obligated to interfere with ongoing operations if they realise that a violation of LOAC is about to occur. Schmitt [246] states that application of Article 28 of the Statute of the International Criminal Court indicates that senior commanders will progressively become more responsible for their subordinates' conduct.
- A potential 'hiatus' in tempo may arise when those on the frontlines start to rely on the intervention of their superiors rather than using their own

discretion before proceeding with a particular course of action. It seems likely that, to avoid such interruptions in tempo, there will be pressure for IDT to be introduced in support of the decision-making process, for instance through the filtering or fusing of information, the generation of response option, plans, schedules, and even the control of weapons.

- It is an open question as to whether senior commanders have access to technology that allows them to observe the relevant aspects of a remote operation to the degree necessary for them to usefully intervene. Given the amount of potential information that such a commander is likely to have to deal with (and require for any lethal target prosecution), and the compressed timeline against which such decisions are likely to need to be made (otherwise he would be using a more conventional weapon), it would seem probable that the only way to meaningfully monitor and respond to any such workload would be through the use of IDT.
- Just as corporations strive to exploit research into human factors and human systems integration to develop interfaces and artificial agents that garner trust in their users, so manufacturers of IDT and UVS will attempt to do the same. In other words, IDT will progressively undertake the difficult work (data fusion, visualisation, etc) so that the decisions made by humans become simpler. The net effect of this ‘sleight of hand’ will be to diminish the possibility of users making truly informed choices.
- Once this occurs, it is a small step to allow UVS to make decisions relating to the use of lethal force may in response to particular defensive circumstances. Clearly, the human commander has made the decision to use the UVS in the first place, which might be considered the crucial discretionary judgement. However, if the circumstances become more complex than initially anticipated or the UVS acts autonomously for some other reason (it perceives an attack to be taking place) there may be an ambiguity in regard to who actually made the decision to engage.

This effectively brings us back where we started, which may mean that the decision on whether or when to use lethal force is handed to IDT sooner rather than later.

6.3.1 The Law of Armed Conflict

The Law of Armed Conflict (LOAC) governs the protection of non-combatants and aims to limit the effect of war on those not directly involved. It sets out a legal framework for commanders in the field and defines the method and means by which warfare can legitimately be employed by state actors. The body of law that underpins the LOAC has been extensively codified in a series of international treaties dating from the 1860’s to the present day. The most significant of these treaties are the 1907 Hague series of treaties, which governs the conduct of operations, and the 1949 Geneva Conventions [217], which governs the protection of people and property.

To some extent, the two branches have now merged under Additional Protocol 1 [217], which was added in to the Geneva Conventions in 1977. As signatories to the Geneva Conventions and its Additional Protocols, countries are obliged to honour their obligations in good faith. Furthermore, many countries also incorporate the provisions of these treaties into their domestic law.[96] Clearly, countries must understand their obligations under these laws to ensure that they do not develop, acquire, or deploy systems that contravene the spirit or the letter of their responsibilities.

The theoretic framework for debate about the morality of specific choices and actions in war is usually attributed to Walzer [288] who established a set of principles that effectively capture general moral principles. The work draws a key distinction between just reasons for going to war (*jus ad bellum*) and just acts in fighting war (*jus in bello*). In this work we deal almost exclusively with the latter, although brief reference to the potential impact of UVS on the principle of *jus ad bellum* is made in the next section. A brief description of Walzer's principles, drawn from [29] now follows.

The **Principle of Military Necessity** states that "a combatant is justified in using those measures, not forbidden by international law, which are indispensable for securing complete submission of an enemy at the soonest moment" [288]. Military necessity requires combat forces to engage in only those acts necessary to accomplish a legitimate military objective. It permits the killing of enemy combatants and other persons whose death is unavoidable and it permits the destruction of property if that destruction is imperatively demanded by the necessities of war. However, destruction of property as an end in itself is a violation of international law as there must be a reasonable connection between the destruction of property and the overcoming of enemy forces.

The **Principle of Humanity**[97] forbids the "use of means or methods of warfare which are calculated to cause suffering which is excessive to the circumstances" [29]. It has also been expressed as averting the infliction of suffering, injury or destruction not actually necessary for the accomplishment of legitimate military objectives.

The **Principle of Proportionality** provides a link between the principles of Military Necessity and Humanity. In simple terms, the principle generally relates to the reduction of incidental injuries caused by military operations and requires that the losses and damage resulting from military action should be proportionate (i.e. not be excessive) in relation to the anticipated military advantage. The proportionality principle, together with the principle of humanity, dictates that "civilians should not be made the object of attack and that while civilian casualties may be an inevitable consequence of an attack, every effort must be made to spare them, and other parties who are non-combatants, from becoming adversely affected" [288]. The principle of proportionality not only requires that an attacker must assess what feasible precautions must be taken to minimize incidental loss,

[96] In the case of Australia these are the Geneva Conventions Acts of 1957 and 1991 [29].

[97] This principle is codified under Article 35(2) of AP-I: It is prohibited to employ weapons, projectiles and material and methods of warfare of a nature to cause superfluous injury or unnecessary suffering.

but must also make a comparison between different methods or axes of attack so as to be able to choose the least excessively destructive method or axis compatible with military success. When making that assessment, the attacker should "naturally take into account likely friendly casualties" [29].

The related **Principle of Distinction**[98] seeks to ensure that only legitimate military objects are attacked. This principle has two components: the first, relating to personnel, seeks to maintain the distinction between combatants and non-combatants or civilian and military personnel; the second distinguishes between legitimate military targets and civilian objects. Military operations must only be conducted against military objectives, including combatants. Non-combatants and civilian objects are protected from attack, that is, they are not legitimate objects of attack. LOAC therefore requires that "belligerents maintain the clear distinction between armed forces and civilians taking no direct part in hostilities; that is, between combatants and non-combatants, and between objects that might legitimately be attacked and those protected from attack" [29].

Together these principles seek to temper the violence and range of war. The **Principle of Responsibility** seeks to tie in the actions of warfighters to morality more generally, by seeking to ensure that the agents of war are held accountable for their actions. For instance, when soldiers turn their weapons against non-combatants, or pursue their enemy beyond what is reasonable, they are no longer committing legitimate acts of war but acts of murder. Other issues that arise from this principle include the morality of obeying orders that are known to be immoral and the status of ignorance (i.e. not knowing the effects of one's actions). In other words, there is at least an aspiration to identify those responsible for deaths in war, even if practical circumstances do not always permit this.

While there does not appear to be a universally agreed definition of "armed conflict" (the point at which LOAC applies) and [246] argues that, "given the means and methods of modern warfare it is no longer sufficient to use a threshold test of actor-based physical confrontation," (almost) all nation-states now accept the basic Principles underpinning the LOAC and that we are legally bound to consider these principles together.

There is not a uniform interpretation of all aspects of LOAC.[99] Moreover, some principles are only codified in Additional Protocol I, which has not yet been ratified by a number of major nations (notably the US) [297]. Consequently, in order to deploy a system that relies upon an IDT designed and/or manufactured by another nation, a country which has ratified the protocol (e.g. Australia) may have to understand the manner in which the vendor nation interprets the LOAC and whether this accords with its own interpretation. This is a further complication and

[98] This principle is codified under Article 48 of AP-I and is complemented by a number of other provisions within the Protocol that relate to the consequences of distinction, including Article 51(2), which prohibits attacks on civilians and Article 52, which defines military objectives.

[99] For instance, a civilian loses the protection of the LOAC "If, and for such time as, he takes a direct part in hostilities" (Article 51 (3) Additional Protocol I). The interpretation of the phrase "Takes a direct part in hostilities" is controversial, with some nations employing a higher threshold test than others.

of particular significance to countries that acquire significant military capability from nations such as the US, as the Statute of the International Criminal Court (ICC), which (for example) Australia has signed and (for example) the US has not, provides that the responsibility for providing the legal means and methods of warfare falls upon the military using the technology [297]. In other words, if the IDT violates the LOAC the forces of the purchasing nation may be held criminally liable not the vendors of the IDT or their national government.

The key message from Principles of Necessity, Humanity and Proportionality is that any weapon developed, acquired, or deployed must not employ any degree of force that is not necessary for the purposes of war. Somewhat obviously, therefore, we may not employ an unlawful weapon on a UVS and abrogate our responsibilities in this regard by implying that the UVS 'made the decision' to use the weapon. Similarly, the Principle of Distinction requires that any deployed system must:

- Distinguish between civilians and combatants;
- Distinguish between civilian and military objectives;
- Direct operations only against military objectives; and,
- Not cause excessive incidental loss to civilians.

It remains an open question as to whether or not IDT will ever be able to successfully identify legitimate military targets all of the time; it challenges humans. Nevertheless, given the rate of technological progression and the level of investment made by militaries around the world into research in this area it seems inevitable that IDT will progress to the point where they could be put to work on aiding and/or making decisions about what objects constitute targets and by what means and manner these targets can be engaged. To this end, let us examine whether UVS might ever be sufficiently competent to undertake such duties.

In order to be of use the division of roles between the human and the UVS will primarily be driven by a combination of human and technological factors, including the cognitive load on any human operator, the degree of trustworthy "intelligence" and automation achievable within the UVS and its weapon, and the potential impact of latencies on the control and actuation of any weapon systems. Based on our deductions from the Principle of Distinction above, these may be expressed as a series of technical requirements for the UVS. It must:

- Positively and accurately identify objects of interest and intended targets;
- Determine and minimise any effects on non-combatants and infrastructure;
- Accurately control the delivery of the weapon wrt to aiming at any targets;
- Maintain a suitable record of action for the purposes of evidentiary hearings;
- Contain no known bugs or malfunctions that might cause 'reckless' behaviour.

In addition to the basic principles of LOAC, which determines the actions that are lawful and therefore permitted, governments then place further limitations on military forces in respect to the application of force (for operational, political, diplomatic, or legal reasons). These conditions are known as Rules of Engagement (ROE). ROE provide authoritative guidance for the application of force in operations and take two forms: actions a soldier may take without consulting a higher authority, unless explicitly forbidden, and actions that may only be taken if explicitly ordered by a higher authority [29].

To comply with any ROE the UVS must also accept additional restrictions in regard to the circumstances under which it may use discretionary lethal force. For instance, it may be necessary to vary the type of targets allowed, the nature or trajectory of a UVS mission, the degree of hostility used to trigger a response and the nature of any response that might be permitted to achieve the desired goal (e.g. a warning shot may be required). Alternatively, we must provide our commanders with the requisite level of understanding of IDT capabilities relative to these functions so that they may take the appropriate precautions prior to and during the execution of their duties.

If for the moment we set aside the complexities of interpreting specific ROE, with the exception of "positively and accurately identifying objects of interest and targets,"[100] technology is currently available to undertake each of the other functions. Consequently, it is not difficult to imagine a system that combines these functions in order to automate the tasks currently undertaken by a human (possibly in conjunction with an automated system), particularly for simple scenarios where complex judgements and ROE do not need to be taken into account (i.e. scenarios that give a UVS very little discretion). To this end, let us now try to imagine how the introduction of such a capability might evolve.

***Stage One* -** A decision-aid is developed that assists the warfighter, allowing him to manage a larger number of tasks or targets. Such a program might allow the human to make goal-setting decisions with respect to UVS missions and navigation, high-level decisions regarding specific target designation and verification and compliance with the ROE. Other tasks, such as target acquisition and tracking, weapon aiming and control, record keeping, HUMS and weapons impact calculations might also be undertaken autonomously. In other words, the UVS would effectively act as humans do today, but would not take the key decisions on target discrimination, designation and verification – humans would.

***Stage Two* -** A UVS is developed that outperforms a human: ROE are then issued to the effect that – aside from exceptional circumstances – warfighters must follow the advice of the IDT. Perhaps experience has shown that under duress or in high workload environments humans make more targeting errors. Regardless, there is now little or no reason for the humans to check the program for compliance. As a result, the role of the human in the decision-making process diminishes and the number of UVS and targets and tasks he can manage become considerable; he

[100] And under certain circumstances this function can also be accomplished using current technology.

devotes little or no time to any particular target. There are still times when a human is called upon to make a decision. If such events occur with regularity, however, the developers of the system will be tasked with developing a routine for handling these. As the capabilities of such a system grow, the need for the human to make decisions diminishes.

Stage Three **-** This begins when the decision is made to do away with the human altogether; perhaps the human input/over-ride causes delays relative to an adversary's capabilities and these have a tendency to result in casualties.

The Principle of Responsibility then requires us to determine who is accountable for any erroneous or 'criminal' actions. This will be the focus of our discussion from here on. First, however, let us consider the potential impact that UVS might have on *jus ad bello* (just reasons for going to war) as UVS might also influence a nation's entry into war.

6.3.2 UVS and Jus ad Bellum

There a few ways in which UVS might affect *jus ad bello* [23][25][250]:

- They could directly threaten the sovereignty of a nation;
- They may introduce ambiguity into who made the decision to attack;
- They might make it easier for leaders who wish to start a war to actually start one;
- They may alter perceptions and ultimately make it harder to win a war;
- They may be used to deliberately carry out war crimes.

Challenge to Sovereignty – Autonomous UVS operating near political hotspots offer considerable potential for starting wars accidentally – or by more nefarious means (e.g. as a result of human manipulation). For instance, a UCAV on a surveillance mission might respond lethally to an Early Warning radar that is legitimately (and passively) protecting its host nation's airspace. Although not strictly a threat to either nation's sovereignty, this may be interpreted as an 'act of war' that leads to a more extensive conflict. Clearly, wars that start as a result of an accident do not fit the principle of *jus ad bellum* [25].

Ambiguity in Attack Decisions – Autonomous UVS with the capacity to make decisions relating to the use of lethal force may have been granted authority to respond to particular circumstances (see the above scenario). Clearly, the human commander has made the decision to use the UVS in the first place, which might be considered the crucial discretionary judgement, and which implies that all subsequent discretionary decisions derive from the principal authority. However, if the circumstances are more complex than initially anticipated or the UVS acts autonomously for some other reason (accidents are covered above) there may be an ambiguity in regard to who made the decision to engage in the first acts of war.

Once again, the commencement of such wars would not fit the principle of *jus ad bellum*, particularly if no formal declaration has been made.

War Crimes – Because the autonomous UVS cannot themselves be held accountable for their actions, some nations or commanders may elect to use them to deliberately conduct activities that constitute war crimes, thereby potentially absolving it/themself of any responsibility.

Lowering the Barrier to War – A central element in the decision-making process for most democratic nations is the estimation of the cost of a war in terms of lives lost, even if these are the lives of soldiers who have volunteered [25]. Consequently, in these nations a political strategy has evolved to limit the number of casualties such that, where possible, military involvement is kept to a 'safe' form. Moreover, significant investment is now made in a range of technologies that lower the risk to and increase the lethality of warfighters (e.g. armour, longer range weapons). This sometimes translates into civilian casualties. To this end, it is clear that the introduction of technologies such as UVS, which are specifically aimed at extending the reach of the warfighter to provide him with greater lethality and more protection, could lower the political barriers of entry into war by offering to 'take the blood out of war'[101] [252].

***Perceptions* –** Autonomous UVS may be perceived as providing those militaries using them as having an 'unfair' advantage and, while not illegal under the LOAC, this perception might be used against (presumably Western) military intervention or even to incite resistance among the local population. Moreover, the media is a powerful weapon in today's warfare and if UVS were to kill innocents this might lead to widespread revulsion. Finally, as coalition forces have discovered recently in the Middle East, there is no substitute for soldiers on the ground that have the capacity to interact with the local civilian population. Ultimately, therefore, wars waged with autonomous UVS may make it harder to win and establish a lasting peace.[102]

Another point made eloquently by [252] is that the vicarious engagement of warfighters in combat through UVS controlled from thousands of miles away effectively combines the previous two points. The net effect on our potential adversaries is to make them believe that we are cowards, thereby raising the likelihood that they will engage us in or attempt to prolong war.

A final note on this score; if we assume that only nations that fight just wars were to develop UVS technologies none of the above would be a concern. Unfortunately, history shows us that "all wars involve at least one unjust (or badly mistaken) nation" [25] so the prospect that UVS will be used to enable such future injustices is a legitimate concern.

[101] Naturally, this is not meant to imply that we should increase the risk to our forces and reduce their lethality in order to improve the Just War considerations. The point simply notes that this is one potential impact of UVS on *jus ad bellum*.

[102] This is perhaps more pertinent to *jus post bellum* (justice after war).

6.3.3 Autonomous Weapons

A full account of what is meant by autonomy and whether UVS can actually possess it requires answers to questions that philosophers have wrestled with for centuries. Regardless of its level of autonomy, however, to some greater or lesser extent a UVS is reliant upon a combination of external human-supervisory commands and onboard sensors and processing to capture, represent, and interpret environmental cues that are then autonomously combined and manipulated (and possibly presented to the user) such that the result is a series of mobility and payload response commands representing mission priorities.

In current parlance, the phrase "autonomous weapon" typically means that the weapon is a 'fire-and-forget' system capable of acting independently of human control, but only to the extent that it can determine mid-course and final trajectories and acquire targets on the basis of certain limited stimuli. There are several such weapons in existence today, some of which use of over-the-horizon and beyond-visual-range systems that include Automatic Target Recognition (ATR) capabilities that allow the weapons to select targets when they enter a designated area of interest. These features preclude full human control.

However, these weapons are really only semi-autonomous, in that they wait for their programming to be satisfied. That is, the determination of location, value and risks are undertaken by humans who ultimately control the weapon. For instance, some weapons must have their target sets pre-programmed before they are launched, although they can be re-programmed in-flight if real time intelligence indicates that the target has already been destroyed or if the situation has changed (e.g. civilian traffic has entered the area). Similarly, HARPY[103] can loiter for several hours before detecting, locking onto, and then destroying enemy radars.

Moreover, the SGR-A1[104] can detect and identify targets within a 4km radius (or 2km at night) and provide either a lethal or non-lethal response. Although the manufacturer indicates that the ultimate decision "should be made by a human not a robot" the system has an automatic mode. Similarly, Israel also has robotically

[103] HARPY is a fire-and-forget, UAV-based weapon system that can be launched from a ground vehicle or from a ship. It is designed to loiter for extended periods during which time it can detect and then destroy radar emitters. No operator input is required, although target verification can be insisted upon: the target radar "lights up," the UAV's onboard sensors acquire and compare the radar signal to a library of hostile emitters, prioritise the threat and (if verification is required/given) the drone enters its attack mode. If the radar is turned off before HARPY strikes, the UAV can abort its attack and continue loitering. If no radar is spotted during the mission, the UAV is programmed to self destruct over a designated area. The latest versions allow visual ID and attack of targets, even after they turn off their emitters.

[104] The SGR-A1 "robot" was jointly developed by the Korea University and Samsung Techwin Corporation. It was deployed along the South Korean border of the Demilitarised Zone (DMZ) in 2007 and uses a combination of electro-optic and infra-red sensors to detect and track mobile targets and to autonomously control, aim and fire a machine gun. The SGR-A1 is also reputed to provide verbal commands to its targets (in Korean) to surrender; apparently recognising the action of holding one's arms high in the air as the act of surrender, which prevents or suspends the automatic firing response.

controlled fifty calibre machine guns along its border with Gaza. The current policy is that "in the initial phases of deployment" there will be a man in the loop; leaving open the prospect of more autonomous operations in the future. There are a wide range of other examples of semi-autonomous weapons and weaponised UVS that include the Phalanx anti-aircraft/missile weapon[105], the Tomahawk anti-ship cruise missiles[106], the anti-submarine Captor mine[107] the TALON SWORDS, GATERS, Fire Ant and MAARS UGVs[108], and the Patriot III missile batteries[109] [23] [64].

At present, it is also possible for many of these systems to be re-tasked by someone from a different command chain to the person who originally initiated its use. Moreover, the person who first used the weapon may not be aware of the new mission or of the rational behind the change. Nevertheless, under the LOAC responsibility for the weapon lies with the person who actually uses it: if he has doubts over the outcome he should not launch it. In reality, therefore, past a

[105] The MK-15 Phalanx Close-In Weapons System (CIWS, pronounced "sea whiz") is an automatic, fast reaction, rapid fire 20mm machine gun that provides US Navy ships with terminal defence against attacks that have penetrated other fleet defences. It is designed to engage anti-ship cruise missiles and fixed wing aircraft at short range without assistance from other shipboard systems. It automatically engages incoming anti-ship missiles and high-speed, low-level aircraft that have penetrated the ship's primary defense envelope. It performs search, detection, tracking, threat evaluation, firing, and kill-assessments of targets autonomously, while providing for manual override.

[106] Tomahawk cruise missiles are launched in the general direction of their intended target. At some distance from the anticipated target location the missile enters a serpentine trajectory (several other patterns are also possible) to search for the target using passive and active radar techniques. Once the target is detected and the onboard algorithms satisfied, the missile locks onto the target and automatically enters its attack phase.

[107] The MK-60 CAPTOR mine is often referred to as the "mousetrap that chases the mouse." It uses onboard acoustic sensors to detect and classify hostile submarines, while ignoring surface ships and friendly submarines. Upon acquisition of a suitable target it launches a modified Mark 46 torpedo, which searches and homes in on its intended target using a circular search strategy.

[108] The TALON SWORDS UGV, developed by Foster-Miller & Qinetiq, is a tracked, semi-portable vehicle that carries an M240, M249 machine gun or a Barrett 0.5 calibre rifle. The system has been operationally deployed. The Modular Advanced Autonomous Robot System (MAARS) is a replacement for the TALON SWORDS and can carry a 40mm grenade launcher or an M240B machine gun. The GATERS (Ground-Air Tele-Robotic System) is a somewhat larger UGV from the 1980's that carried hellfire missiles. Fire Ant (manufactured at the Sandia National laboratory) is also of 1980's vintage and while tele-operated from a navigation stand-point it had an autonomous weapon firing mechanism; the GATERS sensor package was aimed by an operator, but the image processing algorithm autonomously fired the weapon in response to target motion (tanks).

[109] The Patriot missile system uses ground-based radar to detect, identify and track its targets. An incoming missile could be 80km away when the Patriot's radar locks onto it and consequently not visible to a human being, much less identifiable. The Patriot can also operate in a completely autonomous mode with no human intervention as an incoming missile flying at Mach 5 (i.e. traveling at about one mile/sec) does not leave much time for a human to respond, making automatic detection and launch an important feature.

certain point these weapons are no more controlled by humans than artillery or mortar barrages[110] and once the targeting package onboard the weapon is satisfied the target is destroyed. Consequently, at one level, these semi-autonomous weapons would appear similar to autonomous UVS with weapons, except that the IDT onboard the UVS is re-tasking the weapon 'on the fly' instead of a human from a different command chain.

However, the use of fire-and-forget weapons is only considered lawful if they are "equipped with onboard sensors or have access to external sources of targeting data sufficient to ensure effective target discrimination" [71]. Moreover, commanders do not launch these weapons on a purely speculative basis. That is, while there may not be a priori knowledge of the precise location or immediate accessibility of a given target prior to launch, there is usually some intelligence about the target (e.g. it is believed to exist, to be within a specified Area Of Interest (AOI), to be of a particular type, pose a particular threat, etc.). Additionally, the weapons may be safely destroyed if suitable targets are not acquired.

UVS, on the other hand, are often used in intelligence-gathering, surveillance and reconnaissance (ISR) roles, where there is no a priori knowledge of the existence or nature of targets. Given our lack of knowledge regarding the potential diversity, nature and location of any targets that might 'pop up' it is also unlikely that we will have a good understanding of the surroundings and hence the potential effect of weapons on non-combatants and infrastructure, except in specific circumstances (e.g. uncluttered environments such as air-air or underwater engagements).

Another way to describe the difference between the current and future technologies is that the extant weapons represent 'conduit warfare' whereas fully autonomous UVS and their weapons represent 'intermediary warfare.' That is, fire-and-forget weapons represent technology conduits through which warfare is conducted by extending the reach of the protagonists. However, they do not fundamentally alter the terms or conditions of engagement. Newer technologies, on the other hand, promise to remove some of the key elements of the decision-making process in the target prosecution process from the protagonists. Clearly, regardless of the degree of sophistication underlying the weapon's autonomy, there is a need to safeguard against unauthorised persons gaining control of the UVS or its weapons. In fact, surety of such control will likely be one of the underlying tenets of any such fielded system.

6.4 Accountability and Liability

In an earlier section we briefly discussed some of the issues of liability and accountability from the perspective of tort law. Unfortunately, if UVS are to be used operationally and the Principle of Responsibility is to be upheld, who is responsible for breaches of the LOAC involving an autonomous weaponised UVS? Although somewhat far-fetched, the following (based on [256]) will illustrate the matter.

[110] Which cannot be surrendered to and for which there are few moral or legal concerns.

Let us assume that a weaponised UGV has been tasked with some high priority goal. En-route to this objective it encounters adversaries, who surrender. The UGV recognises their surrender and stops to guard them, but also determines that the imposed delay will prevent it from achieving its primary objective. The UGV also determines that if it were to stop guarding the prisoners they would in all probability interfere with (and therefore prevent) it achieving its primary and higher priority goal: it kills the surrendered adversaries and proceeds to its higher objective.

This is clearly a war crime, but who is responsible? Moreover, as the Principle of Responsibility requires agents of war be held accountable for their actions, the argument may reasonably be raised that this alone precludes UVS from making such decisions or taking part in such scenarios. To some even the intuitive implausibility of having to hold a UVS to account will be sufficient to preclude their ever being allowed to make the final decision regarding the lethal use of force. There are a number of issues.

Providing UVS with separate legal personality and holding it 'responsible for its actions' does not imply that it has sentience, intentionality or will. Furthermore, we must consider reasons why humans commit war crimes: hatred, bigotry, racism, malevolence, and so on; all reasons that are characteristically intentional and not something that we expect to ascribe to a UVS. Unintentional reasons (i.e. a "failure to take proper care") really relate to levels of technical competence in the UVS, the users, manufacturers, etc and these are covered in tort law arguments. Interestingly, if a UVS had a moral sense the law would have something to act upon. Even so, infants have little or no moral sense, but are accorded legal personality and hence an implied level of responsibility. Similarly intelligent animals (e.g. dogs) are assumed to know their master's bidding [68] and may be punished for their disobedience.

On this matter, [256] helps us conceptualise the key problems of attributing responsibility to artificial agents by comparing them to child soldiers. Clearly, children have a degree of autonomy and are capable of a wide range of decisions and actions (significantly more so than any existing UVS, in fact). Like UVS, the children will *probably* kill the right person but may not. Regardless, it is widely recognised that, as the children do not fully comprehend the ethical dimensions of what they do, they are not the appropriate objects of punishment. That said, our moral repugnance at the use of child soldiers is perhaps more at the use of children per se rather than at the use of agents of war that have potentially unreliable or unpredictable target discrimination capabilities and a diminished sense of moral responsibility.

As indicated in the section on tort law the legal system ascribes liability by tracing the vector of causation back to the human agency where mistakes were made. The sins of omission and commission are then scrutinised with respect to negligence, recklessness, intentional malfeasance and so on [145]. To understand these issues in greater detail let us follow the approach of [256] and expand upon the collective ideas put forward by [104] [181] [246] [252] [256] and [297] and others and try to attribute responsibility for any accidents or infringements.

An Unfortunate Mistake

The allocation of responsibility is important for the principle of *jus in bello*. Sometimes accidents will happen, but if it is typically impossible to hold individuals responsible for the casualties of war then this violates the Principle of Responsibility. To this end, if the UVS does not over time demonstrate sufficient and predictable capacity to discriminate targets correctly, issues of culpability will almost certainly be raised. Consequently, while it is acceptable to 'apologise and move on' for the accidental destruction of a civilian target, some nations will probably use such an approach for target misidentifications made by the automatic target recognition algorithms (say, when a man with an axe is mistaken for a man with a rifle). However, this strategy is not likely to be acceptable when the violations become indiscriminate or numerous, or when the breaches of LOAC are more complex. Other candidates will be sought.

The User/Commander

The Statute of the International Criminal Court (ICC) provides that the responsibility for providing the legal means and methods of warfare falls upon the military using the technology [246]. To this end, the officer who ordered the deployment of the weapon is held responsible for the consequences of its use. In other words, he accepts the risk that the weapon may go awry when he makes the decision to send it into action, even if the consequences were not intended. Clearly, suitable precautions must be taken by the users prior to any mission.

However, as [256] points out, if we adopt this precedent as the norm for fully autonomous UVS, we neglect that current weapon systems do not autonomously select their own targets, except within the framework of a fire-and-forget weapon. Fully autonomous UVS and weapons, which attack targets that have not been chosen for them, have an inherent 'unpredictability' that must relieve the user from his responsibility – at least to some extent. In other words, the autonomy of the UVS implies that the orders issued by the supervisor influence, but do not prescribe, its actions as the decision-making programs are operating in the real world and make decisions unforeseen by the human operators at time of deployment.

The Acquisition Organisation

Clearly, prior to any acquisition and in-service deployment of an IDT-controlled weapon extensive testing and evaluation will be needed. For example, in order to accept liability the agency must quantify the anticipated performance and reliability of the IDT. To do this it will need a series of metrics, although historically, most of the results associated with artificial intelligence, robotics, UVS, and autonomous systems have been in the form of demonstrations. A trail of documentation would also need to be carefully and accurately recorded as there may be a need to record decisions made in the field for the purposes of evidentiary hearings, court proceedings, improvements, upgrades, etc. This is the case with all modern weapons systems, most of which now crucially depend upon information coming from multiple sensors, filtered and fused using IDT and provided to a human-machine interface (HMI) for a user to make a decision.

The acceptance of liability by governments for sophisticated equipment that can behave in a 'somewhat unpredictable' manner is not a novel concept. However, the IDT for an autonomous UVS may be so complex as to make reproducible testing very difficult, alternatively it may have the capacity to learn from its mistakes. Either way, the more intelligent the decision-making process, the harder it is likely to be to functionally establish or fully test the response of the system to repeatable or verifiable system stimuli. How then can an acquisition agency be held responsible for the actions of an IDT that it cannot fully test? Similarly, how can this agency accept liability on behalf of its nation or the likely users?

The Developer[111]

Given the complications articulated above it is tempting to consider holding those who designed, manufactured, or programmed the IDT responsible for its actions. However, inherent in the nature of an autonomous UVS is that its IDT will make choices other than those directly programmed into it. For instance, when immersed in the real environment, it is highly unlikely that the IDT's decision-making will be entirely predictable. This is not to say that it will be random, but that the internal states of the machine are likely to be so complex that the actions of the machines may be indeterminable [256]. Clearly, the designers of a system should not be held responsible if they cannot predict its actions or the environment in which it will be immersed.

Secondly, it seems likely that the developers will know about the possibility that the IDT could misidentify targets (and hence engage them in error). Similarly, if appropriate to the sophistication of the weapon, it is likely that they will understand that the IDT could potentially develop unresolved or contradictory states that result in other 'unpredictable' behaviour like that described in the UGV scenario at the start of this section; the systems will not be released untested. As companies employ lawyers to advise them on their responsibilities in producing, advertising and selling their products and UVS companies are unlikely to be any different in this regard it seems highly likely that the UVS developers will adequately fulfil their legal obligations in regard to any potential issues of product liability. Thereafter, assuming that there was not a faulty implementation of an otherwise acceptable design, blaming the developers would also seem to be both unfair and impractical.

Thirdly, in order to ensure culpability falls upon the guilty developer we must account for the complex and polymorphic nature an autonomous UVS. For instance, many UVS will comprise architectures where the decision-making elements are distributed across a number of programs and processors with results that are derived from data input from a wide range of sensors, a number of concurrently interacting components, machine and humans, none able to make the decision in its own right. There will, therefore, be interaction between multiple programs developed by multiple programmers (possibly from different software houses) on a variety of processors, operating systems, and architectures (perhaps unknown to each other in advance) and possibly distributed across a network of

[111] This includes the manufacturers, systems integrators, systems designers, soft/hardware engineers, etc.

UVS platforms. Moreover, the sensed data can probably be interpreted from a number of perspectives, each commensurate with different aspects of the integration or mission. Additionally, some military UVS may be designed to be unpredictable so as to inject a degree of flexibility or creativity into the system. Establishing a causal link would seem to be a task of Herculean proportions.

Finally, in all likelihood, the academics who devised the algorithms used in the decision-making process may have had no intention of it being used in this way. Therefore, to allocate any responsibility to them – even if their research was funded by a military organisation – would seem to stretch the bounds of legal or common sense.

Regardless of where responsibility lies, we must try to identify and apply a clear legal doctrine that simultaneously conforms to the LOAC (and any rules of the road), allows UVS development, and progresses its integration into capability such that it can be used with or without weapons. If we do not, those less responsible or less scrupulous, who also wish a capability edge, will likely simply develop and integrate less restrained versions into their militaries.

6.4.1 Legal Personality for UVS

Given the difficulties in attributing responsibility to the humans or their agencies, why not ascribe it to the UVS? To address this we must consider the legal concept of *Agency*. Agency is a highly specialised field of law but may be summarised as: an agent is empowered by their principal to negotiate and make various arrangements on their behalf; thereafter, the principals are bound by the contracts that their agents sign as if they themselves had signed, unless it is possible to prove misconduct. Moreover, the agent's actual authority extends to cases of apparent authority, where the agent has no actual authority but where the principal permits him to believe that he has authority [24]. To this end, as UVS become more sophisticated and are able to perform a range of complex actions on behalf of their human supervisors it is attractive to think of them as agents of their users.

The usual philosophical and legal debate over whether or not legal personality [112] can be ascribed to artificial agents centres around the list of necessary and sufficient conditions that must be met by the artificial agent in order for it to be recognised as an equivalent to a human. However, while UVS may ultimately achieve this status, this is not likely to occur any time soon. Consequently, we may choose to afford UVS *quasi-legal* status and allow them responsibility for a limited set of decisions and actions.[113] In this way, we might consider UVS to hold diminished responsibilities such that the liability for certain decisions could then be transferred to the UVS or differentially apportioned between the UVS and humans or their host organisations.

[112] Typically, a legal persona has the capacity to sue and be sued and to hold property in its own name [68].

[113] Some legal entities (e.g. children, corporations and the mentally impaired) frequently act through agents. Moreover, not all legal entities share the same rights and obligations; some (e.g. marriage) depend upon age whereas others (e.g. voting and imprisonment) are restricted to humans.

Clearly, in order to consider UVS as agents, we must imagine some form of 'contractual' obligation to exist between the user and the UVS. The mechanism through which electronic agents are able to create contracts is typically known as "attribution law" [149], which acknowledges the ability of electronic agents to conclude contracts independently of human review and alter the rights and obligations of their principals in their absence. In essence this law stipulates that a person's actions include those taken by human agents and those taken by electronic ones. In other words, any transaction entered into by an electronic agent is attributed to the person using it, the requisite intention flowing from the programming and use of the IDT. However, legal opinion seems to be divided upon the question as to whether this attribution accords with the general principles of property and contract law: different European courts have produced contradictory rulings [270]. Nevertheless, regardless of the current statutes, the legal principles remain valid.

Another legal concept that is useful is that of Copyright Law[114]. That is, it may be possible to think of the UVS and its human supervisor as holding joint rights to the decisions, in the same way as joint authors are considered co-owners of a single body of work. The human's contribution to the decision would be the input directing the UVS output, which might vary from significant to trivial depending upon the level of human supervision required. The key is that the decisions made by both parties are merged in an inseparable and interdependent manner such that they result in a unitary outcome [33]. There are however problems with this concept. For instance, what if the UVS adaptively learns from its environment and takes decisions independent of the supervisor's or programmer's original input or intention? Under most jurisdictions, the UVS would then become the 'owner' of the decision, because in copyright terms this would be considered a derivative of the original work?

There appear to be five options that may lead to a resolution of the problem; three that require us to apply or slightly modify our interpretation of current 'legal doctrine' (and possibly apply some 'fine tuning' to the existing law) and two that adopt more radical approaches, one requiring us to treat UVS as legal agents of their users, and another where the UVS are accorded separate legal personality.

6.4.2 UVS as Tools

Given the difficultly of ascribing intention to a UVS it is perhaps easier to acknowledge that they are just tools of their supervisors such that they allow lethal engagements to take place. This provides the most straight forward approach to

[114] The author of a work is the initial owner of the copyright in it, and may exploit the work or transfer some or all of their rights in that work to others. The author is generally the person who conceives the copyrightable expression and fixes it, or causes it to be fixed, in a tangible form [32]. However, there are precedents for those who conceive and fix the work in a tangible medium not holding the copyright. Those who pay for the creation of the work, rather than the employee who conceives the work, hold the rights. To this end, it may be useful to think of the human supervisor as a commissioner or employer of the UVS,[114] who in turn is the writer or executor of the work.

the problem, which is to insist on a rigorous interpretation of the current LOAC doctrine. In effect, it disregards the decision-making involvement of the UVS in the engagement entirely, thereby assuming the UVS never plays an active cognitive role in the target prosecution, regardless of its autonomy. The actions of the UVS are then ascribed to its user, regardless of whether or not they are intended, predicted, or mistaken.

While simple in its application, as the sophistication of UVS increases, it will be less and less realistic to assume that they are simply tools of their principals. Hence the limitations will become either unjust or inefficient and will almost certainly stifle the introduction of UVS onto to the battlefield, which in turn may lead to nations ignoring the protocols.

6.4.3 Human Intervention

Another straight forward approach is to simply consider the lethal prosecution of a target illegal unless human intention can be identified at all (appropriate) stages of the process. This is based on similar logic to that used in the preceding section and may in fact simply be a practical outcome of applying such a doctrine.

The likely impact of such a proposal, however, is that operational tempo will be diminished if the UVS must rely upon human intervention, particularly for complex, networked or high workload environments. The commensurate effect will then either be to impede the introduction of UVS into capability or to promote a disregard for certain aspects of the LOAC. As soon as such interpretations are justified by one nation, others will probably follow.

6.4.4 Technical Equivalence

Yet another straight forward strategy would be to accept that technological progress is largely monotonic and that, once UVS have demonstrated their capacity to discriminate between targets to a given level of competence, entry into war accepts that, just as humans can make mistakes, any UVS flexible enough to be delegated our judgements can also be expected to err. This is reinforced by the considerations of software and hardware reliability outlined in the earlier sections on UVS. In the absence of other criteria we could assume that human equivalence is sufficient to use as the threshold test for the level of automation that we are willing to accept as being within the LOAC. However, this again poses problems.

First, with further improvements in sensor technology, processing and fusion, a UVS will be provided with information that is equivalent to or better than that provided to a human. Consequently, in at least some circumstances the UVS will be capable of making superior assessments in regard to the application of lethal force. At this time, the push to insert such systems into military service will increase.

Second, there is a growing perception that aside from any improvements in sensor quality, technological advancement (e.g. agents that can filter and fuse data) can significantly affect the quantity and quality of information available to a commander prosecuting a target. This will enable the human to more precisely determine whether (for instance) non-combatants are to be placed at risk. This has two consequences: the first is to legitimise precision engagement and criminalise collateral damage, such that human equivalence may be inadequate [159]; and, the second is that the commander may increasingly rely upon indirect, filtered information, which may not contain information relevant to his obligations under the LOAC [177]. This issue may become even more acute in networked multinational coalitions when the information supplied by or to a force may be filtered for/from commanders of different nationalities.

Aside from the challenge of defining the various criteria by which we measure human equivalence, the other limitation of this strategy is that while the LOAC accept that civilian casualties are sometimes inevitable, it is also a requirement that the taking of life not be indiscriminate. That is, if we accept that it is legitimate for technology to err in terms of discrimination, some nations may use this approach to introduce indiscriminate targeting or breaches of LOAC in other areas of technology. In other words, while some would probably argue to the contrary, this solution could fundamentally undermine a key principle of the LOAC.

On the other hand, technology will continue to improve and it is reasonable to assume that the agencies who bear responsibility for developing UVS will continue to be charged with devising strategies for rectifying any technical problems with target discrimination techniques. These improvements will present a reference by which the quality of less discriminatory targeting UVS are judged in a way that is little different to other weapons technologies.

6.4.5 UVS as Agents

The final two solutions involve taking the artificial agent theory seriously by treating the UVS either as the legal agents of their users or as independent legal personalities.

The fourth option requires us to treat the matter as if the engagement were akin to a 'contract' between the two opposing parties, with the UVS acting as an agent on the supervisor's behalf. Then, in the same way that a person can be bound by signing an unread contract – or even a contract generated by a computer over which the principal has no direct control – we assume that the supervisor may be bound by the decisions of a UVS acting in accordance with his general intentions even though he was not cognisant of the detail of its actions in terms of any specific engagement. This option, however, requires us to distinguish between autonomy and unpredictability.

In the law of agency, a party's assent is not necessary to form a contract. It is sufficient, judged according to 'standards of reasonableness,' for one party to *believe* that the other party intended to agree; the real but unexpressed state of the other party's mind is irrelevant. In other words, at the point of use a commander will have certain expectations of the way in which a UVS will behave, deploy its weapons, etc. In our case this would mean that we adopt the doctrine that, without ever 'knowing the mind' of the IDT, a 'contract' between the user and the UVS is assumed. If the UVS behaves autonomously, but not erratically or unpredictably, responsibility for its actions would then be in accordance with the current interpretation of the LOAC. Alternatively, if the behaviour of the UVS is deemed erratic or unpredictable then it would be acknowledged that liability for its actions do not fall upon the users.

While attractive, the practical limitations of this approach are significant. For instance, the distinction between autonomous (expected) and erratic or unpredictable behaviour for UVS must be determined. That is, standards of reasonable autonomous behaviour must be established and defined. Aside from the significant technical hurdles of such an undertaking, most militaries are unlikely to want to expose the capabilities of their potentially leading edge technologies. Nevertheless, it may be possible to describe these in broad, but acceptable, qualitative or descriptive terms.

Even if the world's militaries were willing to share the relevant data, the UVS may be so complex as to make reproducible testing impossible, or they may have the capacity to adaptively learn from their experiences and environment. In fact, at present, due to the stochastic nature of most systems, the best that is likely to be achieved is a Monte Carlo simulation to determine performance averages and bounds. This is likely to be cost prohibitive for real systems and the problem will be compounded for multiple or networked systems. Additionally, to achieve our goal of establishing standards of reasonable and erratic behaviour it is not enough to simply determine statistical norms, we must also compare the stochastic performance of the UVS against a series of previously agreed metrics that represent (say) human or other 'normal' behaviour; and most would agree that these do not currently exist. Furthermore, we must then establish which criteria define human equivalence, how many of these are essential as opposed to merely useful, and what we do if the thresholds for several criteria are exceeded, but not others.

6.4.6 Separate Legal Personality

We have already indicated that lethal operations conducted by UVS do not appear to fall within any of the existing exceptions to Walzer's Principles; so why not create a new one? To do this we need to focus on the *fact* of the engagement rather than the *process* of the engagement. Specifically, we would need to accept that human intention need not underlie the use of lethal force in an engagement. In other words, we assume that the user's generalised and indirect intention is sufficient to render the lethal prosecution of a target legal. While reliant upon the technical competence of the UVS to achieve a certain practical standard (e.g. human equivalence) this would, in effect, extend the current interpretation of the Principle of Responsibility to incorporate the UVS, which would need to hold separate legal personality.

Holding separate legal status does not imply that the UVS is considered a person.[115] It simply recognises that being human is not a necessary condition of

[115] Although the technical legal meaning of *person* is "subject of legal rights and duties" [1].

being accorded legal personality.[116] Furthermore, there are well-established precedents for dealing with abstract and inanimate objects that enjoy legal rights and duties (e.g. nation states, the church, gods, shrines, corporations, ships, dead people, trees, and animals). There are also modern (and evolving) theories that define the legal criteria for agency and identity, not on the basis ontological properties (mind, soul, empathy, and the reflexive capacities) possessed by an entity, but on the basis of its ability to communicate [270].

Ultimately the legal status of a UVS will probably be decided as a matter of pragmatism rather than on the basis of "computational substrate or internal architecture" [255]. That is, it is unlikely that such legal personality will come to the courts ready-formed. Furthermore, unless there is demonstrable economic, doctrinal or capability benefit for such a change there is unlikely to be strong internal pressure from the military or legal communities to accord it. It is clear, however, that the vagaries of the 'rules of the road' and LOAC relative to the autonomy of UVS may provide impetus, although this should not be overstated. Quite possibly, the greatest considerations will be the political and financial cost-benefit analyses, as few modern laws are proposed without debate surrounding the relative merits of such packages.

6.4.7 Impediments to Artificial Agency

Artificial agents are now ubiquitous and include avatars (graphical icons that represent a real person), vReps (virtual representatives that are used to humanise online relationships and provide a single point of contact for customer enquiries through natural dialogue), digital buddies (software programs that automate chat with users), and shopping bots (programs that collect and compare information online and recommend products) [149]. These agents are essentially self-contained code units that interact with humans, one another and/or their environment. They often fulfil multiple roles, including information provision (e.g. filtering or fusing information), decision-making (e.g. the generation of goals, plans, and schedules), and action execution (e.g. closure of contracts). Moreover, their functions are similar – and in some cases superior – to the humans they replace. In the civil environment, however, an agent's actions are to a large extent reversible: the contract can be declared void, the goods returned, etc. In the military domain, the consequences of an IDT's actions may not be reversible as they may result in loss of life, destruction of cultural heritage, etc [246].

Some will simply dismiss the notion of agency or separate legal personality for UVS as fantastical and not something deserving serious consideration. Moreover, there are obviously a number of objections to non-humans holding legal status and rights. Several commentators (e.g. [12] [24] [68] [149] [255] [270]) have discussed these from the perspective of artificial intelligence, contract law, and the

[116] Similarly, being human has historically not been considered a sufficient condition to be recognised as a legal persona: in England, prior to the middle of the 19th century, a married woman was not considered to have separate legal status to her husband. Similarly, in the US slaves were considered "non-persons" until the Emancipation Proclamation of 1863 [68].

agents holding rights under various Articles of the US Constitution. The philosophical objections usually include:

- It is only right that natural persons be given legal rights;
- UVS lack soul, consciousness, intentionality, free will, etc; and,
- As human creations UVS can never be more than human property.

Readers are referred to the works above and those referenced therein for a more detailed discussion. Here we simply note that there are clearly contrary intuitions and arguments, and that each depends on values or assumptions that are not necessarily universally shared (i.e. is a child or a human clone a human creation?).

The commentators' various legal objections usually include:

- It is unnecessary as other solutions are adequate;
- The inability to sue an errant agent is not in practice a significant loss;
- The agent has no assets and so any judgement against the agent is meaningless[117];
- In multi-agent systems, which comprise multiple copies of the same code and are in communication with one another, it is not clear whether the entity should be recognised as singular or plural in a legal sense.

Other objections to treating artificial agents as true agents may be interpreted as:

- UVS would lack the legal power to give consent because they are not persons. However, as [68] points out, Roman slaves were not considered persons, but still had the legal capacity to enter into contracts on behalf of their masters.
- The agent onboard the UVS would not have the intellectual capacity or the ability to exchange or represent 'promises' of a nature that a user could depend upon. Here we note that in the civil domain web and software agents are ubiquitous and we interact with them every time we shop online.
- Some legal systems require contracts to exist between agents and their principals: artificial agents are not persons therefore they cannot enter into contracts in their own name. However, as we have already pointed out, in Anglo-American law it is only necessary for the principal to be willing for the agent to bind him as regards third party contracts.
- The final objection, which is raised by [149], relates to the need for the agent to have the mental capacity to comprehend the nature of the act being performed. Given that UVS are not likely to have human-like intelligence at this level of cognition for some time to come, the current legal doctrine would probably need to be modified before artificial agency could be allowed.

[117] Perhaps not the case if the UVS holds legal personality and has insurance purchased on its behalf.

6.4.8 Technical Impediments to Agency

In regard to an IDT onboard the UVS, we may make the following assumptions [145]:

- It will make decisions
- It will be employed for its creative intelligence and judgement
- It will be distributed across an ensemble of programs and processors
- It will be polymorphic and immersed in a changing environment
- Consequences may be unpredictable, surprising and unintended

Unfortunately, as we have seen from the earlier section on UVS, many of the decision-making elements of the UVS are distributed across a number of programs and processors such that no one processor, program, programmer or user will know the full context of certain decisions. Moreover, if the UVS forms a component of an NCW environment, it may be that agents originating at another node in the network are executed within the UVS (or vice versa), such that any users are unaware either when or where the agents are executing or from where they originated. The agents simply cooperate with one another across platforms & operating systems autonomously.

Furthermore, in complex environments, it is not necessarily possible to know (a priori) what concepts (legal or technical) are relevant to any given situation. For instance, an IDT may interpret data from a number of different perspectives and manipulate the information depending upon the nature of the mission or problem. Additionally, some defence UVS controllers may be designed to be unpredictable so as to inject a degree of flexibility or creativity into the system, as predictable systems are not necessarily optimal for military operations.

Developers will naturally strive to achieve ‘best practice’ by implementing basic rules of thumb, keeping their designs simple, providing suitable documentation and creating initially stable designs. Empirical evidence[118] suggests, however, that it is practically impossible to fully debug a sophisticated program on an unblemished processor. Moreover, we are considering the interaction between multiple programs on a variety of processors, operating systems, and architectures across a range of platforms (perhaps unknown to each other in advance) all operating in environments with which they are inherently unfamiliar. To this end, we can be reasonably sure that the UVS will malfunction at some level and that we are unlikely to be able to predict the specific nature or timing of the failures. This will present great challenges to traditional legal analysis in regard to the attribution of responsibility even at the processor level.

The upshot is that the operation of autonomous UVS entails considerable fundamental uncertainty, particularly in complex environments. The failure of a

[118] There are inherent problems with software reliability as, despite the best endeavours of mathematicians, programs cannot yet be verified as correct over an arbitrary set of inputs [145]. And the problem is not limited to software. The processors on which the software runs can also be thought of as programs encased in silicon and they too usually contain bugs. Consequently, from time-to-time, software fails.

UVS may then not always be due to human negligence during their creation, acquisition or operation, particularly if the layout of the card was devised by an automated system or a reinforcement learning algorithm was used to adaptively modify the code or coefficients in some algorithm. As a result it may be beneficial to avoid trying to identify a causal relationship between any injuries caused and the functional component level of the UVS. That is, we may conclude that there are a class of circumstances that allow identification of faults within or by the UVS at the systems level and that not do not require detailed investigation into what, who, when, or why the injury occurred at a component, card, processor or algorithmic level.

Although this may seem unreasonable and it is recognised that the law may hesitate to assign responsibility to a UVS or its IDT when it could more reasonably have been ascribed to someone who *might* have been able to control the outcome it may be preferable to the alternatives, particularly if the best that can be achieved through causal analysis is indeterminate or 'unpredictable.'

6.4.9 Is the UVS the Real Decision-Maker?

Perhaps the most potent objection to an IDT 'taking responsibility' is the question "Is the UVS the real decision-maker?" In order to take the option of separate legal personality seriously, therefore, we must establish whether, given its limited capabilities, the UVS is in fact be the real decision-maker. The key argument in this regard centres on whether the natural decision-maker (the human commander) making the decision to use the UVS in the first place constitutes the crucial discretionary judgement as the power to make these discretionary decisions essentially identifies the principal authority.

However, if we assume that the UVS will probably assume a role in warfighting on the basis that it improves warfighting efficiency and effectiveness – that is, for instance, it allows the prosecution of more targets per person – how can we reasonably argue that the real decision-maker is its supervisor, who may only be making a few discretionary interventions on behalf of the UVS, leaving the technology to cope with the majority of decisions, particularly if the human has no direct contact with the targeting process – i.e. the UVS has undertaken the 'mechanical' tasks required to identify and engage the targets. In these cases, it may be necessary to have a formal hand-over procedure so that we could say "The UVS was the decision-maker until 03:15 UTC on 12th June 2008, whereupon the human took over."

Another reason for objecting to the IDT being held as the real decision-maker is based on the belief that it cannot follow a system of rules (however complex) sufficient to enable it to make appropriate judgements and exercise discretion. Here there are a number of considerations:

Changes of Circumstance: It is not possible to design a UVS that can anticipate all possible targeting decisions and circumstances. How then can we expect it to make good decisions in certain situations? Conversely, if we cannot codify such eventualities, how can we prepare Rules of Engagement (ROE) for humans to make judgements consistent with the LOAC under the same circumstances? For a UVS (as for a human commander), the ROE can therefore be designed to minimise such possibilities. Secondly, the ROE can provide for changes of circumstance by specifying that, if the UVS finds itself unable to carry out its relevant functions, a change in decision-maker (i.e. from the UVS to the human) is permitted.

Decision Outcome: From time-to-time humans wish to improve decision outcomes not simply by recognising novel or more complex approaches within the frame of an existing problem but by re-framing the basis of the problem itself. In this way it might be considered that they have superior decision-making capacity to a UVS. As a minimum, therefore, we must design a UVS with the capacity to recognise and deal effectively with complex novelty and to maintain information regarding any decisions it has made in a form readily interpretable by humans. In this way, we reduce objections to a level commensurate with our (a priori) capacity to codify such eventualities for humans.

Moral Judgements: How would an IDT exercise a moral judgement to implement the LOAC against a simple test of fairness – as opposed to a literal interpretation of what is strictly permissible? This is a particularly significant issue and is discussed extensively by [23] and briefly in the next section on the Ethical Control of UVS.

Legal Choice: In order to reduce the complexities of allocating responsibility, it has been suggested that a UVS can act as a quasi-agent and may be considered legal persona (that might hold insurance). Given this, the UVS may need legal representation in any dispute over liability (as a legal persona it will also have certain legal rights). How could the IDT exercise judgement and discretion regarding over matters pertaining to (say) the settlement of a complex lawsuit? The most straight forward response is that we have already determined that the IDT must maintain a suitable record of action for the purposes of evidentiary hearings, which would seem to satisfy the most basic requirement of this objection. Thereafter a lawyer could be engaged on behalf of the UVS, who could act as 'trustee' for the UVS in regard to any decisions regarding litigation. This would seem to be similar to the representation of minors in acrimonious child custody cases under various jurisdictions.

Before leaving the discussion on legal issues, two final thoughts:

- First, it is often stated that "Humans will always make the final decision" because they have 'mastered logic' (or some derivative of that phrase). The implication is that they can apply rational human judgement to special circumstances in a manner that IDT cannot. While it is true that humans are generally good at pattern recognition – and therefore recognising 'abnormal' circumstances – it should be noted that only *some* humans have mastered logic to a high degree; others have not. On the other hand, IDT now have the edge over most humans when it comes to assimilating, ranking and storing information from a diversity of sources, processing it using mathematical formulae and then interpreting and applying the results of these calculations as actionable outcomes.

- Second, if the UVS holds separate legal personality and has obligations, it will presumably also have rights. Consequently, it might be reasonably argued that it holds some form of legal title over the hardware and software that enables it to operate [254]. At the very least, it might seem reasonable that the insurance company might apply for some form of title to protect its financial interests.

6.5 Ethical Control of UVS

Autonomous UVS will be integrated into military service when they are technically capable of undertaking their defined roles, it makes sound economic and military sense to do this, and there is an appropriate legal framework to accommodate them. There may also be other advantages. For instance, what if we could ensure that UVS make morally superior decisions to a human? Among other things, this might reduce the diversity of interpretations of LOAC, which would not then be made against a backdrop of mainly Western, but nevertheless varying, ethical systems and international cultures, but would be superseded by a more unified and global interpretation of them that would evolve from the specific implementations of what is technically achievable.[119]

Such challenges have held fascination for humans since Isaac Asimov published his three laws in “I, Robot” [27][120] and risks sounding as though we are now crossing the line into the realms of science fiction or fantasy. Nevertheless, if we consider that many automated systems are now in the position of making split-second decisions that have life-or-death consequences it is perhaps not so far-fetched a notion. Furthermore, if we also interpret the challenge from the perspective of what it would take to design an IDT capable of making decisions from the perspective of what the UVS *ought* to do rather than simply allowing it to do what it *may* do,[121] the discussion sounds more plausible.

Let us start by considering an industrial robot. It has limited autonomy and must ‘choose’ between courses of action based on appropriately ‘moral’ judgements, albeit that they are effectively framed a priori by safety standards (e.g. [210] [257]). These systems tend to operate in very limited frames of

[119] We make extensive reference to the work of Arkin and other’s in this short section. The aim is to highlight some of the complexities and opportunities for instantiating the ROE and LOAC-based control of autonomous weapons. For a more extensive treatment of the topic the reader is referred to [19] – [23].

[120] These are: (i) a robot may not injure a human being or, through inaction, allow a human being to come to harm; (ii) a robot must obey orders given to it by human beings, except where such orders would conflict with the First Law; and, (iii) a robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

[121] That is, how might we ensure that the UVS behaviour is in line with what we expect of it in accordance with the LOAC so that it “shows compassion” when the opportunity arises.

reference and it is reasonably safe to assume that the engineers and programmers who design and build the systems may be able to conceive the variety of options the system will be presented with. Consequently, they are able to devise and program appropriate responses. As a result, these systems do not need to specifically evaluate the consequences of their actions on the basis of some moral code or ethical stance; this has been done for them a priori on the basis of (say) the occupational health and safety laws and the common sense of the design team. On the other hand, there are a number of factors that suggest that military UVS will need explicit ethical governance in order to comply with the LOAC [22].

- ➢ The UVS may encounter situations that the designers had not anticipated, either because the UVS is used in a way for which it was not designed or because the designers had not anticipated some complex aspect of the environment;
- ➢ The increasing autonomy of the UVS will allow it to make increasingly complex decisions. This has an impact in two regards:
    - o The variety of context that the UVS could find itself in will preclude a priori interpretation of each and every situation
    - o The number of potential decisions or degrees of freedom available to the UVS will preclude a priori knowledge of each and every potential action
- ➢ The polymorphic nature of the system and the sheer complexity of the systems engineering precludes accurate prediction and testing of UVS behaviour

Despite this, we can articulate some guiding principals for implementing ‘ethical control’ in UVS by following these basic attributes [16]:

- ➢ Consistency – any contradictions in the informing theory must be avoided (i.e. an action must not be simultaneously right & wrong in a given set of circumstances)
- ➢ Completeness – the system must ‘know’ how to act in any ethical dilemma
- ➢ Practicality – the technical implementation must be feasible and we must be able to follow any action recommended
- ➢ Agreement with Intuition – the actions must agree with expert ethicist intuition

Furthermore, it may be that we encode the governance of the UVS relative to the LOAC in line with the ‘decomposition’ strategy outlined in the section on Verification and Validation. That is, rather than attempting to consider the autonomous UVS holistically or at a system level, we devise a strategy that allows us to divide it into its constituent functions and verify that each of these components or functions behave ethically. In this regard [37] presents an interesting methodology, tools and techniques for modelling the heterogeneous real time components of a system as the superposition of three layers: a behavioural layer (specified as a set of transitions), interactions between the behaviours, and priorities (used to choose between the possible interactions).

The central idea – and the one most pertinent to the instantiation of ethical governance or control – is that while complex systems are built through the assembly of components, the main problem is that these components must be integrated in a way that ensures their correct interaction. Aside from obviously describing the key integration challenge for the functional elements of a UVS (and therefore possibly of considerable use in terms of V&V), this methodology also describes the overarching problem for the ethical control of a UVS within the framework of ROE and the LOAC. That is, the guiding principals may be considered consistent, complete and practical when interpreted in human terms, but there is no mechanism for provably demonstrating the composition, integration and software coding of these 'abstract' principles have 'mechanical' integrity. Consequently, the enactment of any physical functionality within the UVS against these 'rules' in a way that preserves the intended properties of the LOAC framework in real time is unprovable. However, the BIP (Behaviour-Interaction-Priority) [37] framework provides a basis for the study of property-preserving transformations or transformations between sub-classes of systems or principles, which in turn allows such component interactions to be verified.

6.5.1 Could UVS Ethically Out-Perform Humans

Humans have an impressive ability to rapidly, spontaneously, and effortlessly recognise and identify a large variety of objects even under unusual conditions. On the other hand, a human's capacity to detect, locate, and track certain objects, particularly in cluttered or attenuated environments, at long ranges, in highly stressful situations, or for protracted periods is inferior to those of sensors combined with intelligent processing. Moreover, the human eye has limited spectral sensitivity, does not see well at night, and its processing can be tricked rather easily [130]. With improvements in sensor, sensor processing, fusion, and communications technology it therefore seems likely that an IDT will be provided with the capacity to acquire, select, represent and retain certain types of information with a capability approaching or exceeding that of a human.

This inevitably leads us to the conclusion that in some circumstances the UVS may be capable of making superior assessments in regard to the target identity and hence the application of lethal force. For instance, during periods of high stress a soldier may feel the need to act in self-defence and may respond lethally to all targets within a given sector, which may inadvertently cause civilian casualties.[122] As IDT-controlled UVS need make no appeal to self-preservation they can value civilian lives above that of their own. This does not negate the likelihood that civilian lives will be lost, but it does indicate a way in which the IDT might ethically out-perform a human. Moreover, before proceeding (or dismissing as fantastical the notion that autonomous ethical governance could out-perform a human), it is instructive to consider the findings of [262] and which were first

[122] Self-defence is a common justification for the exculpation of responsibility for civilian casualties [16].

noted in this regard by [22]. The findings were taken from a sample of 1,320 US soldiers and 447 US marines.

- Only 47% of soldiers and 38% of marines agreed that non-combatants should be treated with dignity and respect;
- 44% marines and 41% soldiers agreed that torture should be permitted to save the life of a fellow marine/soldier and 39% marines and 36% soldiers agreed that torture should be permitted to obtain important intelligence about the enemy;
- 45% of soldiers and 60% of marines stated that they would not report a fellow soldier/marine if he had injured or killed an innocent non-combatant;
- 33% marines and 27% soldiers did not agree that their NCOs and officers made it clear that they were not to mistreat non-combatants;
- Even though they reported receiving ethical training, about 30% of those surveyed reported facing ethical situations in which they did not know how to respond
- Combat experience – particularly the loss of a colleague – was related to an increase in ethical violations.

In addition to the negative aspects highlighted above, there are also a number of ways in which UVS might out-perform humans on the battlefield [19]:

- They can be designed without emotion, which might cloud their judgement;
- The absence of a sense of self-preservation means they can be used sacrificially;
- They are unlikely to be affected by decision bias – the interpretation of (usually contradictory) incoming information such that, particularly in stressful situations, it is fitted to pre-existing scenarios or belief sets of the decision-makers;
- UVS can potentially integrate more information from more sources and more sensors more quickly than humans can (without getting fatigued or distracted);
- The sensors onboard the UVS might be used to report or police human behaviour (independently and objectively – unlike their human counterparts), which might also lead to a reduction in unethical battlefield behaviour over time;[123]

Before any responsible military could proceed with the introduction of such a capability, however, a series of metrics aimed at determining whether or not the

[123] It should be noted, however, that while mobile sensors on the battlefield might be used as policing devices this may have a negative "big brother is watching" effect on some of the troops in combat. This might in turn impact negatively on intra-unit trust and cohesion; although this was not the experience of police forces that introduced video taping into the interview process.

UVS actually delivered greater ethical behaviour on the battlefield would need to be established against the principles of the LOAC. Such a set of criteria might include [228]:

- Does the prospect of lower risk to their own forces result in the more frequent use of military force by governments
- Are autonomous UVS able to achieve greater levels of target discrimination
  - Are there fewer non-combatant casualties
  - Is there a reduction in collateral damage
- Do the IDT improve proportionality
- Are fewer of the forces that use the UVS killed or injured
- Could a captured UVS be used malevolently

6.5.2 Implementation Issues

Regarding the instantiation of ethical governance, there are two primary issues:

- What is considered ethically acceptable; and,
- What can technically be achieved?

In a study conducted by [198], where the researchers have canvassed opinion on the use of lethality by autonomous systems, initial responses to the first of these questions are presented. The survey demographic spans the general public, researchers in the field of robotics, policy-makers, and military personnel. The results of this survey are not discussed here further, but are reported in [23]. It should, however, be noted that ethics are relative: an act carried out by a UVS may be morally acceptable to a programmer, but perceived as immoral by indigenous cultures that witness its actions in theatre [22].

In regard to the second question, and relative to 'everyday' ethics, a number of authors [23] [46] [55] [180] [194] [198] [228] have acknowledged that the LOAC provide us with a 'rigid' framework that reduces the complexity of implementation somewhat. In other words, we neither want nor need the UVS to derive its own code of beliefs from first principles regarding the moral implications of the use of lethal force. Rather we wish it to apply those that have previously been derived by humanity and coded in the LOAC. All the same, we must be able to certify that our UVS behaves in accordance with these coded laws and any ROE, which (fortunately) are designed to be self-consistent, even if not universally agreed upon in terms of specific definition.

Implementing an artificial 'moral agent' that has the capacity to govern the actions of a UVS will involve a broad range of engineering, ethical and legal considerations and a full understanding will require a dialogue between philosophers, legal theorists, engineers, computer scientists, developmental psychologists and other social scientists. Moreover, it would be naïve to assume that such an agent would solve all issues of responsibility; this exercises even the International Criminal Court. Similarly, it is not anticipated that such a system would have the capacity to adequately interpret all situations to the satisfaction of the local commander or user. Under these circumstances it is presumed that they could assume responsibility for over-riding the agent.

In addition to the difficulties outlined in the section on Contextual Decision-Making, therefore, the practical instantiation of an ethical moderator will tax us in a number of ways:

- The creation of such a 'technical conscience' will challenge our understanding of morality to its extreme as, in order to code the conscience, we must formally express our knowledge of the moral framework in a manner that engineers and computer scientists can both understand and express in software/hardware [290];
- Ethical reasoning is based on abstract principles, which often conflict with each other in specific situations. If more than one law, code or principle applies it is often not clear how to resolve the conflict as the favourite tools of logicians and mathematicians (e.g. first order logic) are not usually available [178];
- The LOAC and ROE are intended to be self-consistent and complete. Moreover, they provide a broadly agreed framework upon which we can agree a basis for the development of this agent. However, the premises, beliefs and principles that humans use to make ethical decisions are varied and often intimately linked to religious beliefs [16] (i.e. there are a variety of accepted ethical theories upon which to base our computations).
- Additionally, the interpretations of the LOAC are not yet agreed between even the democratic nations of the world [16];
- If such an agent is able to prevent an IDT from acting in some unethical fashion it must also have the capacity to explain to its supervisor (at least to some degree) the underlying reasons for its logic [14];
- The conditions, premises or clauses are not precise, are subject to interpretation and may have different meanings in different contexts [13]; and
- The actions and conclusions that we obtain from such a system are likely to be abstract, so even if a rule is known to apply the most appropriate action may be difficult to execute due to its vagueness [22].

There are a number of groups working on the development of techniques that represent artificial ethical behaviour (e.g. [7] [14] [22] [55] [180] [194] [198] [228]), most proposing to use a variety of ethical theories (e.g. Kantian, Utilitarianism, Social Contract, Cultural relativism, etc). Several authors have pointed out that a distillation of ethical directions from the LOAC – a 'top-down' approach, if you like – might require us to instantiate a more rigid structure than we would wish; and thereby risk overlooking events or circumstances which then cause the UVS to behave inappropriately because it is bound by a set of rules. Alternatively, we might seek to use an 'optimisation' or 'bottom-up' approach that relies upon adaptive programming or machine learning. Constraints (e.g. the LOAC) could then be applied and the UVS placed in a variety of circumstances (training cases) to make it learn using trial and error. However, this too may have its difficulties.

First, because it does not have a complete 'set of rules' to guide its overarching behaviour the UVS may effectively assemble its own, and, if those are inadequate (or the training data is incomplete), the UVS may once again perform inappropriately.

Second, the UVS will be trained against data selected by their owners and operators. Just as education can enhance cognitive skills it can also create fanatics, dogmatics, cynical manipulators, as well as prejudiced, confused, and selfishly calculating minds [47]. Consequently, while the developer of a reinforced learning algorithm that is aimed at enhancing the capabilities of a UVS would presumably expect it to be immersed into environments and then trained or used in a way that leads to improvements in the IDT's response options relative to the LOAC, there may be those less scrupulous who seek to exploit (and possibly identify) aspects of an algorithm that result in the UVS-equivalent of a fanatic. That is, a system that over-rides some pre-ordained taboo, and otherwise high-priority, action on the basis of some learnt behaviour. This obviously introduces another level of complexity into the problem.

Arkin [22] approaches the topic from the pragmatic perspective of first applying it to limited and specific contexts and, as they learn, applying it more widely to new contexts and applications. The work approaches the design of the weapon control not simply as a problem in ethics, but from the perspective of safety. It uses a hybrid architecture that strictly adheres to the rights of non-combatants by regarding discrimination paramount (i.e. using deontological reasoning) while simultaneously considering proportionality on the basis of military necessity (i.e. using utilitarian reasoning). It effectively starts from an extension of the medical premise "First - do no harm"[124] (i.e. do not engage an enemy until obligated to do so). Based on the derived situational awareness picture, a set of circumstances, C, are defined which are parameterised and characterised on the basis of interpretable stimuli, S. C then comprises vectors to represent the building blocks of situational awareness [96]:

- The perception of the elements in the environment;
- The comprehension of the current situation; and,
- The projection of the future status.

The development of each of these blocks represents increasing levels of complexity. Initially, the interpretable stimuli, S, might consist of a vector expressed in terms of a perceptual class[125], a threshold value for each perceptual class, and so on. Complex, statistical analysis of (say) the geographic distribution and disposition of these classes would then allow us to parameterise and define circumstances that can then be passed to our 'ethical moderator' and acted upon in

[124] As opposed to the "Shoot and ask questions later" or "Apologise and move on" philosophies.

[125] To represent people (non-combatant/combatant), buildings, walls, vehicles, weapons, gestures, etc.

conjunction with the UGV's purpose, P, which is essentially the same as the (parameterised) high level mission or goals of the UVS set by the high-level interaction between the UVS and its supervisor.

A range of actions, A, is then defined for the UVS (e.g. its motion and payload response) that can be factored as two orthogonal components: strength and orientation. Strength would denote the magnitude of the response (e.g. speed or force), orientation the direction of action (e.g. retreat from target, move towards target, warning shot, engage target, etc). The instantaneous response of the IDT would then be expressed as a vector representing each of the degrees of freedom (DOF) for the UVS where weapons targeting and firing are considered within these DOF. We may now write a set of maxims [194]

If $(C \text{ and } P) \rightarrow A$	A is Obligatory for the IDT
If $(C \text{ and } P) \rightarrow \neg A$	A is Forbidden for the IDT
If $\neg((C \text{ and } P) \rightarrow A)$ and $\neg((C \text{ and } P) \rightarrow \neg A)$	A is Permissible for the IDT

Actions can then fall into one of three classes: obligatory (the UVS must act in this way, based on moral grounds), permissible (the act is morally acceptable, but not required) or forbidden (the act is morally unacceptable). Only acts that are deemed to be obligatory may be carried out. In other words, if an action is permissible it may only be carried out if it is considered obligatory in the context of the mission.

Using such a technique, we may now make sense of scenarios such as "UGV under attack and motivated by the need to protect Blue Force (Circumstance) shall try to reduce threat (Purpose) by responding with lethal force (Action)". As previously mentioned a record of action (in this case the UGV's ethical deliberations) would also be required; not just for after-action review, but also for the application of reinforcement learning techniques.

Once again, this does not imply the UVS has sentience or intention as we have effectively just articulated a way to deliver 'artificial morality'. That is, getting a UVS to act as if it were a moral agent. The first step in this process was to understand – and to provide the UVS with cognisance of – the possible harmful consequences of its actions, such that it can select from its repertoire accordingly.

The point of this commentary is to provide an indication of what might currently be possible and to highlight that development of such moral reasoning capability presents us with an interesting conundrum [19].

> If a UVS is given an order such that $(C \text{ and } P) \rightarrow \neg A$ (i.e. the UVS effectively recognises that it has been told to carry out an unethical order), given that the UVS may hold separate legal identity, should it refuse to carry out the order? If it does not who is to blame?

In practical terms, this actually returns us to an earlier discussion regarding location of the crucial discretionary decisions. Clearly, if the UVS is to assume a role in warfighting on the basis that it improves warfighting efficiency and effectiveness – and that warfighting remains for the foreseeable future a human endeavour – it will once again be necessary to have a formal hand-over procedure so that we can say “The UVS was the decision-maker until …, whereupon the human took over.”

Interestingly, if we record the complex decisions being computed by the IDT, which may be necessary data for evidentiary hearings (and to develop and test any ethical governors) it may be possible to combine this recorded data with other ethical reasoning agents to improve and better understand complex matters of liability and reasoning in stressful and cluttered environments. Similarly, we may be able to apply adaptive learning techniques to the ethical reasoning components of the IDT and use a range of techniques that model moral and emotional perceptions such as guilt. This may then in effect allow us to use moral motivation for UVS behaviour, which somewhat ironically, may provide us with the capacity to ‘punish’ the UVS for errant behaviour – or provide the UVS with the incentive to commit war crimes.

6.5.3 Architectural Considerations

According to [184] a good architecture describes the following:

- The overall system organisation
- Ways in which the architecture supports likely changes
- Components that can be re-used from other systems
- Design approaches to allow standardisation of functional components
- How the architecture addresses each system requirement

There are at least four stages (i.e. architectural possibilities) that would allow moral control of an autonomous weapon, each representing a more sophisticated level of ethical governance or control. These are described by [19] as:

Responsibility Advisor - While it may seem fanciful to assume that machines will take over the role of ethical decision making in war, it is more reasonable to assume that, as technology allows, they may be used in some advisory role. Thereafter, when human decision-makers have been put in the position of passively receiving interpretations generated (for instance) by data fusion and hypothesis generation-aid machines and they become less able to recognise emergent problems or be able to interpret the complex ones, they may progressively defer to their autonomous advisors. It is a small step from here to accepting their autonomous governance [20].

This would form a good first step towards any eventual goal of developing an autonomous UVS that acts ethically. Essentially, the advisor would probably form part of the HMI used for pre-mission planning and managing operator over-rides. It could potentially advise (either in advance or during the mission) the UVS operators and commanders of their ethical responsibilities should the UVS be deployed in specific or complex situations. It would then require their explicit authorisation to use lethal force. One could start by designing an advisor that provided guidance on the interpretation of the LOAC and ROE to selected personnel in specific circumstances, thereby retaining the notion that the human supervisors are making the decisions as it is they who must decide whether or not to follow the machine's recommendations. Another benefit of such a system would be that the necessary data required to make such decisions would start to be collected [20]. Additionally, as the interpretation of the LOAC has not yet converged to the point where there is universal agreement, difficult decisions could be flagged and passed to human specialists for later or more detailed analysis.

The ensuing architectures are likely to be instantiated when the performance of the ethical advisor has been shown to outperform human interpretation of the LOAC and ROE. Commanders and supervisors would eventually be compelled to follow the machine's advice or take responsibility for the actions of the machine thereafter. Over time, one would expect the role of the human in the ethical decision-making process to diminish and the number of dilemmas he handles to become considerable; perhaps he devotes little or no time to any particular one. There will still be times when a human is called upon to make a decision, however, and if such events occur with regularity, the developers of the system would need to be tasked with developing improved architectures for handling these, described by [19] as.

Ethical Governor – This is an extension of the previous design that physically transforms or suppresses UVS-generated lethal action to permissible action; either converting it to non-lethal action or permissible lethal action.

Ethical Behavioural Control – This approach constrains all individual control behaviours to be only capable of producing lethal responses that fall within acceptable bounds.

Ethical Adaptor – This architecture would provide the UVS with the ability to update the agent's constraint set and ethically related behavioural parameters, but probably only in a restrictive manner.[126]

To this end, we need to understand the requirements of such architecture. For example, if weapons are ever placed on a UVS, the equivalent of a Weapons Safety Board (WSB) will need to sign off on the system's architecture with respect to the safety of the overall system. To this end, we would need to start by agreeing with the WSB on a level of automation with which they are happy and then work with them to arrive at an approved architecture. We will then have to determine and articulate the instantiation of this architecture against the state-of-the art technologies, such that this instantiation is also signed off. To do this, we will need to define the sensor, technology, and 'intelligence' capabilities (most of which are software) that allow the requisite level of target

[126] Arkin [19] offers the view that such a component would be based upon an 'after-action, reflective review of the system's performance or by using a set of affective functions (e.g. guilt remorse, grief, etc.)' produced by a violation of the LOAC.

detection, discrimination, location, tracking, etc intrinsic within the agreed systems architecture.

Implementation of such a system is likely to be modular with specific code units run on physically separated hardware. It is also likely to have a variety of quarantined time-outs, queries, and messages to shift the weapon control unit back to a safe-operating mode, if required. Distributed systems are likely to introduce another level of complexity. Given the embryonic status of the development process, another desirable property is that the architecture supports growth so that extensions can be added incrementally. For instance, initially we are likely to want to test the systems against a representational yet small sub-set of forbidden and obligatory constraints; as opposed to trying to encode the entire LOAC and interpreting complex ethical situations at the outset [22]. In this way, the basic research required prior to the deployment of any operational systems can be conducted away from the complexities and heat of the battlefield.

Although this certification process is likely to be a lengthy one, we can postulate acceptable generic architectures. For example, it is likely that the governor will make decisions by projecting action into the future on the basis of a model of the UVS, its current and potential behaviour and any environmental dynamics. It will then evaluate the outcomes according to a function or some other selected criteria. This evaluation function then represents the objectives and ethical constraints through the return of high values for plans that meet mission goals without violating the performance envelope of the UVS. Typically, this will involve some form of search through a set of potential plans until an acceptable or feasible plan is found. Consequently, the key is to apply pruning techniques so that only successful outcomes are generated. As with code and system verification and validation there are several major risks:

- The moderation outcome makes inefficient use of resources;
- The moderation outcome could not be generated;
- The moderation outcome generated is not feasible;
- The moderation places non-combatants or civilian infrastructure at risk;
- The moderation places users, supervisors (or possibly the UVS) at risk.

It can be argued that we need to worry about the last one the most. However, to assure ourselves that this (and that the governor is working correctly) is not an issue we do not need to verify the entire system, only its evaluation function. If the evaluation function is correct then the UVS and/or users cannot be placed at risk. Furthermore, it is likely that human oversight will (initially at least) 'double-check' the results.

We may now imagine an architecture or algorithmic sequence, which combines the major Principles of the LOAC (i.e. responsibility, military necessity, target discrimination and proportionality) such as that suggested by [23].

> ***Responsibility*** – Permission for the UVS to use lethal force autonomously or in specific situations (e.g. prior to or at certain points in the mission) is granted by the human supervisor. If relevant, the type or nature of weapon selection is also authorised.

Military Necessity – The UVS makes a determination, based on its observation, comprehension and projection of the circumstances and its assigned mission, as to the need to use lethal force at any instant. It may determine (for example) that it needs to move to obtain a different perspective, fuse data from other on or off-board sensors, or allow the use of lethal force.

Discrimination – Targets and other objects of interest (combatants, non-combatants, buildings, and so on) are accurately identified in a manner that allows positive discrimination between civilians and combatants and civilian and military objectives.

Proportionality – Based on the disposition of the objects of interest and the likely impact of the use of lethal force (weapon selection, firing pattern, etc) collateral damage calculations are performed.

Obligation – In order to satisfy the goal that the UVS is doing 'what it ought' rather than 'what it can' in order to use lethal force the parametric criteria must satisfy the ethical moderator not just that the proposed action is permissible (i.e. that any and all forbidden constraints are upheld), but also that the action is obligatory (i.e. at least one obligating direction must be upheld).

Weapon Control – The UVS then attempts to accurately control delivery of the weapon with respect to aiming at any targets and maintain a suitable record of action for the purposes of evidentiary hearings and adaptive improvement.

Given unlimited computational capacity this problem could then be treated as one of dynamic, constrained optimisation set in a time-varying environment. The governor could continuously re-compute and execute policies over some multi-objective cost function, adaptively learning from its experience. Unfortunately, there are limits to the processing capability available, tasks relating to the autonomous application of lethal force are likely to be highly time-constrained, the constraints and the solutions to any cost functions will only provide good approximations to the specific parameters under consideration, and the optimisation will depend upon the real world variables being optimised. As a result, the best we will likely achieve will be an approximate solution to a precise set of principles.

Furthermore, given the modular architecture proposed, each element of the system would need to detect that it has failed and inform any other components. These requirements are strongly linked to the need for the system to determine conditions under which the prescribed directives are unachievable, either within a required time frame, the broader capability framework of the sensors, weapons, etc, or for some programming shortfall in the ethical moderator. By having this level of self-awareness, and notifying users of such limitations, the human can then over-ride the moderator, as appropriate.

Under these circumstances, or if the ethical moderator has declined to use lethal force for any other reason, the option for operator over-ride should result in responsibility for any actions once again resting with the human. The complication here is that, unless the reasons are appropriately explained to the user, they may not be apparent. However, if the algorithm uses the sequence above this might provide a framework that could help the machine explain its reasoning to the human in terms they can understand. Whether this can be done in sufficiently timely manner for the human to make an informed judgement under the pressures of battle, however, is less clear.

It is clearly a fundamental requirement that the moderator be able to provide permissible and obligatory solutions. To do this it may attempt to recognise those that are not permissible, based on its interpretation of the ROE, its own capabilities, etc. As a result, it is possible that the moderator could work in this domain, defining solutions that it cannot achieve in order to determine those that it can.

However, thus far, we have cast the solution as a binary problem for which the actions are either permissible or not. Relative to the capabilities of a specific UVS there may be areas of grey where solutions are difficult rather than impossible, and that further human judgement needs to be applied. One solution to such challenges is to attempt to explicitly compute the ethical cost functions that are defined in UVS response-behaviour space. These plans may then be treated as input to reinforced learning techniques that can then learn by physically interacting with the ethical deliberations. However, even though the early instantiation of such behaviour bounds are likely to be relatively straightforward and well-understood, the abstract principles of ethical reasoning and the absence of first order logic will likely lead to unknown or un-modelled factors. Consequently, the evaluation function is unlikely to be ‘crisp’ and the application of any reinforced learning techniques complex.

A decision must also be made as to how to structure the ethical governance algorithms. As with UVS planning systems, the computational advantages of using partial solutions are attractive, but present difficulties. For example, if the process is interrupted a partial solution algorithm may not provide a feasible strategy, whereas complete solution approaches should always be able to provide at least one feasible plan; and it is likely to be imperative that a timely solution be available at all times.

As with mission-planners, once operating most ethical moderators cannot simply stop to compute a new plan every time the circumstances demand it. Consequently, at a most basic level, it is also essential that the moderator prescribe solutions that have ‘lives’ longer than the situational awareness predictions of the UVS. If this is achievable the UVS will at least operate ethically within its environment while other plans are computed. In order to achieve this, however, moderation cycles must be performed concurrently with normal system operation and it is impossible to know a priori what update rates will be required for such systems. As a result, if possible, we should incorporate a degree of adaptive or reinforced learning into the prediction component of the adaptive algorithms to allow them to determine their own update requirements. That is, we should allow them to learn from changes in the environment and their own deliberations in order to accommodate a better sampling frequency of sensor inputs and moderation outputs.

However, as previously mentioned, when applying any reinforced learning techniques to ethical moderators, caution will need to be exercised as the UVS will be trained against data selected by their owners and operators. As a result, while clearly aimed at enhancing the capabilities of a UVS, the algorithm could be immersed into environments that train it to over-ride some pre-ordained taboo on the basis of some learnt behaviour.