The development of new military hardware brings new moral questions with it. While the Foreign and Commonwealth Office stated in December 2017 that it ‘does not possess fully autonomous weapon systems and has no intention of developing them’, a 2018 report from Drone Wars found that the UK is actively developing lethal autonomous weapons systems (LAWS). The first LAWS to be developed will likely be unmanned aerial vehicles (UAVs). Initially, artificial intelligence (AI) can be used for UAVs to pilot themselves, until AI reaches a point where UAVs will be able to identify, select and engage targets themselves.

In the future it is possible that more complex and consequentialist moral systems could be applied to LAWS, but AI and machine learning technology is not there yet.

With these ongoing developments in the field, it is pertinent to examine what the best approach would be for the moral use of LAWS. A first step towards this would be to consider a legal framework within which they may operate. US Department of Defense (DoD) Directive 3000.09 comes closest to a codified document that provides a framework for how LAWS should be deployed. It states that LAWS must ‘complete engagements in a timeframe consistent with commander and operator intentions’, demonstrating that the DoD wishes to limit the scope for LAWS, preventing them from becoming independent enough to choose their own engagements.

Patrick Lin, George Bekey and Keith Abney classify LAWS as artificial moral agents or AMA, whose ‘concern for safety’ in their actions require them to have the capacity to make moral judgements. If LAWS engage a target, then they must be functionally moral. All that is required for a functional morality system is the capability to respond to a scenario in which an act has a concern for safety. This is in contrast to an operational morality system, which requires the evaluation of the consequences of those choices. For current systems and those of the near future, there is no need, or ability, to have operational morality.

It could be the case that operational morality may never be required. Wendell Wallach and Colin Allen noted in their 2008 book Moral Machines that moral and ethical behaviour may be more than mathematics; can engineers design a moral framework or is morality also tied into a sensory experience that is not applicable to autonomous systems? A complete moral framework would also have to deal with emotion, psychology and sensation. It is practical enough that any system that is perceived to make moral decisions need only appear to also have functional morality.

A Rules-Based Approach

With this consideration for a functional system, the moral framework which seems most applicable to LAWS is a rules-based one (also known as deontology in moral philosophy). For any rule to be considered ethical in a rules-based system, it must always produce a moral outcome. This works especially well with systems that make limited choices. There is no need to foresee or evaluate the consequences of an action as there would be with an operational morality system.

For consequentialists, LAWS would need an operational morality system, and would have to evaluate the outcomes of actions they take to determine which outcome would yield the greatest good for the greatest number.

Systems with limited choices that follow a rules-based framework may not appear to be especially moral, it is their designers who enable the system to mimic a moral framework.
This is a top-down framework. LAWS using a rules-based system would be able to kill morally as they are only functionally moral. The designers who turn this set of rules into an algorithm are the real moral agents, as obeying those rules is what makes the LAWS moral.

For example, a landmine is an indiscriminate weapon that can kill as a response to an action (being stepped on) and in these cases there is no moral responsibility for the landmine, the moral responsibility lies with the person who set it.

There is then no reason to suggest that systems following a similar and simplistic set of instructions would be able to kill, and if they were not indiscriminate could also be used by an operator to kill morally even though the robot is not a moral agent. The DoD make it clear in Directive 3000.09 that military commanders will continue to be responsible for the actions of their subordinates, including weapons systems carrying out orders.

A problem with applying the deontological approach is moral slavery: LAWS are expected to obey their orders. Lin, Bekey and Abney write: ‘that would collapse (from a deontological perspective) all questions about their ethics into simply questions about the ethics of the military commander, and mutatis mutandis for any other use of autonomous robots as slaves’. It is unclear if, practically, this is a problem, although it implies that LAWS can never be ethical. Would people believe that autonomous systems are slaves? Moral slavery of autonomous systems is best left as a philosophical, rather than practical and security, problem for now.

Consequentialist Counterpoint

Another theory for how to use automated systems morally is the consequentialist approach. For consequentialists, LAWS would need an operational morality system, and would have to evaluate the outcomes of actions they take to determine which outcome would yield the
greatest good for the greatest number.

Depending on the level of rigour applied, this approach is much less practical than a rules-based approach. The reason for this is the great difficulty in predicting consequences of actions. It is far more foreseeable that any rules-based approach would allow for LAWS to not break the rules and therefore remain moral. A working paper submitted by the US delegation to the Group of Governmental Experts on Lethal Autonomous Weapons Systems proposed that properly tested LAWS may even be more compliant with international humanitarian law (IHL) than human soldiers, because the speed of their decision making is significantly faster than humans’.

Systems with limited choices that follow a rules-based framework may not appear to be especially moral, it is their designers who enable the system to mimic a moral framework

If a consequentialist theory is applied to operational morality, it would be difficult for a system to know what outcome would create the greatest good for the greatest number. There is a fallacy in all of consequentialism, of how far one ought reasonably to be able to foresee the consequences of their actions.

For humans, Jeremy Bentham thought that an adequate way to apply consequentialism was to do what is generally seen to be the right thing, and for one to focus on the likely outcome. This ‘Rule of Thumb’ approach bypasses the issue of foreseeability and the similar problem of repetition, where consequentialists must continually rework variations of near identical moral issues to always optimise their actions. This leads to the reality of humanity in consequentialism; the ability to decide if something is, practically, morally right or wrong is a uniquely human trait. One must make decisions based on their own life experience, on what is foreseeable, and the likelihood of those events transpiring. It may be theoretically possible to always make decisions that causes the greatest good, but impossible to do so in practice. It is likely that 19th century moral philosopher Henry Sidgwick would too reject to the idea of LAWS using a consequentialist system, writing that the ‘complexity of [consequentialist] calculations render it likely to lead to bad results in [the wrong] hands’.

The second problem with consequentialism is the phenomenon of scapegoating, where the theory may justify the killing of an innocent to quell the mob. While these examples are theoretical (for academics there are endless mobs milling around somewhere) for LAWS this is a foreseeable scenario.

It is not unforeseeable that this very problem would arise in unstable and volatile regions where LAWS may be deployed. While it might be excusable politically or militarily to engage, even if that were to risk the lives of civilians, it would not be compliant with IHL. An ethical and legal problem that will need to be addressed is to what extent LAWS will be able to operate in cases of self-defence. Where humans may have more options to repel an attacker, these options are not available to LAWS.

If the choice is between shoot and do not shoot, then it is likely that in some cases choosing to shoot will be the consequentialist choice. This is a problem with using LAWS as a first-response choice, as well as with the application of consequentialism to them. The uneasy answer is that choosing to shoot satisfies a consequentialist reading; if consenting to the killing of innocent people saves more than not killing, then there could be no objection. Consequentialists would not see it as a problem for autonomous weapons systems to have a moral code that could justify the killing of civilians.

If one took the opinion that LAWS had to have operational morality to kill, that would put a hard stop on their deployment for the foreseeable future, but it is unlikely that this stance will be taken.

A Rules-Based Approach… For Now

For the foreseeable future, having a series of practical rules to guide the actions of LAWS would allow them to act ethically, even in their lethal capacity. Although these restrictions may stymie their capacity, it is important that the UK continue its endeavour to pursue ethical and justifiable military decision making. It is not worth sacrificing legitimacy for a military advantage.

Properly tested LAWS may even be more compliant with international humanitarian law than human soldiers, because the speed of their decision making is significantly faster than humans’

In the future it is possible that more complex and consequentialist moral systems could be applied to LAWS, but AI and machine learning technology is not there yet. A consequentialist would allow LAWS to make more difficult decisions, and therefore permit them to operate in a broader range of operational spaces. But if the actions they make are not justifiable, then they will only be doing more harm than good in the long run.

Tom Ascott
Tom is a Digital Communications Officer at RUSI.

The views expressed in this article are the authors’ and do not represent those of RUSI or any other institution.