AI and Ethics (2023) 3:281–293
https://doi.org/10.1007/s43681-022-00168-2

ORIGINAL RESEARCH

Meaningful human control of drones: exploring human–machine teaming, informed by four different ethical perspectives

Marc Steen[1] · Jurriaan van Diggelen[1] · Tjerk Timan[1] · Nanda van der Stap[1]

Received: 11 January 2022 / Accepted: 26 April 2022 / Published online: 18 May 2022

Abstract

A human-centric approach to the design and deployment of AI systems aims to support and augment human capabilities. This sounds worthwhile indeed. But what could this look like in a military context? We explored a human-centric approach to the design and deployment of highly autonomous, unarmed Unmanned Aerial Vehicle (UAV), or drone, and an associated Decision Support System (DSS), for the drone's operator. We explore how Human–Machine Teaming, through such a DSS, can promote Meaningful Human Control of the drone. We use four different ethical perspectives—utilitarianism, deontology, relational ethics and virtue ethics—to discuss different ways to design and deploy the drones and the DSS. Our aim is to explore ways to support and augment the operators' capabilities.

Keywords Human-centric · UAV · Drone · Responsible · Artificial intelligence · Human–machine teaming · Meaningful human control · Decision support

1 Introduction

It would be an understatement to say that there is a lively debate regarding the design and application of Artificial Intelligence (AI) systems, and, more broadly, about the role of AI systems in society. Multiple authors have discussed the harms that the deployment of AI systems can do to justice, conviviality, and privacy [e.g., 1–4].

Furthermore, a growing group of people seem to agree on the need for a 'human-centric' approach to the design and application of AI systems. In their *Ethics Guidelines for Trustworthy AI*, the European Commission's *High-Level Expert Group on AI*, e.g., uses the term 'human-centric' to refer to an approach that 'strives to ensure that human values are central to the way in which AI systems are designed, deployed, used and monitored, by ensuring respect for fundamental rights, including those set out in, e.g., the Charter of Fundamental Rights of the European Union' [5: 37]. They aim to use AI systems to *empower people*: to support and augment human intelligence and human capabilities; not to replace people, their dignity or autonomy, or to corrode human faculties.

This view leads us to a key question about a 'human-centric' approach to AI: How can we organize the *collaboration* between people and AI systems? This question deals with Human–Machine Teaming (HMT): the organization of collaboration between people and machines, as teammates, in which they share and coordinate tasks [6] and 'responsibilities'. *Responsibilities* is between inverted commas because it is debatable whether AI systems can have responsibilities, or not. We assume that *only* people can have moral agency and moral responsibility, and that machines *cannot* [7–11]. Moreover, most people would agree that *even if* machines can have some kind of responsibilities, in the (far) future, these responsibilities would be rather different in kind compared to the responsibilities that people typically have.

Below, we will explore various ethical perspectives which can be used to organize HMT and promote MHC. We will explore different application designs for a highly

✉ Marc Steen
marc.steen@tno.nl

Jurriaan van Diggelen
jurriaan.vandiggelen@tno.nl

Tjerk Timan
tjerk.timan@tno.nl

Nanda van der Stap
nanda.vanderstap@tno.nl

[1] TNO (The Netherlands Organisation for Applied Scientific Research), The Hague, Netherlands

autonomous, *unmanned* aerial vehicle,[1] or *drone*, in a military context, and a Decision Support System (DSS) that mediates the drone's operator's interactions with the drone (for an exploration of other types of drones, see [12]).

In a military context, and for such autonomous systems, many have proposed the requirement of Meaningful Human Control (MHC) [13–15]. This requirement typically pertains to systems that can use *lethal force* and is meant to safeguard that their operators can exercise MHC over these systems, e.g., *armed* drones. Whereas this requirement is important, we have noted, that mentioning the notion of autonomous systems that can exercise lethal force can have very polarizing effects in a discussion. Some people believe that we should build such systems, e.g., for strategic reasons, and rely that MHC can be implemented effectively. Others believe that MHC is too difficult to implement, and therefore reject such systems wholesale. Still others advocate putting a ban on such autonomous, armed systems for principled reasons.

Rather than join this polarized discussion, we will focus our exploration on *unarmed* drones: drones with only sensors and radio communication, which are *incapable* of exercising lethal force, but nevertheless function within a morally sensitive context.

Such drones will typically be deployed for reconnaissance tasks, to support troops and commanders to make decisions. It must be understood, however, that these decisions *can* lead to actions that *can* lead to the use of lethal force. *Unarmed* drones can, e.g., be involved in the process of *target acquisition*, that is: the identification and evaluation of potential targets for defensive or aggressive actions. In other words, also *unarmed* drones can become implicated in *armed* activities and the use of lethal force [16]. Nevertheless, our intention of our focus on *unarmed* drones is to put our study *somewhat* apart from debates on the design and use of Lethal Autonomous Weapon Systems (LAWS) [17–22].

We will envision various ways to organize HMT to promote MHC. We will assume that the drone conducts tasks related to reconnaissance autonomously. In addition, the drone presents outputs of its sensors, together with additional information to an operator (HMT), via a DSS. The operator then uses their human capabilities for judgement, and decision-making (MHC).

Our paper's added value is fourfold. First, we explore how HMT can promote MHC. We propose this as a supplement to the body of research into MHC that focuses on 'programming' ethical reasoning into the system (more on that below). Second, we focus on *unarmed* drones, which are incapable of using lethal force. This approach enables us to shed light on some moral issues of military systems that would otherwise be overshadowed by the polarized debate on autonomous lethal weapon systems. Third, we follow a pluralistic approach to ethics in that we turn to four different ethical perspectives: utilitarianism, deontology, relational ethics, and virtue ethics. We propose that this can help to move beyond the default focus on utility-based reasoning (which is dominant in, e.g., computer science) and on deontology-based reasoning (which is dominant in, e.g., law). Fourth, we make our study as practically relevant as possible: we use a realistic scenario and present sketches of what the system *could* look like in practice. Our study is based on collaborations with people who went on military missions, and who shared their experiences with us. In that sense, our work is complementary to research in which the MHC is studied outside a specific application context.

Our paper proceeds as follows: we first discuss the concepts of HMT and MHC. Then we introduce a fictional scenario, which we use for our exploration. Then follow four sections, in which we discuss four different ethical perspectives, which we use to envision four different ways to organize HMT and to promote MHC. We will discuss the benefits and limitations of each ethical perspective. Finally, we discuss several implications of our exploration.

1.1 Organizing human–machine teaming to promote meaningful human control

The requirement of MHC is meant to safeguard that human perception, human judgement, and human decision-making are integrated in the control of the system in such ways that the people involved 'should ultimately remain in control of, and thus morally responsible for, relevant decisions about (lethal) military operations' [23: 1]. The term *meaningful* excludes ways that involve too much human control, e.g., where people need to micro-manage the system, or too little human control, e.g., where people unthinkingly follow the system's actions.

In the context of MHC, a recent article by Shneiderman [24] is relevant. He proposed to view *computer automation* and *human control* not as opposites on one axis, where an increase of one results in a decrease of the other, but as two perpendicular axes. This view provides ways to productively combine high computer automation and high human control; see Fig. 1. We can visualize the ambition to put highly autonomous functionalities in the drone as a move from left to right in Fig. 1: toward high *computer automation*, which comes with a warning *not* to overshoot into excess (right), e.g., where morally sensitive tasks are delegated to the drone, instead of giving them to human operators, who are better able to do these tasks. Likewise, the ambition to give soldiers a decision support system, to enable them to

[1] We focus on flying drones that 'only' have cameras and other sensors, typically used for reconnaissance or surveillance; *not* on drones with weapons, and *not* on drones that can carry loads.

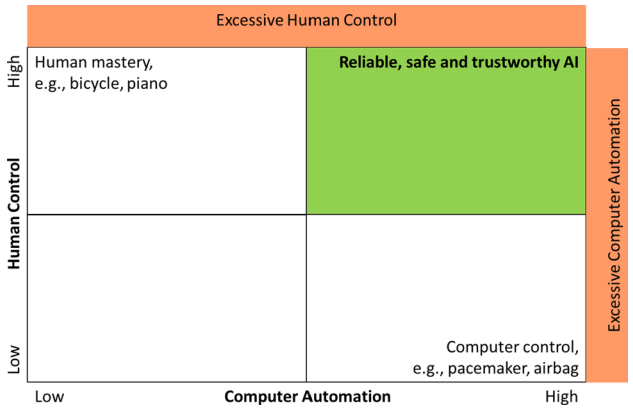

Fig. 1 'Reliable, Safe, and Trustworthy' AI requires appropriate levels of computer automation and human control [adapted from 24]

integrate their perception, judgement and decision-making in the system's control, is visualized as a move from bottom to top: toward high *human control*, which comes with a warning *not* to overshoot into excess (top), e.g., where operators perform too many tasks or repetitive tasks—tasks which the system can do better instead.

Our focus on highly autonomous drones requires us to focus on collaboration between the operator and the drone. We propose that a Decision Support System (DSS) can support this collaboration. It can collect, analyse, and present information in such a manner, e.g., as 'red flags' for specific risks, that operators can take this information into account and can combine it with their professional perception and judgement, to make better decisions. The design of the HMT process, including the interaction between operator and this DSS, is critical. If the system, e.g., (accidentally, unintentionally) encourages operators to consistently disregard its output, or to consistently follow its information, it would undermine its goal of *supporting* decision-making. Indeed, there would be no very little *meaningful* human control. This explains our focus on the ways in which the interaction between operators and the DSS is designed.

Our current aim is to explore how organizing HMT can promote MHC. This may very well diverge from other researchers. Others may understand MHC as pure teleoperation, or as 'programming' ethics into the machine. In contrast, we understand MHC as organizing HMT in ways that enable the people to have MHC; this will typically involve designing procedures for ways to interact with the system, or a specific user interface design to enable operators to exercise MHC.

1.2 Scenario: an unarmed, surveillance drone

The context of our scenario is a (fictional) state in which a separatist, insurgent group uses terrorism against the population, involving serious breaches of security, justice, and peace. The state's government asked for support in the context of the United Nations. In response, the Security Council issued a (fictional) resolution that mandates a group of countries to organize a peace mission to support the government in restoring security, justice and peace. The mission's objectives include preventing conflicts and escalations of violence, and supporting the government's administration, rule of law, and law enforcement. This is a *non-international armed conflict*, in which Common Article 3 of the *Geneva Conventions* is key.[2] This rule prohibits the use of force against anybody who does not participate in the conflict (*non-combatants*); these include citizens, and also soldiers who no longer participate in the conflict (*hors de combat*), e.g., because they put down their weapons, or became injured or ill. These people need to be treated humanely and offered care, without discrimination based on, gender, religion, culture, etc. Moreover, for the task of reconnaissance, the following principles are critical: *distinction*, one needs to distinguish between combatants and non-combatants; *proportionality*, one needs to take into account and weigh the exercise of force in relation to the goal one aims to achieve; and *precaution*, one needs to carefully assess issues, both in the preparation and in the execution of actions that involve the use of force.

The scenario involves a team of two soldiers and one drone. They use the drone for reconnaissance and surveillance, to gain intelligence and 'situational awareness'. The drone is highly autonomous; it flies in designated areas, and avoids other areas; given its specific task, e.g., to survey a series of specific locations or targets, it calculates its route; it uses its sensor data to control and modify its flying patterns, e.g., to fly around a building to get a view from multiple angles; and, in case of a communication malfunction, e.g., when its radio is jammed, it flies back to the launch location before it runs out of power. One soldier is team leader, and responsible for communication with headquarters. The other soldier operates the drone, e.g., monitors the output of its cameras and other sensors; this happens via the DSS, which runs on a tablet-like device.

The deployment of highly autonomously functioning drones brings a series of requirements and challenges, notably: the requirement that human operators are able to understand the system; the limitations or biases in human perception, cognition or judgement; and the management or risks associated to delegating tasks to machines [25].

[2] We discuss *legal* matters only briefly, to provide context. Our exploration focuses on *ethical* perspectives to organize HMT and promote MHC—*not* on legal matters. One remark on Common Article 3; this rule applies out of customary international law.

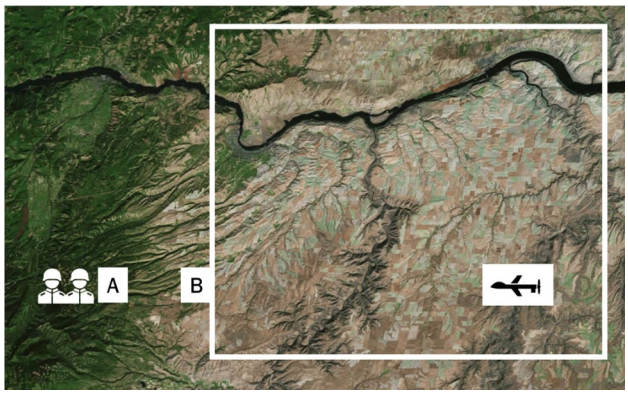

Fig. 2 Scenario: Two soldiers and a drone, tasked with reconnaissance

The team's task is to monitor a specific area, see Fig. 2. In military vocabulary, this is the process of *detect, recognize, and identify* (DRI), which can refer to buildings, objects or people. First, they *detect* a building, object or person that may be of interest; then they attempt to *recognize* it; then they need to *identify* it. Here, the principle of *distinction* is key. Based on the drone's output, an operator assesses whether the people in the picture are combatants or non-combatants. They can also use other data sources, and their own perception and judgement, and use categories like 'enemy troops, threat', 'enemy troops, no-threat', 'non-combatants, important', or 'non-combatants, unimportant', own troops, important' or 'own troops, unimportant'. Such acts of classification can have significant, real-world consequences; they may lead to actions, like 'relay to headquarters for further investigation', which may lead to the choice to use (lethal) force.

The drone provides data to the DSS, which enables the operator to assess the situation and propose a decision. The decisions are made by the team leader. We envision a system that can collect additional information from, for example, databases, and make calculations, for example, about plusses and minuses (see below). Furthermore, we envision that the system presents this additional information on-screen, to support operators in their perception and judgement. So there are three information components that go into the decision support system: the drone's sensor data; additional information from external databases; and human perception and judgement.

Crucially, drones not only matter as tools for surveillance. They also matter in terms of shaping the relationship of the UN mission to the local population. The local population may be very wary about the domestic conflict and about external parties intervening. A UN mission, and the drones they use, can be framed by both sides of the conflict as an 'enemy'. Flying too low over a village, e.g., can disturb the population and can be perceived as threatening indeed—even if the drone is unarmed.

2 Four ethical perspectives

In the next four sections, we will turn to four different ethical perspectives to envision different options to organize HMT and promote MHC. We propose that all four perspectives have value; notably, we will argue that they can be used in parallel. We will envision different options to design and deploy the drone and the associated DSS. We chose to make these options as practical as possible, e.g., with sketches for the user interface. Please to note that these are used for the sake of illustration; they are, by no means, meant to be implemented as such.

Moreover, our explorations can best be understood as thought experiments—as *Trolley Problems*, if you like, but with four variations, with more variables, and with more open ends. We sketch situations with rather broad strokes, to explore what these situations *could* look like (and leave many questions unanswered).

The DSS is meant to preserve the operator's moral agency in the sense that it enables them to exercise responsibility. It does that by enabling two conditions for responsibility, namely information and control [26: p. 12]. *Information* refers to the requirement that operators can access information about the current state of affairs and about possible future states of affairs (foreseeability). *Control* refers to the soldiers' freedom of action, meaning that they can interpret the system's output and use their own judgement. In that sense, the information collected by the drone and the decision support system can be understood as 'moral crutches', a term that Haselager and Mecacci [27] introduced to refer to using AI systems as tools to support and enhance *people's* moral agency—rather than try to put ethics *into* the AI system.

We chose the following four ethical perspectives: *utilitarianism, deontology, relational ethics, and virtue ethics*.[3] For each perspective, we first discuss its particular assumptions and commitments, and then envision the following: the functionality of the drone, where we aim for it to behave autonomously as much as possible (not too much, not too little); and a DSS that mediates the HMT (between soldiers and drone) and enables the soldiers to exercise MHC. The DSS also interfaces with *communication* functionalities, e.g., to retrieve additional information or communicate with people at Headquarters. We follow Schneiderman's [24] proposal to

[3] Our choice is consistent with, e.g., Van de Poel and Royakkers' (2011: 77–105); our discussion of *relational ethics* is, however, somewhat broader than their discussion of *care ethics*.

combine *optimal* computer automation and optimal *human control* to promote reliability, safety and trustworthiness. The autonomy of the drone and the DSS then complement each other.

Each of the four perspectives focuses on different aspects of the *same* situation, and thereby yields different starting points for envisioning HMT and MHC. Our exploration is inspired by Alfano's [28: 14–18] discussion of five key concepts in moral philosophy: patiency[4]; agency; sociality; reflexivity; and temporality. He discusses the relative weights of these different concepts in different ethical perspectives (the same four as we discuss). Based on this, we would like to propose that each ethical perspective can help to draw out different relevant aspects of the same situation:

- A *utilitarianist* perspective focuses on *patiency*, the *potential harms* to other actors of a decision or action; it also deals with how consequences play our over time (*temporality*), and the effects on interactions and relationships (*sociality*).
- A *deontologist* perspective draws attention to *human agency*; it foregrounds and privileges ways to respect, protect and enhance people's abilities to exercise autonomy; and with its focus on rationality, it also emphasizes *reflexivity*.
- Relational ethics (probably unsurprisingly) focuses relationships and interactions between people (*sociality*), and how these change over time (temporality); it also looks at potential harms to these relations (*agency* and *patiency*).
- Virtue ethics focuses on processes of learning and development over time, how people cultivate relevant virtues (*temporality*); it also highlights human *agency* and *sociality*, how people can find ways to flourish and live well together.

2.1 A utilitarian perspective

The *outcomes* of choices and actions are central in a utilitarian approach. For each choice or action, potential positive and negative outcomes are assessed. Jeremy Bentham, a proponent of this perspective, understood outcomes in terms of people's positive or negative experiences; as 'pleasures' and 'pains'. In our case, different options for carrying out the mission can be assessed in terms of positive and negative outcomes to the mission. The best option would be the one with the most or largest benefits, or the least or smallest downsides.[5] People who are involved in the design and application of AI systems typically feel attracted to this type of reasoning because utility functions are common concepts in problem solving algorithms, such as constraint satisfaction, planning, and reinforcement learning [e.g., 29].

Despite their appeal, utilitarian approaches have received criticism, often illustrated with examples like the surgeon who chooses to sacrifice one patient to harvest organs to save five other patients who would die without these organs. Such examples draw attention to a key challenge of utilitarianism: having to compare and weigh *incommensurable* values: values that 'cannot be reduced to a common measure' [30], such as 'safety of own troops' and 'hearts and minds of local population'. Another challenge concerns the scoping of the problem; which outcomes are taken into account, and which are left out of the sum? Think of Peter Singer's [31] example of a child from drowning in a shallow, nearby pond. Most people will rescue this child. But they hesitate to rescue children from malaria or starvation in a distant country—probably because they are less directly visible. We can think of remoteness both in space and in time. We would need to also take into account future outcomes, e.g., the experiences of next generations in the case of climate crisis.

For our case, we will focus on the pros and cons that occur in the area of the mission in which the drones are deployed, and on a timespan of several days after the drone's deployment.

2.1.1 Drone

Typically, we want to design and program the drone in such a manner that it can do as much as possible autonomously. The drone has cameras and other sensors to collect data and it uses software to interpret these data. We assume that it can create a list of possible actions and calculate the utility, or u, for each: the sum total of pros and cons for that action. We also assume that the drone has access to measures that assess parameters like: the safety of own *troops* (t); the local *population's* 'hearts and minds' (p); and the drone's abilities to safely *return* to base (r). We can add weights to these parameters, so that the troops' safety has much weight (5), the population's sentiments average weight (3), and the drone's ability to return home little weight (1): $u = \underline{5}*t + \underline{3}*p + \underline{1}*r$. These weights will need to correspond to the mission's *Rules of Engagement* and various cultural and ethical concerns and values of both the military's home country and the local country.

[4] Patiency refers to being on the receiving end of some action ('undergoer'); it is the opposite of agency.

[5] We ignore the difficulty of making such comparisons. Comparing option A, which is likely to have two large positive outcomes and one small negative outcome, with option B, which is likely to have only one small positive outcome and two large negative outcomes, is relatively easy. We will prefer A over B. Comparing action P, with one positive outcome and one negative outcome, with design option Q, which has one *other* positive outcome and one *other* negative outcome, however, is much harder.

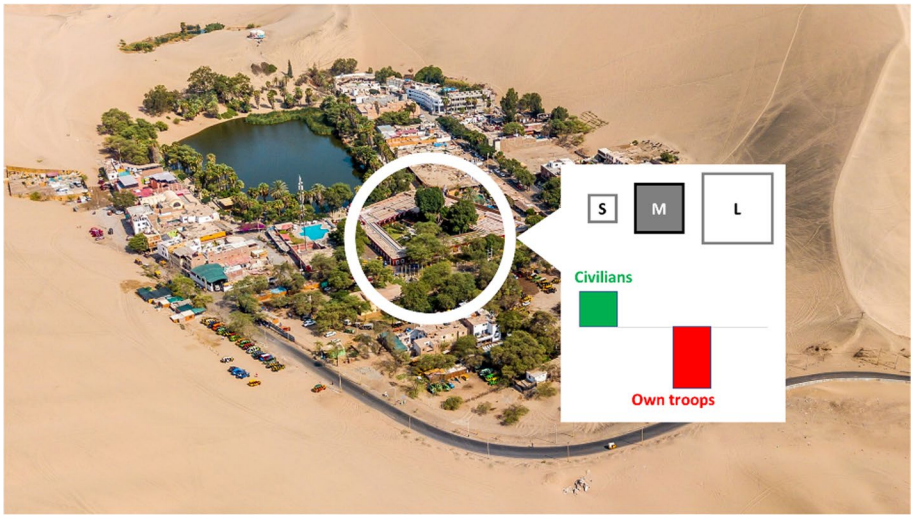

Fig. 3 Decision support system for a utilitarian approach, with green and red bars (plusses and minuses), which operators need to evaluate, and controls for the scope of the assessment (S, M, L)

Fig. 4 Decision support system for a deontological approach, with a pop-up that refers to relevant duties and rights, and a suggestion to communicate and consult with a relevant commander

Now, suppose that the drone, at one moment, must decide between *flying at high altitude*, which is okay for own troops (2), neutral for the population's sentiments (0); and better for the drone's ability to return to base (2) or *flying at low* altitude, which is okay for own troops (2), worse for the population's sentiments, in that it may upset them (-2), and slightly worse for the drone's ability to return to base (1), it would calculate that $u = 12$ for flying high ($5*\underline{2}+3*\underline{0}+1*\underline{2}$), and that $u = 5$ for flying low ($5*\underline{2}+3*\underline{-2}+1*\underline{1}$). The drone will then 'choose' to fly at high altitude.

This example is a gross simplification; in reality, there will be many more parameters, and it will be challenging to determine appropriate weights and reliable values for each parameter. Moreover, such algorithms must deal with *uncertainty* about expected outcomes, or they may need to deploy more complex, non-linear utility functions. Displaying uncertainty, e.g., giving certainty levels for specific likelihoods, can, by the way, support operators in using their judgement and discretion.

2.1.2 Decision support system

Choosing between flying at high or low altitude is a relatively straightforward decision to make (but certainly not trivial in a military context), and it seems reasonable to delegate such decisions to the drone. (Please note that, for MHC, and thus in all four scenarios that we discuss, morally salient decisions are made by people.) Let us imagine a system that supports the soldiers to make decisions according to a utilitarian approach; see Fig. 3 for a schematized interface example.[6]

This interface shows an object that the drone was unable to classify properly, and some information that the human operator can use to make decisions. Based on interpretations of the data that were collected by the drone, the system

[6] Photo by Willian Justen de Vasconcellos (https://unsplash.com/photos/HfLYdUePGyc); also for Figs. 4, 5, 6.

Fig. 5 Decision support system for a relational ethics approach, with information on relevant actors, their functions and relationships, from external databases (The photo of a non-existent person, from thispersondoesnotexist.com)

Fig. 6 Decision support system for a virtue ethics approach, with potentially relevant virtues for the situation at hand and, if available, an exemplar (with exemplary behavior)

assesses that the object is associated with positive outcomes for local civilians, e.g., to protect their safety and 'hearts and minds' (small green bar, positive) and can pose rather large risks to the own troops' safety (large red bar, negative). The interface also shows that a medium-sized scope was used to make this assessment, several kilometers geographically and several hours chronologically. Alternatively, the operator may click S or L to make new calculations using a smaller scope (e.g., less than a kilometer; less than an hour) or a larger scope (e.g., up till 100 km; up till 48 h), respectively. The latter assessment is likely to include more uncertainty.

The colored bars, the labels on these bars, and the selection of scope size are ways to promote transparency of the DSS. The commanding officer may choose to further investigate the situation before making a decision and can use the labels on the bars to guide such investigations. Moreover, such information promotes their accountability; they are better able to explain and justify how decisions were made, based on which information.

The concept of MHC is obvious in the choice to *not* let the drone make decisions about target acquisition, but to put an operator 'in the loop' for human judgement. In isolation, a drone would 'merely' calculate pros and cons, whereas with this way of organizing HMT, with the DSS, human operators are enabled to apply their discretionary competences and make better decisions.

2.2 A deontological perspective

Alternatively, we can use a deontological perspective, which puts human dignity and human autonomy center stage. It starts from the *good will* of moral agents and their *reverence for the moral law*—to borrow phrases from Immanuel Kant, a key figure in deontology. A deontological perspective would articulate duties that one has toward other people (moral patients), and rights of these other people, which would need to be respected and protected. In our case, this would entail evaluating different options by looking at how

they help to fulfill relevant duties and whether they help to respect and protect relevant rights. Thus, we first need to determine *which* moral duties and rights are relevant in our scenario.[7] While acknowledging the difference between *moral* duties and *legal* duties, norms from international law regarding conflict and war, can function as a starting point, notably the *Fourth Geneva Convention*, which deals with the duty for humanitarian protection of civilians in war zones—which ties in with human dignity, a key concept in deontology.

We will focus our discussion on two challenges: to determine which duties and rights are relevant in a specific situation, and to weigh or balance various duties and rights; and to deal with duties and rights in a military context, which is characterized by lines of command, obedience, and compliance.

Regarding the first challenge, it can be difficult for soldiers, in the midst of an operation ('fog of war'), to identify relevant duties and rights. Viewed in the abstract, most people would privilege the duty to protect human rights. In practical situations, however, soldiers will need to take into account and balance various duties and rights. Moreover, these can conflict; e.g., a duty to combat enemy troops and a duty to minimize *collateral damage*. In practice, soldiers are extensively trained to deal with diverse duties, especially because they typically need to make decisions under time pressure.

Regarding the second challenge, we need to appreciate that in a military context, with its chains of command, and the need for obedience and compliance, the concept of human autonomy—a key concept in deontology—is rather complex. On the one hand, soldiers must obey orders. On the other hand, they are expected to apply their discretionary competences in interpreting orders.

A deontological perspective also appeals to people who develop software for autonomous systems; Wallach and Allen refer to it as 'top-down morality' [32, chapter 6]. Proponents of a rules-based approach include, e.g., Thomas Powers [33].

2.2.1 Drone

First, we need to clarify that we assume that the drone is *not* a moral agent; it lacks the capacity to be of 'good will'. We will need to envision a watered-down version of deontology for the drone. We can design and program the drone in such a manner that it follows specific rules, e.g., to stay inside a certain area, e.g., the rectangle in Fig. 2, or to stay outside another area. Such seemingly trivial maneuvers can have real-word effects; military conflicts have escalated from trespassing borders.

Other rules that we can try to put into the drone would be *Rules of Engagement*—rules of a military organization that define the circumstances, conditions, degree, and ways in which specific military capabilities can or cannot be used. *Rules of engagement* typically refer *not* to goals or results, but to means and measures. We can program into the drone rules that forbid or restrict specific behaviors.

Furthermore, deontology typically depends upon moral agents' abilities for practical reasoning, to engage in ethical deliberation. Arguably, a drone does *not* have these abilities. A watered-down version of practical reasoning would be the application of 'if–then' rules. If *this* is the situation, then this or that rule applies. *If* you see civilians, *then* they need to be protected. If you see an eminent threat to own troops' safety, you propose a further examination of that object—which may lead to the operator or commander deciding to proceed to target acquisition.

2.2.2 Decision support system

The moral agents will need to apply their discretionary competences, to complement the drone's reasoning. For this, we can imagine a system with a user interface like the one in Fig. 4.

The system can support soldiers' discretionary competences by providing two sorts of information—which map unto the two challenges discussed above: the challenge of identifying relevant duties and rights; and the challenge of dealing with duties and rights in a context that demands obedience.

We can imagine that the system proposes duties or rights that are likely to be relevant, for example, the duty to protect civilians' rights from Article 4 of the *Fourth Geneva Convention*. It then remains the soldiers' call to assess whether these are indeed relevant, and to interpret these duties and rights appropriately. Furthermore, we envision functionalities to communicate with commanding officers. They can be contacted for real-time consultation; this can create a 'trail' for accountability, which can be useful if, later on, the decision-making process needs to be reconstructed.

2.3 A relational ethics perspective

Utilitarianism and deontology both emerged in the European Enlightenment (although their roots go back millennia) and their current interpretations are based on assumptions that people are independent individuals and that one needs to apply objectivity and rationality in ethics. Currently, we will discuss relational ethics, which can be understood as a

[7] Please note that there is overlap between *moral* duties and rights and *legal* duties and rights. To complicate matters, these categories can also conflict; e.g., avoiding to pay taxes may be legally acceptable, but is increasingly seen as morally inappropriate.

reaction to utilitarianism and deontology, and as a remedy to several of their limitations, notably the challenges of comparing *incommensurable* values, and of combining conflicting duties. Relational ethics draws from ethics of care and feminist ethics [34] and understands people as interdependent (not independent), as involved in various specific and concrete relationships (which are not entirely 'objective'), and as not only rational, but also emotional. Relational ethics typically focuses on specific, concrete situations, rather than on general principles or universal rules. In addition, ethics of care and feminist ethics will typically focus on qualities of relationships, and on the distribution of power.

Interestingly, relational ethics has been put forward in discussions of the design and application of AI, e.g., by David Gunkel [35, 36], Mark Coeckelbergh [37, 38] and Abeba Birhane [39].

If we apply this approach to our case, two challenges stand out. First, the challenge to *choose who to include*, and who to exclude; to identify 'relevant' actors. We encountered this challenge in the other perspectives, but it is even more difficult when it comes to relationships. A relational approach acknowledges that we, together with other actors, are embedded in a web of interdependencies. Second, the challenge of interacting appropriately with these actors. A relational approach focuses on people in *specific* and *practical* situations. Some view this as a drawback; one cannot make (simple, rational, utilitarian) calculations, one cannot follow (general, objective, deontological) rules. 'It depends'. Conversely, this focus can be viewed as an advantage. The right thing to do, indeed, *depends* on the specific and practical context and on the actors involved and their diverse relationships. In face to face situations, we 'read' all sorts of cues (moral sensitivity) and take these into account. When interactions are mediated, however, we need ways to provide these cues. What if the drone gathers information additional to what we see on the screen and presents this?

2.3.1 Drone

A relational ethics approach would design and program the drone in such a manner that it is capable to interact appropriately with relevant actors and stakeholders. This could entail that the drone changes its flying patterns to better relate to the people on the ground. The drone would, for example, fly around a religious building or cultural event, to signal to the people that it 'understands' that flying directly over this building or event would be wrong. The drone would convey a pro-social message to the civilians; comparable to the blue helmets that UN personnel wear to signal their mandate to protect civilians.

The drone could go into different 'modes of engagement', such as 'pro-social' or 'neutral observer', depending on the specific and practical context. This will, however, be rather challenging, because computers are notoriously bad at common sense and open-ended social interactions [40]. At the same time, people tend to respond emotionally to robots or ascribe emotions to them [41]. Moreover, a drone's pro-social behaviors may also have adverse effects, e.g., on the civilians' trust in it, when it first behaves pro-social and then changes its behavior. It needs to be noted that robots' behaviors can, and will, also be deployed to deceive [42].

Another aspect regards the type of information that the drone collects and presents. These data will pertain (also) to diverse relationships, which the human operators will need to interpret and take into account—thereby expanding their situational awareness. One can think of information that clarifies that a hospital is not only there to provide medical care, but may also be used to distribute food and fresh water. It has a place in a local community and mediates various relationships. This type of information enables operators to develop a more fine-grained awareness of the situation.

2.3.2 Decision support system

If soldiers are enabled to apply a relational perspective, they will need to understand not only the drone's images, but also the actors in these images, the relationships between these actors, and the wider context. It will not be possible to fully understand these actors, relationships, and context. The system can, however, provide information that the soldier can use to a *better* understanding. One can imagine a user interface that literally 'puts a face' on specific people. Take, for example, an object that is identified as a hospital. It would be possible to find the name of the hospital's manager, and a picture of them. See Fig. 5. In addition, the user interface could draw from data to add more information; the hospital plays roles in employment, and in food and water distribution. Damage to the hospital would imply also damage to employment, food security and water supply.

These examples (above) are presented as based on correct information. In reality, however, it may be difficult to test the information for correctness. Belligerents 'from the other side' may present themselves as citizens, or use citizens for camouflage. Correctly identifying people 'on the ground' as friend or foe is notoriously difficult. Of course, a relational approach cannot solve this problem. What it can do, however, is provide additional information, including potentially conflicting information, which the soldiers can take into account, as part of their discretionary competences. The added value of a relational approach is in broadening the scope of situational awareness, which could improve decision-making.

2.4 A virtue ethics perspective

Virtue ethics has its roots in ancient Greece, notably in Aristotle's *Nicomachean Ethics*. Since the 1980s there has been a growing interest in virtue ethics, sparked by publications by, for example, Filippa Foot [43] and Alisdair MacIntyre [44, first published in 1981]. A key aim of virtue ethics is to enable people to cultivate relevant virtues: dispositions to think, feel, and act according to virtues that are relevant for the situation at hand; to develop toward their potential (*telos*) and to live well (*eudaimonia*). It aims to promote human flourishing by creating societies in which people can live well together (*polis*). A proponent of this approach is Shannon Vallor; in *Technology and the Virtues* (2016) she advocated drawing from virtue ethics to design and apply technologies in ways that enable people to cultivate technomoral virtues, so they can collectively work toward 'a future worth wanting' (the book's subtitle). Vallor argues that 'technologies invite or afford specific patterns of thought, behavior, and valuing; they open new possibilities for human action and foreclose or obscure others' [45: 2]. When we design and apply technologies, we need to be mindful of how they influence what we can (not) think, feel and do. Others who wrote about virtue ethics in relation to technology design are, e.g., Coleman [46], Ess [47], and Tonkens [48].

Some of the virtues that Vallor discusses are similar to Aristotle's ('cardinal') virtues: courage; self-control; justice, and practical wisdom. Out of her list of virtues,[8] several are especially relevant to our current discussion: humility, 'know what we do not know'; compassion, 'compassionate concern for others'; Flexibility 'skilful adaptation to change'; civility, 'making common cause'; and perspective, 'holding on to the moral whole'. We can add several virtues from a list of military virtues that Skerket et al. [49] put forward[9]: obedience; loyalty; integrity; and perseverance.

For each virtue the aim is to find an *appropriate mean.* Please note that this 'mean' has nothing to do with average or normal—quite the opposite; it refers to an *excellent* expression of a virtue, in a particular context: not too much (excess) and not too little (deficiency). Courage, e.g., is the 'mean' between rashness (excess) and cowardice (deficiency). In a particular situation, you will need to find the appropriate mean for courage, depending on your abilities and on the context.

Imagine that you witness a person being attacked on the street. If you are a frail person, it would be courageous to stay put and phone the police; it would be reckless to intervene. However, if you are an athletic person and know how to handle such a conflict, active intervention would be courageous; it would be cowardly to stay put. You can *cultivate* courage by exercising courage and learning from your experiences. This requires effort and it may feel awkward. Over time, however, you can learn to align thinking, feeling, and actions. A person who has cultivated a virtue will express this virtue 'out of habit', at the right moment, in an optimal form, for the right reasons, and with the right feelings.

The main challenges are to enable moral agents to cultivate relevant virtues. We can understand this challenge as consisting of two elements: to determine *which* virtues are relevant in a given situation, and to find the *appropriate mean* for this virtue, given the situation.

2.4.1 Drone

We need to recognize that virtue ethics assumes that morality happens *within people*—not in machines. So, if we want to explore the idea of implementing virtue ethics into a drone, we will need to use our imagination and make several translations. What could a *telos* or *eudaimonia* look like for a drone?[10] Loosely following Wallach and Colin [32] and Wallach and Vallor [50], we speculate that a drone's *telos* involves contributing to the mission at hand, and that a drone's *eudaimonia* involves collecting and presenting information that its operators find clear, trustworthy, and useful.

Crucially, the cultivation of virtues happens over the course of time; it entails learning from past experiences, active reflection, and adjusting one's thinking, feeling, and actions to better align them to one's *telos* and *eudaimonia*. If we translate this to the drone, it means that it will need to keep track of its contributions to various missions, and learn over time. Moreover, it will need to learn also from other drones; in analogy to how people learn from others, typically from so-called *exemplars*: people who express or exemplify virtues in an exemplary manner.

Moreover, we can understand the attempt to implement virtue ethics as an attempt to implement a type of reinforcement learning that combines a deontological, top-down, rules-based approach with a utilitarian, bottom-up, calculation-based approach [32, Chapter 8, 50]; the system starts with following general rules, and then tries out specific

[8] Vallor discusses the following technomoral virtues: Honesty; Self-Control; Humility; Justice; Courage; Empathy; Care; Civility; Flexibility; Perspective; Magnanimity; and Technomoral Wisdom.

[9] Skerker et al.'s list of virtues: Justice; Obedience; Loyalty; Courage; Wisdom; Honesty; Integrity; Perseverance; Temperance; Patience; Humility; Compassion; Discipline; and Professionalism.

[10] We realize that this can come across as contradictory: first we state that a virtue ethics happens within people—*not* in machines; and then we try to imagine *how virtue ethics could happen in a drone*. We believe we can do this, as part of the thought experiment that we carry out in these four sections.

Table 1 Different ethical highlight different elements of situations, and offer different starting points for the design and application of a highly autonomous drone and a Decision Support System (DSS)

	Highlights	Highly autonomous drone	Decision Support System
Utilitarianism	Potential harms	Calculate plusses and minuses of outcomes	Deal with incommensurable values
Deontology	Human autonomy	Follow general rules; duties and rights	Deal with conflicting duties or rights
Relational ethics	Relationships	Interact more socially, e.g., with citizens	Deal with context and specifics
Virtue ethics	Reflection, learning	Combine bottom-up and top-down 'learning'	Cultivate relevant professional virtues

actions and, over time, learns about the plusses and minuses of these actions, and optimizing its behavior.

2.4.2 Decision support system

If we shift the locus of morality (back) to people, we are (back) on firmer ground to apply virtue ethics. The DSS can support its operators to cultivate relevant virtues. This may happen in two main ways; see Fig. 6. The system can support operators to cultivate specific virtues, typically through trial and error, probably based on data concerning specific operators' current virtues and virtues needed in specific situations. Or the system can support operators to learn from others, notably from *exemplars*. We can imagine a system that presents several virtues that are likely to be relevant in the situation, as a reminder. In addition, it can show a specific case and exemplar that are relevant for the situation, e.g., 'Lieutenant Jones in *Peace Mission*, 2018'. Ideally, the operator knows this case, e.g., through training, so that the *exemplar* function as a role model. Practically, the system needs to access a library of cases and exemplars, and select a relevant case and exemplar.

In virtue ethics, it is critical that people can cultivate relevant virtues over the course of time. Any military operation has briefing moments before operations, and debriefing moments after operations. What is true for the other ethical perspectives, namely that deliberations and outcomes are likely to be discussed during such briefing and debriefing sessions (although, not necessarily in explicit utilitarian or deontological vocabularies), is especially true for virtue ethics—because of its emphasis on learning by doing and through reflection. Virtue ethics puts *practical wisdom* center stage; it functions as a master virtue to moderate and express other virtues [51].

3 Conclusions

Our exploration of four ethical perspectives has made clear that each perspective has its own distinct benefits and limitations. Our proposal is relatively modest and can be modified easily by other researchers: to put an appropriate amount of autonomy (not too much, not too little; see Fig. 1) in the *highly autonomous drone*, so it can behave as if it were a team member (HMT), and to give soldiers a Decision Support System (DSS), which presents the drone's sensor outputs, which the soldiers can use in combination with their professional judgement and deliberation, so that they can exercise Meaningful Human Control (MHC) over the drone. Moreover, our proposal is to combine the four different perspectives—as each offers distinct benefits—see Table 1:

- A utilitarian perspective highlights potential positive and negative outcomes of one's choices or actions, and thus can help to focus on potential harms. To an extent, software is able to calculate plusses and minuses, provided that it has reliable data, but runs into challenges when incommensurable values are at play, and when the analysis' boundaries are questioned.
- A deontological perspective highlights human autonomy and agency, and can help to identify and take into account relevant duties and rights. To an extent, duties and rights can be translated into software. This will, however, run into challenges when they conflict. Moreover, soldiers need to deal with conflicting duties and rights and use their autonomy in a context of obeying orders.
- Relational ethics highlights relationships between actors. It requires viewing each situation as specific and practical, and taking into account a larger context. Key challenges are: to determine which actors to include (and which to exclude), and to understand their interdependencies. Conversely, it can help to design and apply, e.g., drones in pro-social manners.
- Virtue ethics highlights people's abilities to cultivate relevant virtues, and to work toward enabling people to flourish—to live well together. It can help to design and us technologies in ways that support people to reflect and learn. It is hard to 'put' virtue ethics 'into' machines. We therefore explored creating a DSS that enables people to cultivate relevant virtues.

Each of these perspectives is potentially relevant for any given situation. Furthermore, we discussed ways in which one perspective's benefits can compensate for another perspective's limitations; e.g., a relational ethics' emphasis

on relationships between relevant actors can help to better assess pros and cons of a specific decision, or to better find a balance between conflicting duties. We thus speculate that designing a system that enables operators to combine these different perspectives may be an interesting way forward for the design of highly autonomous drones and an associated Decision Support System (DSS).

We acknowledge that our exploration raises more questions than we can currently answer. Further research is needed to study ways to combine the different perspectives in practical situations.

We can envision that people use the four ethical perspectives while preparing or evaluating a mission; the people involved can look at the situation from each of the four perspectives, and take time for each perspective and its implications. It may be worthwhile to experiment with the four perspectives in briefings, debriefings, and training programs. It remains to be seen, however, whether people, during an operation, in the heat of the moment, in the midst of action, are able to use the four perspectives simultaneously. Can they do that consecutively, e.g., of priority or relevance? Or do they need to do that parallel, e.g., by dividing the perspectives over different people? Regarding such practical applications, there is a range of questions that will need to be answered, like *who* prioritizes between these perspectives: an individual soldier, their commander, or somebody higher up in the chain of command?

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. O'Neil, C.: Weapons of Math Destruction. Penguin, London (2016)
2. Noble, S.U.: Algorithms of Oppression: Now Search Engines Reinforce Racism. New York University Press, New York (2018)
3. Turkle, S.: Reclaiming Conversation: The Power of Talk in a Digital Age. Penguin Books, London (2015)
4. Véliz, C.: Privacy is Power: Why and How You Should Take Back Control of Your Data. Transworld Publishes, London (2020)
5. High-Level Expert Group on Artificial Intelligence: Ethics Guidelines for Trustworthy AI. European Commission, Brussels (2019)
6. Peeters, M.M.M., et al.: Hybrid collective intelligence in a human–AI society. AI Soc. (2020). https://doi.org/10.1007/s00146-020-01005-y
7. Bryson, J.J.: Patiency is not a virtue: the design of intelligent systems and systems of ethics. Ethics Inf. Technol. **20**(1), 15–26 (2018)
8. Fossa, F.: Artificial moral agents: moral mentors or sensible tools? Ethics Inf. Technol. **20**(2), 115–126 (2018)
9. Van Wynsberghe, A., Robbins, S.: Critiquing the reasons for making artificial moral agents. Sci. Eng. Ethics **25**(3), 719–735 (2019)
10. Cervantes, J.-A., et al.: Artificial moral agents: A survey of the current status. Sci. Eng. Ethics **26**(2), 501–532 (2020)
11. Sparrow, R.: Why machines cannot be moral. AI Soc. (2021). https://doi.org/10.1007/s00146-020-01132-6
12. Wiseman, Y.: Autonomous vehicles. In: Global, I.G.I. (ed.) Research Anthology on Cross-Disciplinary Designs and Applications of Automation, Vol 2, pp. 878–889. Hershey (2022)
13. Ekelhof, M.: Moving beyond semantics on autonomous weapons: meaningful human control in operation. Global Pol. **10**(3), 343–348 (2019)
14. Umbrello, S.: Coupling levels of abstraction in understanding meaningful human control of autonomous weapons: a two-tiered approach. Ethics Inf. Technol. (2021). https://doi.org/10.1007/s10676-021-09588-w
15. Verdiesen, I., Santoni de Sio, F., Dignum, V.: Accountability and control over autonomous weapon systems: a framework for comprehensive human oversight. Minds Mach. **31**, 137–163 (2021)
16. Ekelhof, M.: Lifting the fog of targeting: "autonomous weapons" and human control through the lens of military targeting. Naval War Coll. Rev. **71**(3), 61–95 (2018)
17. Arkin, R.C.: The case for ethical autonomy in unmanned systems. J. Mil. Ethics **9**(4), 332–341 (2010)
18. Sharkey, A.: Autonomous weapons systems, killer robots and human dignity. Ethics Inf. Technol. **21**(2), 75–87 (2019)
19. Skerker, M., Purves, D., Jenkins, R.: Autonomous weapons systems and the moral equality of combatants. Ethics Inf. Technol. (2020). https://doi.org/10.1007/s10676-020-09528-0
20. Smith, P.T.: Just research into killer robots. Ethics Inf. Technol. **21**(4), 281–293 (2018)
21. Sullins, J.P.: RoboWarfare: Can robots be more ethical than humans on the battlefield? Ethics Inf. Technol. **12**(3), 263–275 (2010)
22. Umbrello, S., Torres, P., De Bellis, A.F.: The future of war: could lethal autonomous weapons make conflict more ethical? AI & Soc. **35**(1), 273–282 (2020)
23. Santoni de Sio, F., Van den Hoven, J.: Meaningful human control over autonomous systems: a philosophical account. Front. Robot. AI **5**, 1–15 (2018)
24. Shneiderman, B.: Human-centered artificial intelligence: Reliable, safe and trustworthy. Int. J. Hum.-Comput. Interact. **36**(6), 495–504 (2020)
25. Chavannes, E., Arkhipov-Goyal, A.: Towards Responsible Autonomy: The Ethics of Robotic and Autonomous Systems in a Military Context. The Hague Centre for Strategic Studies, The Hague (2019)
26. Van de Poel, I., Royakkers, L.: Ethics, Technology, And Engineering: An Introduction. John Wiley and Sons, Chichester (2011)
27. Haselager, P., Mecacci, G.: Superethics instead of superintelligence: Know thyself, and apply science accordingly. AJOB Neurosci. **11**(2), 113–119 (2020)

28. Alfano, M., Moral psychology: An introduction.: Cambridge. Polity Press, Cambridge (2016)
29. Russel, S., Norvig, P.: Artificial intelligence: a modern approach. 3rd edition, Pearson Education, Upper Saddle River (2002)
30. Hsieh, N.-H.: Incommensurable values. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Spring, Berlin (2016)
31. Singer, P.: The drowning child and the expanding circle. New Internationalist (1997). https://newint.org/features/1997/04/05/peter-singer-drowning-child-new-internationalist
32. Wallach, W., Allen, C.: Moral Machines: Teaching Robots Right from Wrong. Oxford University Press, Oxford (2008)
33. Powers, T.M.: Prospects for a Kantian machine. In: Anderson, M., Anderson, S.L. (eds.) Machine Ethics, pp. 464–476. Cambridge University Press, New York (2011)
34. Held, V.: The Ethics of Care: Personal, Political, and Global. Oxford University Press, New York (2006)
35. Gunkel, D.J.: Thinking otherwise: Ethics, technology and other subjects. Ethics Inf. Technol. **9**(3), 165–177 (2007)
36. Gunkel, D.J.: Perspectives on ethics of AI. In: Dubber, M.D., Pasquale, F., Das, S. (eds.) The Oxford Handbook of Ethics of AI, pp. 539–553. Oxford University Press, New York (2020)
37. Coeckelbergh, M.: Robot rights? Towards a social-relational justification of moral consideration. Ethics Inf. Technol. **12**(3), 209–221 (2010)
38. Coeckelbergh, M.: Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. Sci. Eng. Ethics **26**(4), 2051–2068 (2020)
39. Birhane, A.: Algorithmic injustice: a relational ethics approach. Patterns **2**(2), 100205 (2021)
40. Russell, S.: Human Compatible: AI and the Problem of Control. Allen Lane, London (2019)
41. Sparrow, R.: The March of the robot dogs. Ethics Inf. Technol. **4**(4), 305–318 (2002)
42. Danaher, J.: Robot Betrayal: a guide to the ethics of robotic deception. Ethics Inf. Technol. **22**(2), 117–128 (2020)
43. Foot, P.: Virtues and Vices: And Other Essays in Moral Philosophy. Blackwell, Oxford (1978)
44. MacIntyre, A.: After Virtue, 3rd edn. Duckworth, London (2007)
45. Vallor, S.: Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford University Press, New York (2016)
46. Coleman, K.G.: Android arete: Toward a virtue ethic for computational agents. Ethics Inf. Technol. **3**(4), 247–265 (2001)
47. Ess, C.: Ethical pluralism and global information ethics. Ethics Inf. Technol. **8**(4), 215–226 (2006)
48. Tonkens, R.: Out of character: on the creation of virtuous machines. Ethics Inf. Technol. **14**(2), 137–149 (2012)
49. Skerker, M., Whetham, D., Carrick, D. (eds.): Military Virtues. Howgate Publishing, Havant (2019)
50. Wallach, W., Vallor, S.: Moral machines: From value alignment to embodied virtue. In: Liao, S.M. (ed.) Ethics of Artificial Intelligence, pp. 383–412. Oxford University Press, New York (2020)
51. Steen, M., Sand, M., Van de Poel, I. (2021). Virtue ethics for responsible innovation. Business and Professional Ethics J, **40**(2), 243–268