SOFF Presentation 1

The concept of 'meaningful human control' (MHC) originates from discourse on autonomous weapons systems
(AWS). It emphasizes the notion that humans must remain in a position of control or oversight over the decision-
making of a lethal system. In other words, such types of systems should not be able to execute lethal action 2
without human intervention. The quote you see here is by scholar and political military strategist Thomas K. Adams
who communicates the difficulty of formulating a practical solution while also preserving the ever-increasing
processing rates that accompany increased automation.

Although technological innovations have always played a key role in military operations, autonomous weapons 3
systems (AWS) have received asymmetric attention in public debate as well as academic discussions—and for good reason. As these systems are designed to carry out more and more tasks once in the domain of human operators, questions regarding their autonomy and potential recalcitrance have sparked discussion. Debate highlights a potential *accountability gap* between their use and who, if anyone, can be held accountable. At the international level, discussions about how to exercise control over the development and deployment of these autonomous military systems have been underway for over a decade. Still, there remains very little consensus as to what constitutes a sufficient level of control.

The concept of meaningful human control (MHC) has emerged in discourse to encompass this ideal of human 4
control over autonomous systems. Various approaches have been taken to define a sufficiently robust notion of MHC that addresses technical requirements proper training for use designer-user engagement, operations planning, design requirements, and the responsibility of designers. Each of these approaches provide insight into how MHC over these types of systems can be understood and attained. Although they are generally proposed as isolated frameworks for attaining MHC, they share some underlying precepts. Approaches that emphasize the operational planning and military context of use provide a strong contextual landscape for understanding MHC. Whereas other approaches may focus on design histories, designer intentions and plans, or the responsibilities of designers and supraindividual agents. They provide cogent arguments for designing these systems with both backward- and forward-looking responsibility. Still, they largely focus on a single level of abstraction at the opportunity cost of the other.

What I want to discuss in this short presentation is that MHC can only be attained if militaries and industries work 5
closer together as a supraindividual agent. As a way of marrying these often-isolated projects of defining MHC, I propose a two-tiered approach to understanding MHC. First, it doesn't make sense to divorce discussions of AWS from actual and often trivial military operations; AWS exist within this landscape, not outside it. One must therefore situate AWS within their operational context in order to understand them. However, this does not mean there are no accountability gaps in terms of technical (fully)AWS. The question of design remains important for determining the responsiveness of a system to the relevant moral reasons of the relevant agents, thereby creating the design level. By coupling these two levels, we can account for the technical and full autonomy of certain types of AWS. Many of the issues associated with (fully)AWS can thus be rendered non-issues.

The Operational Level of Control 6

In this level of MHC we can begin by grounding it on military operation practice that both supports and constrains targets in areas of operations. This level of control focuses on higher level of organization and operational control of the military as a supraindividual agent. This addresses the fact that the 'autonomy' of AWS (and of any human agent in the military, such as soldiers) is necessarily constrained by such operations. The result of these constraints is that 'full' autonomy, which is often construed in discussions on AWS, is not 'full' in the sense that is often implied (e.g., self-determining agents). Instead, it is restricted to various operational decisions and planning a priori to deployment and operations.

We can use a case of conventional air operations to frame human involvement in operations through a dynamic targeting process to illustrate this. By framing the role of human agent decision-making within distributed systems, we can outline ways policymakers and theorists can determine how military planning and operations *actually* function. AWS can then be deployed within the context of use of these practices. Characterizing the human role in military decision-making,

Pre-mission Briefing

Before the mission is undertaken, the air component receives a briefing with information on mission execution. Such briefings are often highly detailed with information such as “target location, times, and munitions”; however, they are less detailed when we consider dynamic targeting in situ. Such information is distributed to various domains of operations to specialists, who then vet and use it in more detailed planning. The executers of the mission, in this case fighter pilots, are then brought in for briefing on the mission details. The pilots take the time to study the information provided while also taking care of any last-minute preparations for execution.

To find the target for operations, intelligence and data are required. Such targets are pre-programmed in the navigation systems of both the fighter jet and the payload. Whereas a dynamic target requires in situ data collection, the task here involves arriving at the preprogrammed “weapon’s envelop (i.e., the area within which the weapon is capable of effectively reaching the target)”. This process is displayed on the operations heads-up display

Once the operator arrives within the weapon’s envelope, onboard systems aim to positively identify the target confirmed during operational planning. This ensures payload delivery complies with relevant military and legal norms In this case, targets were preplanned and confirmed. For positive target identification, the operator usually does not engage in visual confirmation; instead, they refer to onboard systems and the validation that took place during operational planning to ensure lawful engagement of the identified target. Even in this fixed case of pre-planning, the human pilot does not need to attend to anything else during this phase other than arriving within the weapon’s envelope

The operator tracks the target within the weapon’s envelope to ensure the continuity of positive identification. This also provides concurrent updates regarding the position and status of the target. In the case of a static target (e.g., a military compound), tracking is relatively straightforward and involves simply entering the weapon’s envelope as in the fix phase.

During this phase, the relevant rules of engagement (RoE), laws of armed conflict (LoAC) and other targeting rules are invoked to ensure lawful targeting and deployment. These also address other considerations, such as issues related to collateral damage and risk factors that may result to one’s own forces. In this predetermined and validated target case, the legal and military experts who vetted the target permit the pilot to simply input relevant data into the vehicle and weapons payload delivery systems to ensure proper execution. Given the visually impairing weather conditions, any further collateral damage estimates cannot be attained. Planning at pre-mission stages validated that collateral damage estimates were low and were conducted according to the norms that govern them. The human pilot thus does not actively participate or intervene beyond piloting the vehicle into the weapon’s envelope.

Once the operator enters the designated weapon’s envelope, the onboard computer suggests to the pilot the most opportune time for releasing the payload to ensure effectiveness. This suggestion is based on its knowledge of the capabilities of the equipped weapons system. Given that the payload system itself is GPS guided, there is no need for any other forms of targeting based on visual identification. Once the pilot authorizes the release of the weapon, the munitions guide themselves to the target.

At this point, the results from the previous stage are assessed to determine the effects of the strike. Of course, a visual assessment from the pilot can be impaired by a number of factors (weather conditions, in this case). Similarly, visual assessments of collateral damage from the vantage point of a pilot may fail to accurately reflect the efficacy of the strike and its consequences. In the case of aerial engagements such as this, ground support forces may be required for a more accurate assessment of engagement.

In considering MHC then, it appears that most (if not all) of the performance latent to each step is beyond the pilot's control. It could be argued that this is emblematic of contemporary aerial operations more generally. While the pilot can be seen as in direct operational control of some of the operation, piloting the craft to the weapon's envelope and engaging in weapons release, this type of control is not sufficiently meaningful. This is because the pilot lacks full "cognitive clarity and awareness" of the situation within which they are participating. The privation begs the underlying question of whether the pilot actually possesses levels of clarity and awareness sufficient enough to be deemed substantial in a meaningful way.

Discussions at the pilot level could provide some future insight both for operations employing AWS as well as modern aerial crafts. But these would converge on the operator, which is the wrong vector. Alternatively, such discussions should emphasize how the military as a supraindividual agent (i.e., an organization) can have MHC over targeting operations. Because of this, the ongoing international debate on AWS focuses overly much on the deployment stage of AWS and their relations to individual operators. In doing so, the debate attempts to locate the vector for MHC between those two agents (AWS-human). But it ignore the broader covariance of the distribution of labour between agents within a military complex that determines decision-making practices. The steps outlined above, particularly the pre-mission briefing stage with its collateral damage and proportionality assessments, are largely sidelined in these discussions.

This approach shows the need for a distributed notion of MHC to accurately account for numerous decision and measures undertaken by different agents in the broader decision-making mechanism before deployment. Different agents have different levels of control over any given vector in the process. Any sufficient conception of MHC must therefore reflect this. Of course, this does not negate the role that human operators play. Rather, it positions the role within the larger distributed network of decision-making. Here, 'full autonomy' is not full in the sense that is commonly intuited. It is constrained by the larger apparatus within which it forms a part.

The Design level 8

In this level, MHC is the co-variance between the behavior of a system and the intentions or reasons behind an agent's decisions and actions. Systems can be designed in ways that permit agents to forfeit some of their direct operational control while still possessing global control over the system itself. This means that more, rather than less, levels of autonomy may (in certain cases) permit more salient control of a system. As mentioned in the preceding section, more direct operational control has little meaning in the desired sense for autonomous systems. With this understanding, clearer lines of accountability can be drawn when humans remain 'in-the-loop' over a system. As tracking the relevant reasons behind an agent's decisions is a necessary condition for MHC, the retention of humans 'in-the-loop' allows MHC.

This approach to MHC is functionally comprehensive in its breadth, which looks beyond individual systems to the whole sociotechnical infrastructure wherein systems are embedded. Although the specific design and deployment of systems implicate important factors in understanding MHC, they cannot be understood in isolation from the infrastructures, organizations, and other agents who are inextricably connected to their design, deployment and use (which is the military-industrial complex). The approach focuses on the design level because it describes MHC as something that can be designed *for* by engineers. In other words, MHC are technical design requirements—not only for the system itself, but also for relevant sociotechnical infrastructures.

In order to design for MHC, two necessary conditions must be met: tracking and tracing. Satisfaction of these two 9
conditions permits a more comprehensive conception of MHC that reaches beyond that of solely end users. Here, a level of meaningful control is extended to agents such as designers and policymakers along with organizations and states. With this control comes clearer lines for attributing responsibility. For the sake of time I won't go into the depths of the design level, because this is fundamentally a function of engineers and designers, if you want to know more about this please feel free to contact me.

MHC as design thinking and design engineering as MHC 10

As mentioned in the introduction, the central premise of a possible ban on AWS is grounded in a certain level of autonomy that results in an accountability gap in the event of recalcitrance. But we have to be very clear what we mean when we speak about autonomy for AWS. There are currently 5 level sof autonomy as understood in the literature. The least problematic stage is level 1. Levels 4 and 5 are arguably the most problematic. Both are seen as dangerous due to how an AWS selects a target (i.e., systemic opacity, computer vision, etc.) and its technical ability to do so as a function of various targeting norms and the rules of engagement. Level 4 questions the cognitive clarity of a human operator who has veto power when determining the validity of the target(s) chosen by the system. Regardless, Level 5 is typically the subject of debate as it is considered the descriptor of full autonomy in terms of AWS.

Here, we can already begin to tease out some of the issue with problematizing autonomy per se. There are convincing arguments contra AWS other than the supposed accountability gap proposed by the above ordinance, such as the dehumanization of war and its deleterious effects on human dignity. But it appears that actual military operations planning and deployment intuitively constrains the autonomy of any given agent, soldier, or AWS to being a function of a larger a priori plan. This plan bears little, if any, intrinsic operational value outside their functional capacity to carry out such plans. Of course, this does not extricate the AWS deployed within such constraints from the possibilities of limitless actions or wanton recalcitrance. As a predicate for technical design requirements, technical design must reflect both the proximal and distal intentions and goals of relevant agents within the deployment envelope. These would be the commanders who employ such weapons in their area of operations as well as potential human operators who may be engaging with them symbiotically on the ground (e.g., aerial AWS such as fully autonomous drones). Regardless, the system capacity to respond to the relevant moral reasons of relevant agents must be considered a foundational variable in the weaponeering decision-making process for any given context of deployment in the pre-mission stages.

The procedural process of operational planning and target identification form the higher, or meta-, level of MHC as clearer lines of causality can be conceptualized. This culminates in weapons release and efficacy assessments. Similarly, the design level of MHC is functionally dependent on a systems understanding of both tracing design histories as well as tracking the responsiveness of autonomous systems to the relevant moral reason(s) of the relevant agent(s) in the design and use chains for such systems. Theoretically speaking, both levels are predicated on systems or networks of interconnected parts. Similarly, both LoA feed into one another despite their different scopes. Within the operational level, the bounds in which weaponeering decisions are made prior to deployment are contingent on the functionality of the system itself in order for it to be chosen as the most salient means for carrying out the intended mission. But how such technical responsiveness to the on-the-ground needs for successful mission completion is not contingent on those types of pre-mission assessments. System-level recalcitrance can jeopardize the overall level of MHC even when the system is bound by the operational level of control.

Weaponeering decisions must thus be reflected at the design level in order for those decisions to be sufficiently salient prior to deployment. In this sense, the operational level feeds down into the design level by supplying the norms, objectives, and intentions necessary for deployment to be lawful. These are also necessary for the operational level to be holistic in terms of the sufficiency of control. Likewise, the various agents that are essential to pre-mission planning operations form part of the population of relevant moral agents (or collectively as a supraindividual agent); these agents permit the design level to actually design AWS so they are sufficiently responsive to the reasons and intentions of the actor(s) who makes the weaponeering of AWS permissible—and thus under a priori MHC on both levels. Of course, this would mean a closer military-industrial partnership that uses these agents as stakeholders for whom systems can be designed (coupled with the relevant RoE and LoAC).

One scenario that is often discussed in the literature contra AWS is that of an AWS killing civilians. Within this scenario, we can begin to trace reasons for dismissing it. In order for an AWS to kill a civilian on the ground, the civilian must fall within the weapons envelope delimited prior to deployment. The killing is not mala in se to the extent that collateral damage assessments are agreed upon pre-deployment under existing norms for proportionality. To some degree, the killing of civilians is not necessarily equivocal to recalcitrance as it can be traced back to the briefing information. If we imagine that an AWS kills civilians disproportionately even within the weapons envelope and even against explicitly acceptable damages determined in pre-planning, this can be construed as technical recalcitrance. This is because it can be traced back to the relevant agents within the design and use histories of the AWS to determine whether the system was designed in such a way as to be maximally responsive to the relevant intentions of those agents. If such is shown to not be the case, then the AWS was under MHC. It is thus not a viable option for weaponeering decisions and its deployment was unlawful overall (this is a good vector for thinking about ban criteria). If relevant agents such as designers and users (commanders, AWS designers/programmers, proportionality specialists, etc.) are capable of understanding the capabilities and consequences of the system, then they may be said to be in possession of MHC. They have MHC both in their weaponeering decisions on the operational level, as well as in design decisions at the design level. Divorcing one level from the other leaves open vectors from which accountability gaps can arise.

Many of the technical issues presented as mala in se against the development of AWS, such as increased autonomy or the targeting of civilians, are only problematic if decoupled from responsible design, actual military planning, and actual operations practices. When these are taken into account, the augmentation of autonomy is necessarily constrained by many—if not all º of these processes. In certain cases, autonomy can increase rather than decrease the ability to have MHC. If these systems are designed so as to be maximally sensitive to the relevant moral reasons of the relevant moral agent(s), then they likewise augment MHC rather than lessen it.

Divorcing the operational level from design leaves design impotent and potentially recalcitrant. Divorcing the design level from operations leaves operations with an opaque and nebulous lethal tool that may result in poor (if not unlawful) weaponeering decisions. We can think about systems and, more specifically, these various levels of abstraction as co-constituting one another. This permits their inherent complexity to be modeled more easily. As a consequence, we can design for complexity rather than leaving design decisions ad hoc afterthoughts.