Toward Meaningful Human Control of Autonomous Weapons Systems through Function Allocation

Marc C. Canellas & Rachel A. Haga
Cognitive Engineering Center
School of Aerospace Engineering
Georgia Institute of Technology
Atlanta, Georgia, 30332-0250
Email: {marc.c.canellas, rachel.haga}@gatech.edu

***Abstract*— One of the few convergent themes during the first two United Nations Meeting of Experts on autonomous weapons systems (AWS) was the requirement that there be meaningful human control (MHC) of AWS. What exactly constitutes MHC, however, is still ill-defined. While multiple sets of definitions and analyses have been published and discussed, this work seeks to address two key issues with the current definitions: (1) they are inconsistent in what authorities and responsibilities of human and automated agents need to be regulated, and (2) they lack the specificity that would be required for designers to systemically integrate these restrictions into AWS designs. Given that MHC centers on the interaction of human and autonomous agents, we leverage the models and metrics of function allocation – the allocation of work between human and autonomous agents – to analyze and compare definitions of MHC and the definitions of AWS proposed by the U.S. Department of Defense.**

Specifically, we transform the definitions into function allocation form to model and compare the definitions, and then show how a mismatch between authority and responsibility in an exemplar military scenario can still plague the human-AWS interactions. In summary, this paper provides a starting point for future research to investigate the application of function allocation to the questions of MHC and more generally, the development of rules and standards for incorporating AWS into the law of armed conflict.

Keywords— Autonomous weapons systems, meaningful human control, function allocation, human-automation interaction

I. INTRODUCTION

The next generation of military technology has many names: autonomous weapons systems (AWS, [1]), lethal autonomous weapons systems [2], lethal autonomous robots [3], [4], killer robots [5], [6], terminators [7], and cyborg assassins [7]. Whatever the name, the technology, in principle, is a relatively simple combination of new and old capabilities: new software and hardware capable of making decisions without humans are being attached to established lethal weapons capable of killing humans. This convergence of technology supply and military demand has enabled the existence of AWS [8]. The increasing amount of media coverage, formal international discussions, legal essays, and technical articles can be attributed to the growing awareness that in the near future, technology itself will not determine the constraints on military combat and casualties [2]. Instead the limitations will only exist in restrictions imposed on the use of military technology such as AWS. This responsibility is likely to fall entirely to the domains of ethics, law, policy, and empathy.

While research has begun to address these concerns, the previous work has (1) been inconsistent in what authorities and responsibilities of human and automated agents need to be regulated, and (2) lacked the specificity required for designers and engineers to implement these restrictions in a systematic and meaningful way. Therefore, this paper will show how the domain of cognitive engineering, specifically function allocation, can provide technical support for the discussion and development of appropriate constraints.

As defined by the U.S. Department of Defense (DoD), an AWS is "a weapon system that, once activated, can select and engage targets without further intervention by a human operator" [1]. There have been significant public and literature debates about how autonomy will affect warfare and how the law of armed conflict can adapt to these changes. To this end, there have been two United Nations Convention on Certain Conventional Weapons (CCW) Meetings of Experts focused on AWS, the first in May 2014 and second in April 2015.

One of the few convergent themes during the CCW Meetings of Experts for AWS was the requirement that there be meaningful human control (MHC) of AWS [9]. MHC was introduced as a term by the British non-governmental organization, Article 36 [10], as an organizing principle "that those who plan or decide on an attack have sufficient information and control over a weapon to be able to predict how the weapon will operate and what effects it will produce in the context of an individual attack, and thus, to make the required legal judgements" [11]. The United Nations Institute for Disarmament Research (UNIDIR) discussed that MHC could be a beneficial framing concept for international debate because it is (1) more precise than discussions of "human-in-the-loop" or "judgment," (2) consistent with the laws of armed conflict by discussing human responsibility for decisions made, and (3) broad enough to integrate human-machine interaction [9].

Since its introduction, MHC has been a popular concept among nation-states, non-governmental organizations, and researchers for framing discussions on AWS [9]. However, regardless of the popularity and utility of the current discussions of MHC, there is a critical need for a precise

definition and operationalization of MHC. As stated by UNIDIR, "If states wish to move from using [meaningful human control] simply to structure policy discussion to using it as a basis for an international norm, further work will be needed to develop a shared understanding of how such control is operationalized" [10].

The challenge faced in developing an operational standard for MHC is explained by Anderson *et al.* [12]: "The double challenge here is, on the one hand, for States to apply sufficiently clear and robust standards and rules,... as sophisticated, modern autonomous weapon systems are gradually fielded. And, on the other hand, ensure that standards and rules... that States develop today will be equally relevant or adaptable for the future systems which will be developed ten, twenty, thirty years from now."

This paper argues that the foundation for a precise, comprehensive and robust definition of MHC is found in the cognitive engineering discipline, whose primary focus is on the interaction between automation and humans, particularly in complex and dynamic domains. Of specific interest to this work is cognitive engineering's concept of function allocation. Function allocation is defined as the allocation of work within teams of human and automated agents [13]. In a series of papers Feigh and Pritchett first derived five requirements for effective function allocation from a critical literature review [13], to which Pritchett, Kim, and Feigh outlined a modeling framework [14] and eight metrics [15] for evaluating human-automation function allocation.

To show how function allocation can be utilized to frame MHC such that it is sufficiently precise to be applied to a specific system, while simultaneously comprehensive enough to adapt to the inevitable evolution of AWS, we first review the current ambiguities and questions regarding proposed definitions of MHC. Then we review the requirements of effective function allocation, methods for modeling and measuring function allocation, and show how they relate to MHC. Based on the established modeling and measuring techniques for function allocation, we then compare and analyze the definitions of MHC and the DoD AWS definitions. We conclude the paper by introducing a framework for incorporating operational and technical considerations into the process developing rules and standards for AWS.

II. Background

A. Current Definitions of Meaningful Human Control

Meaningful human control definitions are fundamentally exclusionary. They are intended to be a set of minimum requirements for sufficient information and control such that any weapon which does not enable this information and control should not be allowed to operate in armed conflict. Horowitz and Scharre [16] provide the canonical examples of the two extremes of MHC:

> Consider a person who sits in a room and is supposed to press a button every time a light bulb in the room goes on. If the person does this as instructed, and a weapon fires each time the person presses the button, a human has fired the weapon, but human control over the weapon is far from meaningful. Alternatively, the Platonic form of meaningful human control is when a person swings a sword, axe, or knife -- or uses their bare hands -- to directly end the life of an enemy combatant.

To begin to precisely define the differences between these two cases, Table 1 shows three proposed definitions of MHC from Article 36 [11], ICRAC [17] (Sharkey [18], a member of ICRAC, provides similar autonomy requirements), and Horowitz and Scharre [16]. While these definitions reveal progress toward a definition, there are still limitations that must be addressed. (Also, though outside the scope of this paper, there is still discussion as to the general applicability of MHC to the law of armed conflict [9], [12].)

Horowitz and Sharre raised the operational concern that definitions of MHC must practically consider the current accepted weapon use by not ruling out "the use of weapon systems that today are used without controversy, including systems that make civilian casualties less likely" [16]. If these practical considerations are not accounted for, then the definition would have a low likelihood of success at the national or international level. Horowitz and Scharre criticized the ICRAC definitions from this operational perspective, stating that ICRAC's definitions are an "idealized version of human control divorced from the reality of warfare and the weapons that have long been considered acceptable in conducting it." ICRAC's first and third MHC requirements of a human commander would seem to eliminate the use of unguided munitions which do not contain a guidance system and simply follow their ballistic trajectory. Once the bomb or missile or bullet is launched/shot at an area containing enemy combatants, a non-combatant may enter the target area, however, the commander would not be able to 'react' to the change (ICRAC Req. 1) nor suspend or abort the attack (ICRAC Req. 2).

Furthermore, even if there was agreement on what operational concerns to address and regulate, in many cases the wording is insufficiently precise to be practically applied by designers. Consider for instance the terms 'adequate contextual information', 'sufficient time for deliberation', or 'properly trained.' While they all are well intended requirements, there are no systematic ways to determine if these metrics are met. It is highly unlikely that any two designers, much less nations, would interpret them in the same way.

The current states of these definitions of MHC leave many additional issues to be considered. In concluding discussions of proposed MHC definitions, authors consistently express a need for further debate and further precision (e.g. [9], [10], [16], [17], [19]).

B. Requirements for Effective Function Allocation

Progress in defining MHC will require the use of precise definitions, models, and measures. MHC in the context of lethal autonomous weapons systems is, at its heart, a question of function allocation. A function is an activity to be

Table 1. Summary of proposed definitions of meaningful human control.

Article 36 [10]	**ICRAC** [17]	**Center for a New American Security** [16]
	PREFACE	
Requirements for meaningful human control over individual attacks include, but are not necessarily limited to:	ICRAC hold that the minimum necessary conditions for meaningful control are:	Meaningful human control has three essential components:
	REQUIREMENTS	
Information – a human operator, and others responsible for attack planning, need to have adequate contextual information on the target area of an attack, information on why any specific object has been suggested as a target for attack, information on mission objectives, and information on the immediate and longer-term weapon effects that will be created from an attack in that context.	First, a human commander (or operator) must have full contextual and situational awareness of the target area and be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack.	1. Human operators are making informed, conscious decisions about the use of weapons.
Action – initiating the attack should require a positive action by a human operator.	Second, there must be active cognitive participation in the attack and sufficient time for deliberation on the nature of the target, its significance in terms of the necessity and appropriateness of attack, and likely incidental and possible accidental effects of the attack.	2. Human operators have sufficient information to ensure the lawfulness of the action they are taking, given what they know about the target, the weapon, and the context for action.
Accountability – those responsible for assessing the information and executing the attack need to be accountable for the outcomes of the attack.	Third, there must be a means for the rapid suspension or abortion of the attack.	3. The weapon is designed and tested, and human operators are properly trained, to ensure effective control over the use of the weapon.

performed by an agent (e.g. selecting or engaging a target). Therefore, function allocation determines how to allocate work within teams of human and automated agents [13]. Based on a review of the function allocation literature, Feigh and Pritchett [13] derived five requirements for effective function allocation:

1. Each agent must be allocated functions that it is capable of performing.
2. Each agent must be capable of performing its collective set of functions.
3. The function allocation must be realizable with reasonable teamwork.
4. The function allocation must support the dynamics of the work.
5. The function allocation should be the result of deliberate design decisions.

The only way to adhere to the fifth requirement of effective function allocation, that function allocation should be the result of deliberate design decisions, is through the use of effective models and measures of function allocation. Based on reviews of modeling [14] and measuring [15] techniques for function allocation, this subsection provides justification for their specific application to the operationalization of MHC.

Two examples of modeling function allocation from [14] are: modeling the teamwork and individual work through abstraction hierarchy, and modeling teamwork as defined by function allocation and team design. The modeling of teamwork through function allocation and team design is most applicable to the proposed definitions of MHC as that is the level of specification of the definitions in MHC in Table 1 and the DoD definitions of AWS. This modeling provides explicit notation of which agent (e.g. human operator or AWS) has control over which function. From the explicit notation, different specifications of teamwork (e.g. different definitions of MHC and definitions of AWS) can be compared.

From those models, Pritchett, Feigh, and Kim [15] developed 8 metrics for evaluating function allocation which could be derived from human-in-the-loop simulations or understandings of real operations: 1) workload, 2) stability of the work environment, 3) mismatches between responsibility and authority, 4) incoherency in function allocations, 5) interruptive automation, 6) automation's boundary conditions, 7) function allocations limiting human adaptation to context, 8) and mission performance.

With the high legal and political concerns of AWS, mismatches between authority and responsibility are particularly relevant to MHC. Authority describes which functions an agent is asked to perform and responsibility describes which outcomes an agent will be accountable for in an organizational, regulatory or legal sense [13]. The exemplar of a gap between authority and responsibility is when automation is assigned to execute a function in an operational sense but human will be held accountable in an organizational and legal sense for its outcome [14]. One common result of mismatches between authority and responsibility is that the human is unable to assess whether automation is correct, leading to overtrust or undertrust in the automation's actions [13].

C. Parallels between Function Allocation and Meaningful Human Control

The status and history of function allocation research is particularly relevant to MHC. In the following three ways, we show that function allocation research can enable novel, technical perspectives on MHC issues.

1) Subjective-Objective Principles

Foundational principles of the law of armed conflict, such as proportionality, still rely on combinations of subjective-objective standards for determining key components [20]. Function allocation is a field with a similar combination of subjective understandings and objective metrics. Designing

Table 2. Summary of requirements for effective function allocation categorized into taskwork, teamwork, and collective work [8].

Effective Function Allocation Types	Design Considerations	Resulting Issues if Requirement is Not Met
TASKWORK: Each agent is able to perform each of the taskwork functions assigned to him/her/it (Effective Function Allocation Requirements 1 & 2)	Design functions around both the automation and the human's capabilities.	Brittle automation: where emergencies must be handled by humans and there is little support for off-nominal conditions
		High workload spikes during off-nominal situations [22] and excessively low workload during normal operations in between the spikes leading to humans becoming out of the loop [22], [23].
		Leftover allocation (automate as many functions as technology will permit, and assume the human will pick up whichever functions are leftover) often results in a human assigned to monitor automation or the environment for conditions beyond which the operate; functions in which humans are ineffective [24]
	Consider authority-responsibility	Gap between authority and responsibility: automation is assigned to execute a function is an operational sense but human will be held accountable in an organizational and legal sense for its outcome. [25]
		Without being able to assess whether automation is correct, humans often overtrust or undertrust the automation [26]
TEAMWORK: Each agent is able to perform each of the teamwork functions assigned to him/her/it (Effective Function Allocation Requirement 3)	Consider automation as a team member.	Automation and/or human unable to anticipate each other's information needs and provide information at useful, non-interruptive times [27]
		Clumsy automation which interrupts its team members because it cannot implicitly sense information about whether other team members would benefit from an interruption [28]
COLLECTIVE WORK: Each agent is able to perform the collective set of taskwork and teamwork functions. (Effective Function Allocation Requirements 3 & 4)	Dynamic analysis of the physical environment. Function allocation should support how the work environment is managed by the agents.	In complex work environments where many functions interdependent and coupled, some couplings may be hidden. Unaccounted-for couplings can result in insufficient coordination, idling, and workload accumulation.
	Examine the tradeoff between maintaining predictability for the human versus applying complex automated capabilities and dynamically allocating functions [29]	Adaptive function allocation (dynamically changing function allocation) can aggravate environmental unpredictability.

function allocation has long been considered an art [21]. For example, desired attributes of automation include that it should be "a good team member" and "not clumsy" with general heuristics used to evaluate criteria [15]. Only recently has modeling and measuring advanced to states where researchers could start identifying good/bad function allocation using objective metrics [14], [15]. This push toward more quantitative assessment of function allocation is similar to the push for operationalization of MHC [9].

2) Exemplar Human-Automation Issues

Given the number of experiments and simulations used to examine these issues for function allocation and human-automation interaction, many issues have already been identified and explored. In addition to the literature review of Feigh and Pritchett [13], Sheridan and Parasuraman [30] review specific automation-related incidents and accidents.

3) Focus on Minimum Requirements

The particular relevance of the requirements for effective function allocation -- and the derived models and measures – is that, "instead of seeking an optimal function allocation, [the authors] identify those requirements that any function allocation should meet" [13]. Just as MHC definitions are intended to set the minimum requirements for "sufficient information and control" of a weapon [11], these requirements for effective function allocation set general minimum standards for dividing work between human and automated agents within a team.

III. Modeling and Evaluating Function Allocation Designs for Meaningful Human Control

There are numerous methods of modeling, measuring, and evaluating the function allocation of complex systems. From our review of the function allocation literature we have selected one method of modeling and evaluating to exemplify the ability of function allocation to assist in the process of developing the rules and standards of MHC. A table describing the implementation of function allocation (Tables 1-4 of [14]), what we term function allocation form, is used to model the proposed definitions of MHC and the policy definitions of AWS set out by the DoD. Once the definitions are transformed into function allocation form, the definitions are compared to determine potential conflicts. Then, to measure and evaluate the definitions we examine the potential for mismatches between autonomy, authority, and responsibility. Since we do not have computational models like [15] we do not explicitly count the number of mismatches. However, the discussion of the mismatches shows the value of the technique.

A. Comparing Definitions through Function Allocation Form

The MHC definitions in Table 1 and the official DoD definitions of AWS [1] were transformed into function allocation form with results in Table 3. To explain the transformation, we use the DoD definition of a semi-AWS:

Table 3. Transformation of proposed meaningful human control definitions and U.S. Dept. of Defense (DoD) definitions, into function allocation form.

	Proposed Meaningful Human Control Definitions		DoD Definitions		
Function	**Article 36**	**ICRAC**	**AWS**	**Human-Supervised AWS**	**Semi-AWS**
Tracking and identifying targets	Automation*	Automation*	Automation*	Automation*	Automation
Select targets	Human	Human	Automation	Automation	Human
Information about target	Human	Human	Automation*	Human*	Human*
Information on mission objectives	Human	Human	Automation*	Human*	Human*
Information on weapon effects	Human	Human	Automation*	Human*	Human*
Cueing potential targets	Automation*	Automation*	Automation*	Automation*	Automation
Prioritizing selected targets	Automation*	Human	Automation*	Automation*	Automation
Timing when to fire	Automation*	Automation*	Automation*	Automation*	Automation
Engage targets	Human	Human	Automation	Automation	Automation
Terminal guidance	Automation*	Automation*	Automation*	Automation*	Automation
Intervene and terminate engagements	Human*	Human	Automation*	Human	Human*

* Interpretation of whether the function would be allocated to a human or automation as the definition was unclear.

> Semi-autonomous weapon system: A weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator.
>
> This includes:
>
> Semi-autonomous weapon systems that employ autonomy for engagement-related functions including, but not limited to, acquiring, tracking, and identifying potential targets; cueing potential targets to human operators; prioritizing selected targets; timing of when to fire; or providing terminal guidance to home in on selected targets, provided that human control is retained over the decision to select individual targets and specific target groups for engagement.

The definition of semi-AWS is explicit in describing which functions are assigned to a human agent or autonomous agent. Each of the functions listed in the semi-AWS definition were listed in Table 3 and then designated, per the definition, to a human agent (Human) or automated agent (Automation). Where definitions were unclear about whether a function should be allocated to a human agent or automation agent, an interpretation was made and denoted by the use of an asterisk*.

For the definitions of MHC from Article 36 and ICRAC, in written form the similarities or differences were difficult to see, however, in function allocation form, the similarity and differences are clearer.

With regards to selection of targets and information provided to the operator, both definitions agree. In contrast, ICRAC explicitly states that humans must be perform: 1) prioritization of selected targets and 2) intervening and terminating engagements.

For the definitions of AWS, the increased automation from semi-AWS to AWS is also more evident in function allocation form. The definition for semi-AWS explicitly states that human control is retained over selection of targets while human-supervised whereas human-supervised AWS only requires humans to be able to intervene and terminate engagements.

Table 3 shows that the proposed definitions from Article 36 and ICRAC are in conflict with the U.S. DoD definitions. Specifically, should either Article 36 or ICRAC's definitions be adopted in the laws of armed conflict, all levels of the DoD's AWS would be illegal. Based on Table 3, the suggested issue of discussion between the groups is how to allocate the functions: selecting targets, prioritizing selected targets, and engaging targets.

B. Evaluation of Mismatches Between Authority and Responsibility

To show how function allocation can provide critical analysis of proposed definitions of MHC, this subsection uses the metric, "mismatches between responsibility and authority" to examine a real military scenario. From an operational and technical standpoint, the question of concern is: do the proposed MHC definitions, as they are written, actually ensure their intended goal of requiring sufficient information and control of a weapon operating in armed conflict?

As an example of this potential concern, we use the example of Rules of Engagement for the U.S. Army of anticipating attack when encountering a potential threat [31]. A soldier is required to "not base anticipatory force on a mere hunch that the person is hostile" [31] and instead determine if someone's intentions are hostile based on the SALUTE format:

- Size – How many individuals are you facing?
- Activity – What are they doing? Pointing a weapon?
- Location – Are they within arms range? In a prepared firing position? Entered a restricted area?
- Unit – Are they wearing a uniform? Part of an organized armed force?
- Time – How soon before they are upon you?
- Equipment – Are they armed? With what? Range and lethality of his weapon?

The following is a scenario adapted from a real incident with the U.S. Marine Corps in Somalia in 1993 [31]:

> Imagine a soldier sitting in front of a text-based console at a military base monitoring an autonomous weapon system miles away protecting a convoy of officials as they move through a dangerous city. In this city, adults have been seen handing grenades to children and persuading them to use them against convoys. As the convoy moves, verbal warnings are constantly given out by the AWS to stay away from the convoy with the threat of deadly force. Suddenly, a boy, ignores the warnings and runs toward the convoy. At this point, the soldier's text-only screen comes to life: "Target approaching convoy. Target ignored warnings. SALUTE factors support identification as hostile. Target 5 seconds from convoy. Engage with aimed fire?"

In this scenario the soldier is responsible for determining whether to engage and the legal considerations of the engagement, but the AWS has the authority over analyzing the information. Importantly, most of the information requirements put forth by Article 36 (see Table 1) were satisfied but yet, there is still a mismatch between authority and responsibility. The soldier is provided some level of contextual information and the SALUTE factors did support engagement in the real-life scenario, but there is likely inadequate contextual information for the soldier to have all the responsibility.

IV. Framework for Developing AWS Rules and Standards

It can be further argued that the effective function allocation literature provides a framework for designing rules and standards for AWS. In addition to Anderson *et al.* who stated that "rules and guidelines for the development of autonomous systems...could be based not only on legal requirements, but also policy considerations" [12] we argue that technical considerations are also fundamental to effective rules and guidelines that can practically implemented and enforced.

First, law and policies should be transformed into function allocation form. Specifically, laws are the rules and standards to be followed compulsorily (e.g. the law of armed conflict: necessity, distinction, proportionality, and humanity [12]) whereas policies are the objectives or goals of rules and standards (e.g. the proposed definitions of MHC and AWS). This allows for the incorporation of accountability and moral responsibility standards [16], based on what the law of armed conflict states regarding allowable actions during armed conflict. These rules state what armed conflict should and should not be; including, for example, the specific rules of engagement (see [31]).

Then, operational and technical considerations can be used to determine what rules and standards of AWS will be effective. Operational considerations based on the realities of the use of force include that rules and standards should 1) cover the use of weapons across all of the various ways they may be used, and 2) reflect a realistic vision for how weapons are used by soldiers in all domains [16]. Additional operational considerations can be found in [9].

Technical considerations should be based on the effective function allocation literature. The principles [13], models [14], and measures [15], of function allocation provide an important basis for ensuring that the rules and standards are instituted for AWS can be technologically adhered to.

In an important practical sense, separating legal, policy, operational, and technical considerations allow for the two sides of the AWS debate to clarify their perspectives: the *atechnists*[1], who believe that technology (sensors, artificial intelligence, etc.) will never be capable of adhering to the law of armed conflict (e.g. [5], [6], [32–34]), and the *technists*, who believe that someday robot technology may advance to a point that will enable AWS to adhere to the law of armed conflict (e.g. [3], [35–37]).

For example, this framework could be used to clarify discussions such as those found in Docherty [6], which first states, "it is highly unlikely that a robot could be pre-programmed to handle the infinite number of scenarios it might face so it would have to interpret a situation in real time;" then states, "the test [of whether an attack is proportional] requires more than a balancing of quantitative data, and a robot could not be programmed to duplicate the psychological processes in human judgment that are necessary to assess proportionality." These two statements, and many others in the atechnist literature, conflate what is technically feasible, whether AWS can be preprogrammed to interpret situations in real-time, versus philosophical concerns over whether humans, and only humans, can adhere to the law of armed conflict. (See [35] for arguments that humans are not very good at adhering to the law of armed conflict.)

V. Discussion and Conclusion

The roles and responsibilities of technologists with respect to the development of autonomous weapons systems (AWS) is nearly unparalleled in the direct relationship to armed conflict and human rights. The widespread public, political, and academic discussion regarding the safe operations of AWS shows that there are significant social implications of operationalizing the rules and standards for AWS. As Anderson *et al.* [12] states: "it is quite rare for an international-law related question to arise before it actually becomes a real-life drama. There is therefore a unique (although probably short-lived) opportunity to get it right; to develop the rules and code of conduct for such systems before they are field on the battlefields of the world in large numbers."

In the interest of getting it right, this paper provided the foundational connection from the legal and policy considerations for MHC of AWS to the technical considerations of human-automation interaction described by

[1]The labels of the two sides use the Latin root, "tech-" which refers to technology. For example, atechnist refers to being against technology, specifically the use of technology which interprets the law of armed conflict.

effective function allocation. By transforming proposed definitions of MHC and operational definitions of AWS into function allocation form, we were able to examine the conflicts between what MHC definitions would regulate and what the DoD AWS definitions are attempting to build, and then through an example scenario we examined mismatches between what automation may be assigned to do and what the human will be held accountable for.

There is also much more to be done to address the legal, policy, operational, and technical challenges of AWS using effective function allocation. Future research in this area should include more advanced models, metrics, and even simulations (e.g. [38]) to evaluate real and hypothetical case studies of AWS operations.

ACKNOWLEDGMENTS

The authors thank the Sam Nunn Security Program at Georgia Tech, Dr. Sy Goodman, and Dr. Margaret Kosal for facilitating the initial development of this paper. We also thank the Dr. Amy Pritchett and Dr. Karen Feigh of Georgia Tech's Cognitive Engineering Center for teaching us the importance of human-automation interaction. We also thank the two anonymous reviewers for their instructive comments and remarks which helped to improve the paper.

REFERENCES

[1] U. S. D. of Defense, "Directive 3000.09: Autonomy in Weapon Systems," United States of America: Department of Defense, Nov. 2012.

[2] M. Hagerott, "Lethal Autonomous Weapons Systems (LAWS): Offering a Framework and Suggestions," in *CCW Meeting of Experts on Lethal Autonomous Weapons Systems*, 2014.

[3] R. Arkin, "Lethal Autonomous Systems and the Plight of the Non-combatant," *AISB Quarterly*, no. 137, 2013.

[4] G. E. Marchant, B. Allenby, R. Arkin, E. T. Barrett, J. Borenstein, L. M. Gaudet, O. Kittrie, P. Lin, G. R. Lucas, R. O'Meara, and others, "International governance of autonomous military robots," *The Columbia Science and Technology Law Review*, vol. 12, no. 7, pp. 272–315, 2011.

[5] B. L. Docherty, *Mind the Gap: The Lack of Accountability for Killer Robots*. Human Rights Watch, 2015.

[6] B. L. Docherty, *Losing Humanity: The Case Against Killer Robots*. Human Rights Watch, 2012.

[7] D. Garcia, "The Case Against Killer Robots: Why the United States Should Ban Them," *Feogn Affairs*. Foreign Affairs: Council on Foreign Relations, May-2014.

[8] M. Bowden, "How the Predator Drone Changed the Character of War," *Smithsonian Magazine*, pp. 1–3, 2013.

[9] "The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward," United Nations Institute for Disarmament Research (UNIDIR), 2014.

[10] A. 36, "Killer Robots: UK Government Policy on Fully Autonomous Weapons," Article 36, Apr. 2013.

[11] A. 36, "Key areas for debate on autonomous weapons systems," Article 36, May 2014.

[12] K. Anderson, D. Reisner, and M. C. Waxman, "Adapting the Law of Armed Conflict to Autonomous Weapon Systems," 2014.

[13] K. M. Feigh and A. R. Pritchett, "Requirements for Effective Function Allocation A Critical Review," *Journal of Cognitive Engineering and Decision Making*, vol. 8, no. 1, pp. 23–32, 2014.

[14] A. R. Pritchett, S. Y. Kim, and K. M. Feigh, "Modeling Human–Automation Function Allocation," *Journal of Cognitive Engineering and Decision Making*, vol. 8, no. 1, pp. 33–51, 2014.

[15] A. R. Pritchett, S. Y. Kim, and K. M. Feigh, "Measuring Human-Automation Function Allocation," *Journal of Cognitive Engineering and Decision Making*, vol. 8, pp. 52–77, 2014.

[16] M. C. Horowitz and P. Scharre, "Meaningful Human Control in Weapons Systems: A Primer," Center for a New American Security, Mar. 2015.

[17] D. Garcia, "ICRAC statement on technical issues to the 2014 UN CCW Expert Meeting," in *CCW Meeting of Experts on Lethal Autonomous Weapons Systems*, 2014.

[18] N. Sharkey, "Towards a principle for the human supervisory control of robot weapons," *Politica & Società*, vol. 3, no. 2, pp. 305–324, 2014.

[19] N. Sharkey, "ICRAC celebrates successful fulfillment of its 2009 mission." International Committee for Robot Arms Control, May-2014.

[20] J. D. Wright, "Excessive' ambiguity: analysing and redefining the proportionality standard," *International Review of the Red Cross*, vol. 94, pp. 819–854, 2012.

[21] T. B. Sheridan, H. P. Van Cott, D. D. Woods, R. W. Pew, and P. A. Hancock, "Allocating functions rationally between humans and machines," *Ergonomics in Design: The Quarterly of Human Factors Applications*, vol. 6, no. 3, pp. 20–25, 1998.

[22] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.

[23] M. R. Endsley and E. O. Kiris, "The out-of-the-loop performance problem and level of control in automation," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 2, pp. 381–394, 1995.

[24] R. W. Bailey, *Human performance engineering: A guide for system designers*. Prentice Hall Professional Technical Reference, 1982.

[25] D. D. Woods, "Cognitive technologies: The design of joint human-machine cognitive systems," *AI magazine*, vol. 6, no. 4, p. 86, 1985.

[26] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 39, no. 2, pp. 230–253, 1997.

[27] E. E. Entin and E. B. Entin, "Measures for evaluation of team processes and performance in experiments and exercises," in *Proceedings of the 6th International Command and Control Research and Technology Symposium*, 2001, pp. 1–14.

[28] K. Christoffersen and D. D. Woods, "Advances in human performance and cognitive engineering research," E. Salas, Ed. Bingley, UK: Emerald Group, 2002, pp. 1–12.

[29] C. A. Miller and R. Parasuraman, "Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 49, no. 1, pp. 57–75, 2007.

[30] T. B. Sheridan and R. Parasuraman, "Human-automation interaction," *Reviews of human factors and ergonomics*, vol. 1, no. 1, pp. 89–129, 2005.

[31] M. S. Martins, "Rules of engagement for land forces: a matter of training, not lawyering," *Mil. L. Rev.*, vol. 143, p. 1, 1994.

[32] B. L. Docherty, *Shaking the Foundations: The Human Rights Implications of Killer Robots*. Human Rights Watch, 2014.

[33] A. Krishnan, *Killer robots: legality and ethicality of autonomous weapons*. Ashgate Publishing, Ltd., 2009.

[34] N. Sharkey, "Death strikes from the sky: the calculus of proportionality," *Technology and Society Magazine, IEEE*, vol. 28, no. 1, pp. 16–19, 2009.

[35] R. C. Arkin, "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture," Mobile Robot Laboratory, College of Computing, Georgia Institute of Technology, 2011.

[36] P. M. Asaro, "Modeling the moral user," *Technology and Society Magazine, IEEE*, vol. 28, no. 1, pp. 20–24, 2009.

[37] J. Canning, "You've just been disarmed. Have a nice day!," *Technology and Society Magazine, IEEE*, vol. 28, no. 1, pp. 13–15, 2009.

[38] A. R. Pritchett, "Simulation to Assess Safety in Complex Work Environments," in *The Oxford handbook of cognitive engineering*, J. D. Lee and A. Kirlik, Eds. New York, NY: Oxford University Press, 2013, pp. 352–366.

