Computational Modelling of the Proportionality Analysis under International Humanitarian Law for Military Decision-Support Systems

Zurek, T.; Woodcock, T.; Pacholska, M.; Van Engers, T.

DOI
10.2139/ssrn.4008946

Publication date
2022

Document Version
Final published version

Computational modelling of the proportionality analysis under International Humanitarian Law for military decision-support systems

Tomasz Zurek[1 [0000−0002−9129−3157]⋆], Taylor Woodcock[1⋆⋆], Magdalena Pacholska[1 [0000−0002−7217−0565]⋆⋆⋆], and Tom Van Engers[2 [0000−0003−3699−8303]]

[1] T.M.C. Asser Institute, R.J. Schimmelpennincklaan 20-22 2517 JN The Hague
[2] Complex Cyber Infrastructure, Informatics Institute, University of Amsterdam
t.zurek, t.woodcock, m.pacholska @asser.nl, T.M.vanEngers@uva.nl

Abstract. The emergent use of AI to aid military decision-making is increasingly contentious, raising questions about the implications for compliance with international humanitarian law (IHL) during targeting. This paper presents a computational weighting model that compares incidental harm with military advantage expected from an attack, as mandated by the IHL proportionality rule. It also outlines preliminary legal and ethical issues around the use of the model, especially those related to explainability and the 'reasonable commander' standard.

Keywords: International law · rule of proportionality · formal model · military AI.

1 Introduction

Under international humanitarian law (IHL) an attack is prohibited if it is expected to cause incidental harm excessive to the concrete and direct military advantage anticipated. The determination of excessiveness requires weighting two incommensurable values and therefore has been subject to extensive debate. Some point out that autonomous weapon systems are unlikely to be able to execute proportionality assessments, and on this basis should not be developed and used for this purpose. Attempts have been made to represent the factors and weighting mechanism in the form of mathematical formulae [18, 15], or a grading scale [3, 10]. advances debates through an initial attempt at the development of a computational model of proportionality assessments carried out for

⋆ Tomasz Zurek received funding from the Dutch Research Council (NWO) Platform for Responsible Innovation (NWO-MVI) as part of the DILEMA Project on Designing International Law and Ethics into Military Artificial Intelligence.
⋆⋆ Taylor Woodcock received funding from the Dutch Research Council (NWO) Platform for Responsible Innovation (NWO-MVI) as part of the DILEMA Project on Designing International Law and Ethics into Military Artificial Intelligence.
⋆⋆⋆ This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101031698.

attacks in armed conflicts. This model constitutes an element of a more complex framework comprising a hybrid decision-support system (DSS) that supports commanders during the targeting process. Key to our proposal is the hybrid approach consisting of data- and knowledge-driven components, namely: (1) the cognitive element, responsible for recognizing objects, predicting results of actions, and evaluating decision options, and (2) the knowledge-based element, responsible for the application of legal rules, the balancing of options, and the decision-making mechanism. The cognitive elements of this system, which are not outlined in this paper, reflect probabilistic functions, related in particular to the distinction between military objectives and unlawful targets [4, 11, 12]. Our model presumes that it will function as part of a broader system, in conjunction with a pre-existing cognitive function capable of distinguishing targets in line with the requirements of IHL. In this paper, we do not aspire to model the proportionality rule (discussed further in section 2), but rather the weighting exercise required by this rule. Analysis of how this weighting exercise can be reflected in a computational format is addressed in section 3. Preliminary legal issues are then raised in section 4. The authors emphasize that the objective of this paper is to present a method of computationally representing the proportionality assessment under IHL. Discussion of legal issues is preliminary and serves as a starting point for further examination into computational representation of the proportionality rule, an issue to be taken up by the authors in future research. The controversy surrounding the use of fully autonomous systems and the unlikelihood of military use of fully autonomous systems for targeting in the foreseeable future warrant conceptualising this model as a DSS. Based on signals intelligence (SIGINT, i.e. sensor data reflecting the circumstances of the analyzed situation), this model would present a qualitatively-supported binary confirmation or denial of a commander-proposed attack plan, allowing commanders to critically analyze this recommendation and make the final decision to attack. Targeting is a highly complex process that involves multiple considerations, evaluations and measures, of which the proportionality assessment is only one. This computational model is not a fully-fledged targeting system. Rather, it presumes that the other dimensions of targeting (such as distinction of targets and taking feasible precautions to spare civilians) have already been undertaken (e.g. by a commander, or perhaps in the future by other computational models). This DSS is envisaged to be used to support proportionality assessments as the final step prior to launching an attack. We hope that the computational model developed below will be a springboard for discussions around the technical and legal complexities surrounding the development of AI-enabled DSS.

2 Proportionality in Attack: Elements and Application of the Rule

Proportionality is an ever-elusive concept with a plurality of meanings across various legal regimes. In contemporary IHL regulating the conduct of military operations in armed conflicts, proportionality is both a principle, and a rule [9].

In its broadest sense, the principle of proportionality [6, 11] concerns all military operations and encompasses both the affirmative obligation to take feasible precautions in attack, as well as the rule of proportionality. All assertions, analysis, and discussion in this paper pertain exclusively to the narrowly construed rule of proportionality in attack, set forth in black-letter IHL as follows [4]: "Launching an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated, is prohibited." While the existence of the rule – often used as a textbook example of the reconciliation of military necessity with considerations of humanity [12] – is widely accepted, the precise boundaries of most of its elements remain debated. Given its scope and purpose, with regard to contested aspects, this section adopts an interpretation supported by the majority of scholarship; the minority position being referenced in the footnotes.

2.1 Comparing Apples and Oranges (the What)

The application of the proportionality rule requires a juxtaposition of incommensurate values, that is, on the one hand 'loss of civilian life, injury to civilians and damage to civilian objects' (often jointly referred to as Incidental Harm (IH))[7], and on the other 'the concrete and direct military advantage anticipated (MA).' Neither IH nor MA is precisely defined, and for neither a closed objective catalogue of relevant considerations exists. That said, the IH is contemporarily understood to encompass the following types of harm: (i) loss of life by civilians[3], not directly participating in hostilities; (ii) physical and severe mental injury [7], [17] to civilians not directly participating in hostilities; (iii) damage to civilian objects, including harm to the civilian use of 'dual-use' objects [13]. Depending on the situation, the assessment of the IH might involve one of the above-mentioned categories or a combination thereof, including the foreseeable reverberating effects of the attack on the civilian population and civilian objects [13, 17, 11]. So construed, IH needs to be weighted against the concrete and direct MA anticipated. MA, understood as "any consequence of an attack which directly enhances friendly military operations or hinders those of the enemy" [3], needs to be distinguished from psychological or political gains which are excluded from the equation [13, 11]. The criteria of "concrete and direct" are commonly understood as requiring the MA to be both identifiable and quantifiable, and not solely of speculative value [17, 11]. While the MA "must be perceived in a contextual fashion",[12] it is generally considered to include the following [4, 13, 11]: (i) ground gained (including not only terrain but also specific facilities or locations); (ii) annihilating or weakening the enemy armed forces; (iii) diverting an enemy forces' resources and attention; (iv) denying the enemy the ability to benefit from the military objective's effective contribution to its military action; (v) one's own force preservation; (vi) lowering the morale

[3] On the special categories of civilians excluded from the scope of IH see: [6] para 5.12.3.2 and [12], 80-82.

of enemy forces; and (vii) protection of civilians (such as foiling enemy attacks directed at civilians).

2.2 The weighting exercise (the How)

Whatever the MA entails under specific circumstances, it should be made for the attack "considered as a whole" and not only from isolated or particular parts thereof [11–13]. That does not mean that the concept should be extended to an entire armed conflict, rather it must remain a distinct operation therein. Furthermore, what matters for the weighting exercise envisioned by the proportionality rule is not the actual MA achieved, but rather its anticipated scope. Opinions on how to account for the likelihood of the MA materializing (or simply put, the attack succeeding) vary [10], but it is generally accepted that "the 'concrete and direct advantage anticipated' is not the value of the target wholly in the abstract but rather its abstract value relative to the likelihood of in fact neutralizing or destroying the object."[5] The degree of uncertainty of IH occurring, however, does not need to be factored in.[4] "[O]nce [IH] is expected, it must be calculated into the proportionality analysis as such; it is not appropriate to consider the degree of certainty as to possible [IH]." [17] Finally, to bar an attack, the disproportion between IH and MA has been understood as having to rise to the level of excessiveness, interpreted as "significant imbalance." [3, 18, 11, 12][5]. In short, under the proportionality rule, the attack is prohibited only if, before launching an attack, a reasonable military commander based on the reliable information available to him/her at the time would consider the IH to be significantly imbalanced in relation to the identifiable and quantifiable MA multiplied by the likelihood of achieving it. Whilst the determination of excessiveness might not be "amenable to a precise or mathematical tabulation"[1], but the following section nevertheless proposes quantitative approximation thereof to support decision-making. After all, proportionality is a zone, not a standard of precision [2].

3 The model

The model presented in this paper is envisioned to form part of a complex hybrid DSS for targeting by the military. The model should reflect the requirements of IHL, especially those related to observing the explicitly formulated targeting standards, and, since the proportionality rule is the object of constant debate, leave open the possibility of subsequent modification. The key assumption underlying our framework is the distinction between the cognitive elements and the reasoning-based elements in the decision-making process. Currently, we witness the rapid development of data-driven approaches, such as deep learning neural networks. Whilst these techniques are suited to drawing inferences from

[4] For an opposing view see [7, 10, 17].
[5] For opposite see [16].

vast bodies of (big) data, they suffer from inherent limitations and drawbacks, in particular relating to explainability and predictability. Although the issue of explainability of data-driven models, as well as the use of explainable surrogate models e.g. in [20], is the object of in-depth research, some tasks are so inherently evidence-based that most current models cannot perform acceptably. A prime example is the proportionality analysis, where excessiveness is assessed to the standard of a reasonable military commander (discussed further below in section 4). Where a data-driven model is used to inform the commander's decision-making, the opacity of the system would inhibit the commander from scrutinising the reasonableness of the proportionality assessment, and justifying an attack afterwards.[6]

The use of a hybrid DSS harnesses the strengths of data-driven approaches in making sense of big data, as well as those of knowledge-driven approaches, which allow for comprehension of the reasoning that produces the system's output and thus may better support compliance with the proportionality rule under IHL. As such, the model presented here – comprising the knowledge-driven component of the DSS - supports both model and decision explainability, by facilitating ex ante review by the commander of how the system produces output, allowing for assessment of the legality of the attack by subjecting the proportionality decision to the standard of a reasonable commander, as well as providing explanations of why particular attacks were considered proportionate ex post. On this basis, the authors argue that the DSS for military targeting should be composed of at least two parts:

1. The first component is responsible for the cognitive aspect of decision-making, by which we understand the process of interpreting the input signals, predicting the results of actions, evaluating particular decisions or actions in the light of moral values etc. This part of the system might be constructed using a data-driven (Machine Learning, or ML-based) paradigm, and would contain: a decision option generation module (the mechanism generating the set of possible decision options); the prediction module (the mechanism predicting the result of a given decision in the actual circumstances); and the evaluation module (the mechanism evaluating the expected consequences of a particular decision made in particular circumstances in light of pre-determined values).
2. The second component is responsible for the reasoning process and also includes the weighting exercise. This part of the system should rely on the knowledge-based paradigm.[7].

[6] While the commander need not understand the logic or specificities of the DSS, and most military systems are in fact highly technical, the commander should be able to review and critically evaluate the reasons given by the system for a particular determination of proportionality.

[7] In the literature the reasoning process is sometimes divided into a so-called epistemic part (determining what is true) and a practical reasoning part (determining what to do) Here we assume that both of those parts are elements of the reasoning part, but we are aware that the epistemic part can be correlated with the cognitive part in a complex way. This topic we leave for future research.

This pairing allows us to harness the advantages of both AI paradigms. ML-based functions can be used for the tasks for which rules or principles of conduct are hard to express and require induction from complex case descriptions (evaluating the decisions in the light of different values, etc.), whilst the knowledge-based mechanisms are used for the processes requiring transparency, explainability and predictability. This paper focuses on a particular element of the reasoning component, namely the weighting exercise. Hence, we assume that the cognitive part of the system is already prepared and that the system can: (i) distinguish a decision space and lawful from unlawful targets; (ii) predict the results of decisions with their probabilities; and (iii) evaluate those results in the light of different values.

3.1 Weighting exercise

How is the weighting exercise required by the proportionality rule performed in practice? Suppose that a commander, on the basis of SIGINT, the general circumstances of the situation, commonsense, specialist knowledge and training, and experience identifies a set of courses of action (COA) and predicts their expected results. With such knowledge, every option is evaluated in light of the anticipated MA and expected IH, which is compared to assess possible excessiveness. Those options which are considered to violate the proportionality rule, i.e. where the expected IH is excessive with respect to the anticipated MA, held to the standard of a reasonable commander, are eliminated.

In order to represent this proportionality weighting exercise, the model must relate and compare two different dimensions: the level of MA with the level of IH. Both evaluations are highly contextual and controversial from a legal and moral point of view. It is not acceptable to strictly compare the number of hurt combatants, destroyed military objects, etc. to the number of hurt civilians, destroyed civilian objects etc. Hence, in order to make those two dimensions comparable, we have to introduce an intermediate notion representing a kind of abstraction of the input data.

Although values are the object of debate in many research disciplines, we assume a very general definition of value taken from [21]:

Definition 1 *Value is an abstract (trans-situational) concept which allows for the estimation of a particular action or a state of affairs and influences one's behavior. Consequently, on the basis of such a definition, we assume that particular values can be satisfied to a certain degree.*

Values can be seen as abstractions of particular situations and they can be promoted (or satisfied) to a particular extent by those situations (state of affairs, actions, etc.). Our model represents the proportionality analysis through comparison of the degree to which decision options support particular values.

In order to represent the weighting exercise ordinarily conducted by a commander, we assume some basic concepts:

Definition 2 *Input data is understood as everything that can influence the decision and allow for the prediction of the results of a decision, including SIGINT and the general circumstances of the analysed situation. Input data will be denoted by a vector* X[8].

Definition 3 *By decision option we denote a pair containing* X *(input data vector) and a particular decision* d *(obtained with the use of option generation module) available in the circumstances described by a given input data vector. Let* $S = \{s_1, s_2, ...\}$ *be a set of all available situations, i.e. all possible decisions to be made in given circumstances.*

In defining value, we assume that this may be satisfied to a certain level by a particular state of affairs or action. By $v_x(s)$ we denote the level of satisfaction value v_x by decision option s[9].

In order to obtain a value satisfaction level, we introduce a function which, on the basis of a given situation, returns the level of the value's promotion:

Definition 4 *Suppose a function* $\Phi_{v_x}(s)$ *which returns the value satisfaction level* $v_x \in V$ *by a situation* $s \in S$*. For example,* $\Phi_{v_x}(s_1) = v_x(s_1)$ *By* $\Phi = \{\Phi_{v_x}, \Phi_{v_y}, ..\}$*, where* $v_x, v_y, ... \in V$ *we denote a set of functions.*

Functions from the set Φ represent the process of evaluating the expected results of a given decision and are elements of the evaluation module in the cognitive part of the system.

Although in this paper we do not introduce any particular function, and we assume the functions from set Φ, below we propose how such a set of functions could be obtained: Since the proportionality rule assumes that the weighting exercise should be performed on the basis of *potential* IH and MA, a mechanism is required that can predict the result of a decision and one that, on the basis of predicted results, can assign the levels of satisfaction of the analyzed values to particular results of the decision.

In order to represent the 'weighting exercise' two values are necessary: the *IH* (civilian life, health, infrastructure, etc.) and *military advantage*. Let V will be the set of values and $V = \{v_{IH}, v_{militaryAdv}\}$.

Note that since value, by definition, is something positive, the value v_{IH} should be inversely proportional to the level of harm to civilians.

Obviously, the value-support level of a particular result is not easy to estimate; as noted earlier, we cannot say that if the number of enemy combatants killed is greater than the number of protected civilians killed, then the IH is not excessive in comparison to the MA and the attack is lawful under IHL. However, these and other factors influence the value-support level.

Once the COA-specific value-satisfaction levels are obtained, we perform the weighting exercise:

[8] In order to preserve the generality of the model, we do not impose any particular format, type, or dimensionality of the data.

[9] We are not going to impose here any particular scale of the levels of satisfaction of values, but for the sake of this paper we assume that this is represented by a number.

Since for a particular decision option s_1, the value-promotion level $v_{IH}(s1)$ is inversely proportional to the level of IH, the level of satisfaction of $v_{militaryAdv}(s1))$ is proportional to the MA, then:

Definition 5 *IH will be excessive in comparison to the MA if the level of promotion of values life of civilians and civilian objects will be lower than military advantage:*

$v_{militaryAdv}(s_1) > v_{IH}(s_1) \Rightarrow excesive(s_1)$

Since the levels of satisfaction of v_{IH} and $v_{militaryAdv}$ are obtained on the basis of ML-based mechanisms, then the model presented here may be considered too general to be sufficiently informative and explainable. Therefore, we have to extend our system by introducing a more complex value system.

In order to extend our system, following other researchers (for example [19]), we assume that values have a hierarchical character, i.e. there are some more specific values which make up a more general value.

On the basis of the above we assume that v_{IH} is influenced by:

- $v_{civilianLife}$ inversely proportional to the loss of life by civilians not directly participating in hostilities. The level of satisfaction of this value can be obtained by a function $\Phi_{v_{civilianLife}}$
- $v_{civilianHealth}$ inversely proportional to injury to civilians not directly participating in hostilities (including physical and mental health). The level of satisfaction of this value can be obtained by a function $\Phi_{v_{civilianHealth}}$
- $v_{civilianObjects}$ inversely proportional to damage to civilian objects. The level of satisfaction of this value can be obtained by a function $\Phi_{v_{civilianObjects}}$

Let us denote by a set $V_{IH}(s_1)$ a set of levels of satisfaction of all components of value v_{IH} by decision option s_1.[10]
Similar to the above, $v_{militaryAdv}$ is influenced by:

- v_{ground} ground gained. The level of satisfaction of this value can be obtained by a function $\Phi_{v_{ground}}$.
- $v_{disruptingEnemyActiv}$ disrupting enemy activities. The level of satisfaction of this value can be obtained by a function $\Phi_{v_{disruptingEnemyActiv}}$
- $v_{divertingEnemyResources}$ diverting an enemy force's resources and attention. The level of satisfaction of this value can be obtained by a function $\Phi_{v_{divertingEnemyResources}}$.
- $v_{denyingEnemyBenefit}$ denying the enemy the ability to benefit from the military objective's effective contribution to its military action (or simply complete or partial destruction of enemy military targets). The level of satisfaction of this value can be obtained by a function $\Phi_{v_{denyingEnemyBenefit}}$.
- $v_{OwnForcePrteservation}$ own force's preservation. The level of satisfaction of this value can be obtained by a function $\Phi_{v_{OwnForcePrteservation}}$
- $v_{LoweringEnemyMorale}$ lowering the morale of enemy forces. The level of satisfaction of this value can be obtained by a function $\Phi_{v_{LoweringEnemyMorale}}$

[10] Note that a set is denoted by capital letter V.

– $v_{CivilianProtection}$ protection of civilians. The level of satisfaction of this value can be obtained by a function $\Phi_{v_{CivilianProtection}}$

Let us denote by a set $V_{militaryAdv}(s_1)$ a set of levels of satisfaction of all components of value $v_{militaryAdv}$ by decision option s_1.

How can we model the influence of these considerations on the more general value (MA)? Note that these do not necessarily have an equal influence on the general satisfaction level. In order to represent the differences in importance of particular values we assume a set of weights:

Definition 6 *Let* $\Psi = \{\Psi_{civilianLife}, \Psi_{civilianHealth}, \Psi_{civilianObject,...}\}$ *be a set of weights related to particular values. Each weight is represented by a number, representing the relative importance of a value. Greater weight represents greater importance.* Ψ_{IH} *denotes weights of components of value* v_{IH} *and* $\Psi_{militaryAdv}$ *denotes weights of components of value* $v_{MilitaryAdv}$*.*

On the basis of the above, we can assume that the level of satisfaction of value v_{IH} is equal:
$v_{IH} = v_{civilianLife} * \Psi_{civilianLife} + v_{civilianHealth} * \Psi_{civilianHealth} + v_{civilianObjects} * \Psi_{civilianObjects} = \Sigma_{i \in V_{IH}} v_i * \Psi_i$
Analogically, the value satisfaction level $v_{militaryAdv}$ is equal:
$\Sigma_{i \in V_{MilitaryAdv}} v_i * \Psi_i$
The value satisfaction levels v_{IH} and $v_{militaryAdv}$ resulting from a decision are then weighted sums of their components. Such a model increases explainability of the system by introducing an additional layer of decision evaluation. By using a weighted sum, we assume the linear influence of specific values on the more general one. This is admittedly a simplification, but one that the authors make for the sake of simplicity and clarity. In order to introduce a more sophisticated way of representing weights, we can use the functions of weights introduced in [22] which can be seen as a generalisation of weights introduced in this paper allowing for representing nonlinear weights.

3.2 Adding probabilities

As mentioned in Section 2.2., there are two main approaches to considering the uncertainty of the results of a decision. Unlike [18], we present a model of the approach in which only the uncertainty of MA is taken into consideration. Whilst this approach is a dominant one in legal scholarship, from a computational point of view it can be seen as controversial, as it compares the ostensibly certain IH with uncertain MA.

Such an approach is only seemingly controversial. The key point is in the understanding of the decision's uncertainty. We can distinguish here at least two levels of uncertainty: one level is the uncertainty whether a particular decision would bring about a desired consequence, and the second level is the uncertainty of the input data and basic assumptions. The proposed approach takes into account the likelihood that certain assumptions will hold true, rather than the probability of such. For example, if there is a decision to kill a high-level leader of

an organized armed group party to a conflict, then the uncertainty will relate to the chance that this leader will be at the expected location, not the probability of success of the action. Note that even if this leader will not be at the place and the action would not bring about the desired results, the IH will be the same.

The probability of the action's success and the probability of occurrence of a particular IH should be carefully discussed; the authors leave this for future exploration.

If we assume that by π_{s_1} we denote the certainty that the assumptions of decision s_1 are true, then the weighting exercise can be expressed by:

$v_{militaryAdv}(s_1) * \pi_{s_1} > v_{IH}(s_1) \Rightarrow excesive(s_1)$

Moreover, unlike in [18], we do not assume that we can estimate the probability of damage to every single actor and object involved in the military operation, but rather introduce one general level of uncertainty. Our approach is less nuanced than Schmitt's et al. one, but has an additional benefit as being much more feasible from a technical perspective.

4 Discussion of preliminary legal issues and conclusions

Both the content and the application of the proportionality rule are subject to extensive debate in scholarship and practice. This debate reflects the complexity of this rule and why discussions about automating the proportionality weighting exercise are fraught. Nevertheless, this paper seeks to prompt further discussions about the feasibility of computational representation of the proportionality analysis. The authors reiterate that this brief discussion only introduces some of the relevant legal issues to be explored in future research.

Within the model presented in this paper, values function as an intermediate concept representing an abstraction of the targeting situation and the concept connecting ML and knowledge-based parts of the system. This more readily allows for the weighting exercise inherent in the proportionality assessment to be represented in a computational format, as it facilitates comparison of value satisfaction levels rather than of particular considerations of IH and MA, which can be difficult to compare from both a legal and technical perspective. In the broader hybrid system, the value satisfaction levels are based on the ML component, which remains a black box. Nevertheless, this operates in conjunction with the knowledge-based component, which explicitly calculates the levels of satisfaction of IH and MA and makes a comparison of the two in an intelligible way, reflecting the explainability inherent in such a hybrid system.

Given the nature of the proportionality rule, it is widely accepted that there is no objective amount of IH that can be considered disproportionate; rather, proportionality assessments must be made on a case-by-case basis in light of the prevailing circumstances surrounding the attack, weighting both IH and MA. As the rule is not results-based but engages the ex ante decision-making process of the commander, proportionality assessments are held to the standard of the reasonable military commander. This standard reflects the 'experience, training

and understanding of military operations' ordinarily possessed by an individual in this position [14]. Whilst some argue that this is a purely subjective standard, and proportionality assessments no doubt entail a number of subjective judgments, compliance with the proportionality rule is better thought of as necessitating a 'semi-objective' standard [11]. Such a standard captures that determinations of proportionality by commanders in attacks 'must be objectively reasonable... based on the actual information held by the attacker' [11]. It allows for the fact that, when faced with identical circumstances, different military commanders may come to different proportionality decisions. The proportionality rule thus allows for a number of different findings that an attack will be proportionate along a spectrum of reasonableness, reflected in the statement above that proportionality is a zone. This conception of the reasonable commander standard raises questions about whether a commander relying on an AI-enabled DSS for proportionality analyses can continue to act objectively reasonably in conducting proportionality assessments. In practice, reasonableness requires that a commander is expected to take into account all reliable information available, as well as training and past experience, in light of the prevailing circumstances in which the attack is carried out. The model presented in this paper could support a proportionality analysis by increasing a commander's situational awareness. Nevertheless, a potential challenge that could impede the commander adhering to the standard of reasonableness is that the system may be limited by difficulties in dealing with the dynamic and unpredictable context of the battlefield. This justifies conceptualizing the model as a DSS, rather than a decision-making system, as a commander should remain involved and ultimately take the final decision in proportionality assessments.

An open issue for future research is how to mitigate potential cognitive biases that might arise with respect to the use of such a model by commanders to ensure the output of the model can be subjected to critical evaluation, in light of the complexity and scale at which the DSS operates. A further unexplored question is whether or not there exists an inherent moral dimension when commanders conduct proportionality assessments that could be accounted for in the model. Proportionality assessments necessarily entail value-laden decision-making. One perspective is that the inherent morality of the balancing exercise required by the proportionality rule mandates the valuation of innocent human lives such that what is at stake must be fully accounted for in the decision-making process, meaning this a task that must be conducted by humans.[11] An alternative view is that although computational systems cannot fully realise the value of human life, the ML-based evaluation tools may be trained on the basis of the set of data which has been labelled by humans, who presumably can assess the value of human lives, thus allowing for a comprehensive evaluation of proportionality in the light of MA and IH. In future work, the authors intend to engage in a more comprehensive discussion of the problem of uncertainty of the anticipated results of the decision, in-depth analysis of the legal aspects of the use of hybrid

[11] See generally the debates about autonomous weapons and the need for human control in the lethal use of force, e.g. [8].

DSS for proportionality assessments and, eventually, experimental analysis of the system, especially its cognitive components.

References

1. Prosecutor v. Strugar, IT-01-42, prosecutor's pre-trial brief pursuant to rule 65 ter (e) (i), 27 august 2003, para. 152 (2003)
2. Public Committee against torture in Israel v. the Government of Israel, HCJ 769/02, para. 58 (2005)
3. HPCR manual on international law applicable to air and missile warfare, produced by the program on humanitarian policy and conflict research at Harvard University (2009)
4. ICRC, Customary IHL database, rule 14 (2009)
5. Prosecutor v. Gotovina et al., it-06-90-t, prosecution's public redacted final trial brief, 2 august 2010, para. 549 (2010)
6. Office of the General Counsel, U.S. Dep. of Defense Law of War Manual (2015)
7. Proportionality in the conduct of hostilities: The incidental harm side of the assessment. Chatham House, December 2018 (2018)
8. Asaro, P.: On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. International Review of the Red Cross **94**(886), 687–709 (2012). https://doi.org/10.1017/S1816383112000768
9. Bartels, R.: Dealing with the principle of proportionality in armed conflict in retrospect: The application of the principle in international criminal trials. Israel Law Review **46**(2), 271–315 (2013). https://doi.org/10.1017/S0021223713000083
10. van den Boogaard, J.: Proportionality in international humanitarian law. Ph.D. thesis, UvA-DARE (2019)
11. Cohen, A., Zlotogorski, D.: Proportionality in International Humanitarian Law: Consequences, Precautions, and Procedures. Oxford University Press, (2021)
12. Dinstein, Y.: The principle of humanity. In: Mujezinović Larsen, K., Guldahl Cooper, C., G., N. (eds.) Searching for a 'Principle of Humanity' in International Humanitarian Law. Cambridge University Press (2013)
13. Gisel, L.: The principle of proportionality in the rules governing the conduct of hostilities under international humanitarian law, ICRC International Expert Meeting 22-23 June 2016 (2016)
14. Henderson, I., Reece, K.: Proportionality under international humanitarian law (ihl): The 'reasonable military commander' standard and reverberating effects. Vanderbilt Journal of Transnational Law (3) (2018)
15. Humphrey, A., See, J.E., Faulkner, D.R.: A methodology to assess lethality and collateral damage for nonfragmenting precision-guided weapons. ITA Journal pp. 411–419 (2008)
16. Sandoz, Y., Swinarski, C., Zimmermann, B.: Commentary on the additional protocols of 8 june 1977 to the geneva conventions of 12 august 1949. ICRC (1987)
17. Schmitt, M.N.: Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations. Cambridge University Press, 2 edn. (2017)
18. Schmitt, M.N., Schauss, M.: Uncertainty in the law of targeting: Towards a cognitive framework. Harvard National Security Journal **10**, 148–194 (2019)
19. Schwartz, S.: Value priorities and behavior: Applying a theory of integrated value systems **8**, 119,144 (2001)

20. Schwartzenberg, C., van Engers, T.M., Li, Y.: The fidelity of global surrogates in interpretable machine learning. In: BNAIC/BeneLearn 2020 Proceedings (2020)
21. Zurek, T.: Goals, values, and reasoning. Expert Systems with Applications **71**, 442 – 456 (2017). https://doi.org/http://dx.doi.org/10.1016/j.eswa.2016.11.008
22. Zurek, T., Mokkas, M.: Value-based reasoning in autonomous agents. International Journal of Computational Intelligence Systems **14**, 896–921 (2021). https://doi.org/https://doi.org/10.2991/ijcis.d.210203.001