2017 9th International Conference on Cyber Conflict
Defending the Core
H. Rõigas, R. Jakschis, L. Lindström, T. Minárik (Eds.)
2017 © NATO CCD COE Publications, Tallinn

Update to Revolving Door 2.0: The Extension of the Period for Direct Participation in Hostilities Due to Autonomous Cyber Weapons

Tassilo V. P. Singer
Research Associate
Public and Administrative Law, Public International Law,
European and International Economic Law
University of Passau, Germany
tassilo.singer@gmx.de

Abstract: The rule concerning the direct participation in hostilities (DPH) by civilians is one of the most controversial rules within the law of armed conflict. While civilians are generally guaranteed protection by the principle of distinction, DPH provides for a loss of protection 'for such time as they take a direct part in hostilities'. This temporal component of DPH poses significant challenges in light of the use of autonomous cyber weapon systems (ACWS), as the active part of the civilian can be reduced to the second of activation.

Autonomy in this context means that the ACWS is able to operate without any human control, rendering human influence on the actions impossible. As soon as the ACWS is fully self-operating, the common interpretation based on the original wording of the DPH-rule would suggest that DPH is no longer possible. Consequently, a civilian hacker using ACWS cannot lawfully be attacked after control of the system has been relinquished even if damage occurs or the attack is recognised later. As a result, civilian hackers are legally privileged without any discernible justification. In order to remedy this unsatisfactory situation, this article suggests an extension of the relevant period of time for DPH to the whole period of the operation of the ACWS to solve the 'Revolving Door 2.0' problem. It is further submitted that concerns that such an extension would unduly broaden the scope of the DPH-rule can be met by the requirement that, additionally, the regular cumulative criteria for DPH have to be fulfilled.

Keywords: *direct participation in hostilities, loss of protection as a civilian, timeframe of direct participation in hostilities, autonomous cyber weapon systems*

1. INTRODUCTION

The protection of civilians under IHL ceases following their direct participation in hostilities (DPH), according to Article 51(3) AP I. However, the protection is regained after return or the end of the activity.[1] This transition regularly poses problems in practice, but the revolving door of protection and civilian DPH is even more critical when the use of autonomous cyber weapons (ACWS) is involved.

Imagine critical parts of a state's military networks like unmanned combat aerial vehicle (UCAV) control have been hacked, leading to the UCAVs being turned against the state's own troops. After the type and source of the attack has been identified, it turns out the malware used has been operating for months without any human control. In such a case, the predominant legal interpretation is that the perpetrator must not be attacked as a civilian directly participating in hostilities. According to the common understanding, the loss of protection of a civilian is terminated after the last moment of control, even if the attack continues or its effects occur later.[2] This article will focus on the discrepancy between current interpretation and reality, and argues towards an extension of the time period of DPH as a solution to the dilemma posed by the use of ACWS by civilians.

2. DIRECT PARTICIPATION IN HOSTILITIES

According to the principle of distinction set out in Articles 48 and 51(2) AP I,[3] civilians may not be the object of an attack, and such an attack does not form a part of the military advantage for a conflict party.[4] However, if civilians decide to take up arms and fight against a conflict party, the balance between humanity and military necessity[5] would be negatively affected. It would be unacceptable to grant protection to civilians while they attack combatants, but the latter may not counter the attack due to the general rule of distinction. Therefore, the rule of Article 51(3) AP I provides for a loss of protection for the period of time during which civilians directly take part in hostilities to sanction the participation in hostilities by an originally protected person.[6] However, the interpretation of – and state practice on – the rule for DPH is fairly controversial.[7]

As the meaning of DPH is neither specified in treaty law nor clarified by sufficient state practice or international jurisprudence,[8] the rule is open for interpretation under Article 31 of the Vienna

1 *ICRC, Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law*, Nils Melzer (ed.) (May 2009), 67.

2 Marco Roscini, *Cyber Operations and the Use of Force in International Law* (1st edn OUP 2014) 209; Heather Harrison-Dinniss, 'Participants in Conflict – Cyber Warriors, Patriotic Hackers and the Laws of War' in Dan Saxon (ed), *International Humanitarian Law and the Changing Technology of War*, 251-278, 274-276.

3 *Legality of the Threat or Use of Nuclear Weapons* (Advisory Opinion) [1996] ICJ Rep 226, para. 78.

4 Yoram Dinstein, *The Conduct of Hostilities under the Law of International Armed Conflict* (2nd edn, 2010) 123-125.

5 Id., 4-5.

6 Michael N. Schmitt (ed.), *Tallinn Manual on the International Law Applicable to Cyberwarfare* (Cambridge University Press 2013), 104-105, 118-119; Roscini (n 2), 203.

7 ICRC (n 1); Program on Humanitarian Policy and Conflict Research at Harvard University, *Commentary on the HPCR Manual on the International Law Applicable to Air and Missile Warfare (hereinafter HPCR Manual)*, Rule 28, 119.

8 ICRC (n 1), 41.

Convention.[9] Direct participation derives from the notion of 'taking no active part in the hostilities' of Common Article 3 GC I–IV, so active and direct participation share a common background.[10]

Article 51(3) AP I has three elements. First, the civilian has to take a direct part, which is an element of directness and or immediacy.[11] It may be better to choose a narrow interpretation of the directness to increase the protection of civilians.[12] This can be countered, however, by arguing that a wide interpretation of the rules would encourage innocent civilians to stay away from the hostilities as far as possible.[13] The second element is 'in hostilities', and indicates a certain nature or threshold of harm. Finally, 'for such time' implies a time-frame[14] and indicates only a temporary loss of protection. The content of this period is the decisive legal question concerning autonomous cyber programs. If one interprets the wording strictly, it only refers to the actual conduct of the civilian. After this period of direct participation, a civilian generally regains protection.

According to the ICRCs study on DPH, three criteria have to be met cumulatively under Article 51(3) AP I: the threshold of harm; a causal link between the act and the intended or inflicted harm; and the belligerent nexus, meaning the act must directly be related to the hostilities.[15] The threshold of harm means the required 'specific acts of war which by their nature or purpose are likely to adversely affect military operations or capacity'[16] or to cause a certain damage to protected targets.[17] Besides potential attacks on military targets,[18] it remains unclear whether it is necessary that damage to protected civilian targets actually occurs, or if potential danger to civilian targets suffices.[19]

On direct causation, the ICRC study determines that either the act has to be one causal step between act and effect,[20] or the relevant contribution of the civilian must constitute an integral part of a coordinated military operation.[21] However, the direct character is not precluded if the 'effects occur' with a delay, meaning 'some time after the malware is inserted'.[22] Also, due to the 'integral part' requirement, group-based actions which alone would not suffice for DPH can be taken into account if these are connected to a specified operation.[23]

The third criterion, the belligerent nexus, means that the relevant act has to be linked 'in some direct way [...] to the armed conflict'[24] and consequently has to be 'specifically designed to

9 Vienna Convention on the Law of Treaties (VCLT), 1155 UNTS 331.
10 *The Prosecutor v. Akayesu*, ICTR, case no. ICTR-96-4-T (1998), paras. 175, 182, 582; *Public Committee Against Torture in Israel et al v The Government of Israel et al*, HCJ 769/02, Supreme Court, 11 December 2005 (hereinafter *Supreme Court*), para 34; ICRC (n 1), 43.
11 *Supreme Court* (n 10), para. 35.
12 Id., para. 34.
13 Ibid.
14 Michael N. Schmitt, 'Deconstructing Direct Participation in Hostilities: The Constitutive Elements' (2010) International Law and Politics, Vol. 42, 728-729, 738; Supreme Court (n 10), para. 38.
15 ICRC (n 1), 46; 47 ff, 51 ff, 58 ff.
16 Id., 47.
17 Ibid.
18 Compare Schmitt (n 14), 715-716.
19 See Roscini (n 2), 204-205.
20 Id., 206.
21 ICRC (n 1), 51.
22 Roscini (n 2), 207.
23 *HPCR Manual* (n 7), Rule 29, 120, para. 3; id, 207.
24 Schmitt (n 14), 735.

directly cause the [...] threshold of harm in support of a party and to the detriment of another'.[25] The design of the act does not depend on the subjective intent of the actor.[26] However, the ICRC study suggests that:

> the conduct of a civilian, in conjunction with the circumstances prevailing [...], can reasonably be perceived as an act designed to support one party [...] by directly causing [...] harm to another party.[27]

These findings have been criticized for various reasons.[28] First, the term 'threshold' in the first criterion was considered misleading because the issue should be the nature of harm and not a set threshold.[29] Second, the outcome that one causal step is necessary can also be objected to, as it would be impractical and contradictory to ask for an immediate consequence of the act and allow for a temporal distance.[30] Third, the requirement of direct causation was welcomed, but seen as used too restrictively. Schmitt holds that the integral part criterion[31] should not only be limited to coordinated military operations, but should also extend to individuals.[32] Thus, the act of the civilian must be more generally 'an integral part of the conduct that adversely harms one party and benefits the other party to a conflict'.[33] Concerning the belligerent nexus, one can criticise that the act must be linked directly to the causation of such harm and the causation of the benefit, and Schmitt suggests an alternative approach focusing on a link of the act either to the support or the detriment of one party.[34]

Even if there is disagreement concerning the exact content of the three criteria recommended by the ICRC, the critics agree that at least the three constitutive criteria can be applied[35] and have to prevail independently of the exact legal content in every situation of DPH by a civilian.

For the discussion of the particular problem of ACWS, an international armed conflict has to prevail[36] and a civilian has to act. A negative definition of who should be considered a civilian can be found in Article 50 AP I.[37] Also the discussion regarding organised armed groups (and a continuous combat function in this context)[38] can be disregarded.

25 ICRC (n 1), 58.
26 Id., 59; Schmitt (n 14), 735-736.
27 ICRC, (n 1), 63-64.
28 Compare as overview of critical views on the ICRC study by Boothby, Schmitt, Watkin and Hays Parks and a response by Melzer: *New York Journal of International Law and Politics* 42 (2009-10).
29 Schmitt (n 14), 716.
30 Id., 728.
31 ICRC (n 1), 51.
32 Schmitt (n 14), 729 ff.
33 Id., 739.
34 Id., 736.
35 Id., 738; *Tallinn Manual* (n 6), 119.
36 The rule is viewed as customary international law and can be applied in non-international armed conflicts: Jean-Marie Henckaerts, Louise Doswald-Beck, *Customary International Humanitarian Law*, Vol I, Rule 5, 17; *Supreme Court* (n 10), para. 38.
37 Referring to Article 4 A (1),(2), (6) of GC III and Article 43 AP I.
38 Compare Roscini (n 2), 200 f.

3. DIRECT PARTICIPATION IN HOSTILITIES BY THE USE OF AUTONOMOUS CYBER WEAPONS

A. Technical Peculiarities

Under these legal circumstances, a big challenge is posed by new cyber tools like conditioned and delayed code, scripts or malware[39] which have autonomous behaviour. After insertion or after the start of the program's working process, the malware's capabilities may include self-guidance, self-reproduction, and even redefinition and adaptation. Many variations of such software are conceivable.[40] All of these have in common that they require only a very short period of human interaction with the software in order to start process. Human participation can be reduced to pressing 'enter', to sending an email, or to integrating the malware somewhere. From the moment when the malware operates independently and is not controllable by a human any more, a cyber tool can be called autonomous.

Software and programs can calculate extremely quickly and thereby make decisions in less than a fraction of a second. In the last decade, the speed of data transmission via cable or satellite has increased enormously, correlating with a growing interconnectivity worldwide. As the amount of software and code has expanded exponentially, so have their weaknesses, as the potential contact surface for attacks has increased.[41] Tools for encryption of data and communication have also spread out and become more sophisticated, further complicating retracing and attribution.[42] These factors, combined with certain malware which itself can act independently without any human control, pose significant challenges for detection and set the factual framework for the envisaged problem.

Malware can contain multiple tools with sub-functions (as weapons) separately and independently from one another comparable to a weapon system[43] in the original sense. The exact qualification of a cyber tool as a weapon, a weapon system, or more generally as a mean of warfare does not matter for the legal problem discussed here as long as the cyber tool has the potential to cause the required damage to constitute a DPH. As the law of armed conflict applies to every weapon system,[44] this is also true of the DPH-rule in the context of ACWS.

B. Transfer of the Constitutive Criteria to ACWS

The constitutive criteria have to be transferred to the use of ACWS by a civilian. In the aforementioned situation, the manipulation of the UCAV control systems by an ACWS leads to friendly fire by the UCAVs. As military targets have been attacked, the required threshold of harm is met. Arguably, the manipulation of the military control system could be considered sufficient in itself, even if no material damage is caused but military operations have been hampered.[45] The same would apply if an autonomous program was able to penetrate a military

39 Malicious software, compare *Tallinn Manual* (n 6), Glossary.

40 Compare CERT-UK, *Common Cyber Attacks: Reducing the Impact*, 2015, last viewed 22.12.2016, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/400106/Common_Cyber_Attacks-Reducing_The_Impact.pdf ; US-CERT, National Cybersecurity and Communications Integration Center, *DDos Quick Guide*, last viewed 22.12.2016, available at: https://www.us-cert.gov/sites/default/files/publications/DDoS Quick Guide.pdf.

41 William A. Owens, Kenneth W. Dam and Herbert S. Lin, *Technology, Policy, Law and Ethics* (1st edn., National Academies Press 2009), 80-108.

42 Ibid.

43 For differentiation compare: *HPCR Manual* (n 7), Means: Rule 1 (t); Weapon: Rule 1 (ff).

44 *Nuclear Weapons* (n 3), para.86.

45 *Tallinn Manual* (n 6), 119.

communication network and interrupt the communication links by continuously changing the IP-allocation of the central communication hub.

A civilian starts an ACWS by using a computer and software to send or activate the tool. This act by the civilian has to be viewed as the last causal step before the ACWS causes the damage. The ACWS, if programmed properly, will run continuously until it reaches its goal without any means for the human to control it. An ACWS checking the military network for the UAV control and finally using an exploit in the code to manipulate or take over control is comparable to a fired missile, even if there is a much longer interval upon effect. An additional causal step is not possible or necessary. Even if this view is not shared, the act of activation (and possibly the preparation of the ACWS) by the civilian is at least an integral part of the specified operation and thus constitutes the required direct causation.

Finally, the act of the civilian to start the ACWS to harm the UAV control is at least designed to harm the adversary party in the conflict. Generally, the acts of the ACWS have to be viewed as one complex in relation to the belligerent nexus. If this complex is designed to either benefit one party or directly cause harm to the other, the nexus is given.

C. Acts of Participation in Hostilities

The relevant acts of a civilian directly participating are often the preparation, the act of direct participation itself, and the return.[46] Transferred to autonomous cyber weapon systems[47] the elements could look as follows.[48] The ACWS first has to be prepared, a process which could consist of coding or at least an acquisition process. During the preparation a target or a framework for potential targets has to be set. It is problematic to distinguish between having just to prepare the software without further participation, and the necessary integral part, meaning the preparation of a malware with direct causation.

During the preparation the civilian cannot generally be lawfully attacked.[49] The question then is when exactly the civilian becomes a legitimate target in the latter case. One could draw the line and find a sufficient act of preparation as soon as the software is shaped for implementation or specified[50] in order to cause a certain form of damage (e.g. the malware is shaped to the specifics of a common control-software for system-processes in military networks).[51] Ergo, the software has to have the potential to damage relevant targets.[52] The process has to be developed to the point that the damage can occur even if the code is not yet sophisticated enough to avoid any detection. However, the monitoring of someone preparing an ACWS is hardly ever possible to accomplish given the speed, the internationality, and the interdependence of cyberspace.

46 ICRC (n 1), 65; *HPCR Manual* (n 7), 118, Rule 28-29; *Tallinn Manual* (n 6), 121.

47 Compare DPH and the use of unmanned systems: Dorota Banaszewska, 'Kombattanten und Zivilisten weit weg vom Schlachtfeld' in Robert Frau (ed.), *Drohnen und das Recht* (Mohr Siebeck 2014).

48 See, for a different description for types of actions: Roscini (n 2), 204.

49 Compare ICRC (n 1), 53, 68; *HPCR Manual* (n 7), 120, para. 3; Bill Boothby, ' "And for such time as": The Time Dimension to direct participation in Hostilities', [2010] *International Law and Politics* 42, 748, 752; Roscini (n 2), 201 (concerning continuous combat function), 207-209.

50 See ICRC (n 1), 52-54; Roscini (n 2), 208.

51 It is necessary to differentiate between military targets, where the likelihood definitely suffices and civilian targets which may require a damage to occur. Roscini (n 2), 205-206.

52 See ICRC (n 1), 47; Roscini (n 2), who refers to 'objective likelihood that the act will result in such harm', 206.

D. The Problem of DPH Using ACWS – Revolving Door 2.0

The act of participation itself might be observable, e.g. a civilian shooting at a military convoy with a gun. Looking at the use of ACWS, the situation is far more difficult. The period wherein the civilian is active could be only a few seconds and the necessity to act is often just the process to let the ACWS start its self-guidance; the effect and damage usually occur later. In practice, the ACWS becomes recognisable to the victim only in this moment due to detection and retraceability problems.[53]

Unfortunately, due to the regular time delay the action of the perpetrator will commonly have been over for weeks or even months. In such a situation, the civilian would not directly participate in hostilities any more, and consequently would enjoy legal protection again. A comparable situation is the revolving door problem known from the war against insurgents in Afghanistan and articulated in the 'farmer by day, fighter by night' problem.[54]

However, in this case the problem is raised to a new level. Even if all available surveillance tools are used, the identification of the origin and the proof of the use of ACWS itself is much more difficult in this short amount of time. Because the circumstances in the context of ACWS are even more challenging and the technical and legal limits even tighter, the problem could be called revolving door 2.0.

The decisive factor is the time-frame 'for such time' which is a prerequisite of the legal exception of DPH. In a case concerning DPH, the Israeli Supreme Court quoted the AP I commentary, proposing that the time-frame should neither be interpreted too narrowly nor too widely.[55] An international group of experts found that DPH contains:

> all actions immediately preceding or subsequent to the qualifying act. In a cyber operation, this period might begin once an individual begins probing the target system for vulnerabilities, extend throughout the duration of activities against the system, and include the period during which damage is assessed to determine whether re-attack is required.[56]

Concerning delayed effects, the majority found:

> that the duration of the individual's direct participation extends from the beginning of his involvement in mission planning to the point when he or she terminates an active role in the operation. [...] Note that the end of the period of direct participation may not necessarily correspond with the point at which the damage occurs.[57]

53 Compare: Marco Roscini, 'Evidentiary Issues in International Disputes Related to State Responsibility for Cyber Operations', [2015] *Texas International Law Journal*, 234-238; Wissenschaftliche Dienste des Bundestags, *Anwendbarkeit des humanitären Völkerrechts auf Computernetzwerkoperationen und digitale Kriegsführung (Cyber Warfare)*,(2015) WD 2 – 3000 – 038/15, 10-11; Different view: Russell Buchan, Cyber Warfare and the Status of Anonymous under International Humanitarian Law, [2016] *Chinese JIL*, 767.

54 *Supreme Court* (n 10), para. 40; *HPCR Manual* (n 7), Rule 28, 119, para. 5; ICRC (n 1), 70-71; Boothby (n 49), 753-758.

55 *Supreme Court* (n 10), para. 34; Jean De Preux, *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*, 516, para. 1679.

56 *Tallinn Manual* (n 6), 121.

57 Ibid.

If one sticks closely to the wording and just considers the time during which the civilian is active as the moment of losing protection, the period of DPH ceases after activation[58] and the civilian directly participating by using ACWS will again become legally privileged. The civilian may not be attacked after the action,[59] or at the time when nearly all of these actions will be noticed due to the practical peculiarities of the autonomous behaviour of the ACWS.

By a narrow understanding of the legal criteria, the time-frame in which the person is directly participating and the time when one can recognise the actions differ widely. Therefore, sticking to the original, restrictive understanding[60] would mean creating a legal grey area. A civilian must be afraid of becoming a legitimate target only during preparation and in the short period of activating the ACWS. Thus, a civilian user of ACWS could continue such actions and is encouraged to do so without the threat of retaliation. This could lead to an escalation of the use of ACWS by civilians in future conflicts. Hence, a different legal perception and understanding of the criterion 'for such time' is needed.

4. SOLUTION FOR THE REVOLVING DOOR 2.0

A. Proposal for Solution

For the solution of this dilemma, it is proposed that the legally relevant time-frame be extended from the act to the whole action, meaning the ongoing cyber operation. As long as the ACWS is working, the person responsible for activating the program has to be considered as directly taking part in hostilities.[61] As soon as the damage has occurred, meaning the end of this period, the time-frame to react has to be narrowed down until the moment when an appropriate and proportional reaction is no longer possible depending on the single case. This means that, until the program has reached its final destination and stopped its damaging purpose, plus an appropriate and proportional time to react after recognition, the civilian loses their protection based on Article 51(3) AP I.

In theory, ACWS could operate endlessly and so in theory the protection could never be regained. However, this would be in contrast to the exceptional character of the rule and the wording indicating a temporary period ('for such time'), which implies the end of the loss of protection, too. Therefore, the extension of the time period 'for such time' has to be understood as applying only for the time that the ACWS is actively operating and cumulatively coinciding with the set constitutive criteria for DPH.

B. Possible Critique and Replies to the Critique

Nevertheless this view can be criticized for several reasons. The ICRC states that:

> any extension [...] beyond specific acts would blur the distinction made in IHL between temporary, activity-based loss of protection (due to DPH), and continuous, status or function-based loss of protection [due to combatant status or continuous combat function].[62]

58 ICRC (n 1), 65, 67-68.
59 Buchan (n 53), 766-767.
60 Critic: Boothby (n 49), 743.
61 Compare: Dinstein (n 4), 148; Ibid., 758; *Tallinn Manual* (n 6), 121.
62 ICRC (n 1), 44-45.

Concerning remote attacks, the ICRC guidance states that:

> the duration of direct participation in hostilities will be restricted to the immediate execution of the act and preparatory measures [...].[63]

This apparently rejecting statement, however, does not have to mean that the use of an ACWS cannot extend the period of DPH. If the execution phase also encompasses 'the period over which the [malware] is installed or deployed' or 'the period over which [...] the targeted systems are compromised' and 'to the time period over which the victim actually suffers the effects',[64] the use of ACWS leads to a DPH by a civilian.

The arguments on a narrow or wide interpretation of the DPH also apply here: a narrow understanding of 'for such time' increases the protection of civilians using ACWS. The criticism that there is no military necessity to attack a civilian 'who is no longer playing a role in the operation'[65] aims in the same direction. The script or code of the ACWS may still be running while the civilian is attacked.[66]

On the other hand, a narrow interpretation would lead to a privilege for a civilian using such ACWS. These civilians will be encouraged to continue participating in hostilities,[67] when they realise that their actions cannot or do not provoke sanctions. This contravenes the original purpose of the protection of civilians in exchange for refraining from hostilities.

A compromise could be to fall back on criminal law and to try to get hold of the person instead of attacking them.[68] However, this requires territorial control or the cooperation of the host state of the perpetrator, which is hardly likely, and does not counter the threat of a continuing ACWS.

A civilian who wants to desist from attacks by ACWS may have to inform concerned conflict parties about the ACWS's existence in order to regain protection,[69] as the danger posed by such tools can be unlimited in time.

Finally, the military necessity can be found in the balance of humanity and military necessity itself.[70] This 'subtle equilibrium' has to be preserved[71] under all circumstances. Additionally, the prevention of future attacks could be considered necessary, too.

Often the situation is compared to a civilian placing a mine or an IED, who is regarded as not directly participating after its return, completing the action (the revolving door problem),[72] but these two situations are only superficially comparable. A minelayer could in theory be under surveillance during preparation, and a mine is physical and can be found by technical

63 Id., 68.
64 Owens, Dam, and Lin (n 41), 90.
65 Roscini (n 2), 209.
66 Dinniss (n 2), 274-276.
67 *Supreme Court* (10), para. 34.
68 Id., para. 40.
69 Boothby (n 49), 757.
70 Compare: Ibid., 767; Buchan (n 53), 768.
71 Dinstein (n 4), 5.
72 Dinniss (n 2), 275-276; Buchan (n 53), 767.

means. With ACWS, the fog of war is much denser; ACWS are non-physical and cyberspace works at high speed. The interconnectivity offers endless possible connections and there are many possibilities to hide the origin of an attack in the data transfer chain. If the perpetrator's computer is not linked to the Internet, state hacked or continuously observed, a civilian DPH using ACWS cannot be watched in nearly any case.

C. Arguing in Favour of an Extension of the Time Period

The period during which the civilian loses protection is not unlimited due to the necessary prevalence of the cumulative requirements. In particular, the loss of the belligerent nexus due to a provable withdrawal could prohibit a loss of protection concerning the hit of an arbitrary target.

Another argument in favour of the extension can be based on the wording of the rule. The phrase 'for such time' does not prevent an expansive interpretation if one views the running ACWS as the continuing act of participation. The use of ACWS can be considered as an extended arm of the human acting behind it. The system as a tool fulfils the set duties or frameworks and acts as the human would. Thereby the ACWS substitutes the human element of directness and immediacy.[73]

The same applies to the need for the requirement of only one causal step.[74] As soon as the malware is activated, it works autonomously and no other causal step by a human is needed. Even the ICRC guidance points out that causal proximity and temporal proximity do not have to coincide.[75]

The act of activation of the ACWS is a physical act[76] and an integral part to cause the result, which should not be underestimated. As soon as an ACWS is set free it will act independently and can rarely be stopped, comparable to an artillery shell or an unsophisticated rocket. The activation is the last step that the human has available to refrain from an attack using an ACWS. One could argue that the belligerent nexus is not given, as an arbitrary target is chosen by the ACWS, which was not originally intended by the civilian. Therefore the establishment of the necessary link could also be unintended. But even if the ACWS does not attack a specifically designated target, as long as there is a link to the armed conflict (and some argue that as long as it is either in support of a party or to the detriment of another, not both), the required nexus exists for the civilian. By using an ACWS a danger is created, which damages unforeseeable and unintended targets exactly because of its autonomy.[77]

From an objective perspective, this solution is favourable. The victim recognises that something is harming its systems, but whether this is directly controlled by a human or not is unknown to the victim. Even if the system does not have a set framework which substitutes for a kind of ongoing human control, the perpetrator creates a danger of an unforeseeable ACWS acting with less limits in cyber space. It could theoretically target endlessly and indiscriminately, potentially resulting in civilian casualties.

73 Compare *Supreme Court* (n 10), para. 35.
74 Roscini (n 2), 206.
75 Schmitt (n 14), 728; ICRC (n 1), 55.
76 Compare Ibid., 56, 57; Schmitt (n 14), 732.
77 Schmitt (n 14), 735-736.

The bad faith of the civilian to use such cyber tools also militates against a narrow interpretation of the time-frame.[78] Besides the necessary tools and abilities, an in-depth knowledge of the functioning of the ACWS, and possibly of the targeted systems and their security barriers, is also required – possible attack points, the logic of the whole network, and so on. The level of knowledge increases with the sophistication of a cyber operation. If an attack causes not only digital damage but also physical damage as a consequence of a manipulation of system controls, detailed knowledge of these processes is needed to prepare such a tool. A civilian using ACWS has to be considered able to foresee and understand the damage they might cause, and thus the consequences of their actions. For this reason, there is no legal need for a regaining of protection by the law, except where there are contravening acts of withdrawal by the civilian, like providing information about the operating ACWS.

'[G]rey areas should be interpreted in favour of finding direct participation' to enable 'a clear distinction between civilians and combatants'.[79] It would also be inequitable to restrict the range of time concerning a civilian using ACWS. The rule of equity is known especially in the Anglo-American legal system and requires that 'those who seek equity shall come with clean hands'.[80] The civilian makes use of the protection of humanitarian law but nevertheless acts to its detriment by getting involved in hostilities. Therefore, they cannot seek equitable treatment, meaning they cannot be protected by international law, while at the same time violating it by the same acts.

An extension of the time-frame would also provide for legal equality with the perpetrator.[81] Otherwise the civilian would be privileged by law due to using ACWS or delayed or conditioned attack tools. The use of delayed attacks would legally be protected, even if this conduct violates the principle of distinction in its negative understanding (whom not to consider as a protected civilian).

Finally, the requirement that the constitutive criteria have to prevail for implying a direct participation of a civilian restrict an inadmissible or unlawful extension of DPH.

5. CONCLUSION

The proposed solution for an extension of the DPH rule applying to a civilian using ACWS is practical. It restores the balance between the protection of civilians and military necessity by preventing legal privileging of wilful perpetrators. Nevertheless, a lawful attack on a participating civilian has a high prerequisite: an attacking party has to be able to attribute and identify the individual[82] with a high level of certainty and proof[83] as the life of the perpetrator is at risk due to the consequences. This and the determination of the DPH criteria have to

78 Compare Boothby (n 49), 759-760.

79 Michael Schmitt 'Direct Participation in Hostilities and 21st Century Armed Conflict' in H. Fischer (ed.) *Crisis Management and Humanitarian Protection: Festschrift für Dieter Fleck* (2004), 505, 509.

80 *Diversion of Water from the Meuse, Netherlands v Belgium*, Judgment, PCIJ Series A/B No 70, ICGJ 321 (PCIJ 1937) [Individual Opinion of M Hudson] 77; *Military and Paramilitary Activities in and against Nicaragua, Nicaragua v. US of America*, (Dissenting Opinion of Judge Schwebel) [1986] ICJ Rep 259 [268]; compare also the Latin legal quote 'nemo auditur propriam turpitudinem allegans'.

81 Compare: Boothby (n 49), 757.

82 Wolff Heintschel von Heinegg, 'Cyberspace- Ein völkerrechtliches Niemandsland' in: Schmidt-Radefeldt/ Meissler (Hrsg.), *Automatisierung und Digitalisierung des Krieges*, (Nomos 2012) 159, 172.

83 *Supreme Court* (n 10), para. 40; Roscini, *Evidentiary Issues* (n 53), 254; Schmitt (n 14), 736.

be conducted with 'all feasible precautions', according to Article 57(2)(a)(i) AP I. If there is doubt, the person has to be treated and protected like a civilian.[84] In reality, a possible practical consequence could be a targeted cyber counterstrike on a civilian computer or computer network instead of a targeted lethal strike on an identified civilian. All in all, the problem of cyber operations by civilians will increase in the future and more sophisticated and effective cyber means will be available. The suggested solution should be recognised as one answer to react to these significant threats, especially those posed by autonomous abilities of cyber tools.

REFERENCES

Dorota Banaszewska, 'Kombattanten und Zivilisten weit weg vom Schlachtfeld' in Robert Frau (ed.), *Drohnen und das Recht* (Mohr Siebeck 2014).

Bill Boothby, ' "And For Such Time As": The Time Dimension to Direct Participation in Hostilities', [2010] *International Law and Politics* 42, 741-768.

Russell Buchan, 'Cyber Warfare and the Status of Anonymous under International Humanitarian Law', [2016] *Chinese JIL*, 741-772.

Heather Harrison-Dinniss, 'Participants in Conflict – Cyber Warriors, Patriotic Hackers and the Laws of War' in Dan Saxon (ed.), *International Humanitarian Law and the Changing Technology of War* (Martinus Nijhoff Publishers 2013) 251-278.

Jean De Preux, 'Protocol I – Article 43' in: Yves Sandoz, Christophe Swinarski, Bruno Zimmermann (eds.), *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (Martinus Nijhoff Publishers 1987).

Yoram Dinstein, *The Conduct of Hostilities under the Law of International Armed Conflict* (2nd edn., Cambridge University Press 2010).

Wolff Heintschel von Heinegg, 'Cyberspace – Ein völkerrechtliches Niemandsland' in Schmidt-Radefeldt/ Meissler (eds.), *Automatisierung und Digitalisierung des Krieges* (Nomos 2012) 159-174.

Jean-Marie Henckaerts, Louise Doswald-Beck, *Customary International Humanitarian Law*, Vol I, (Cambridge University Press 2005).

ICRC, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law*, Nils Melzer (ed.) (ICRC 2009).

William A. Owens, Kenneth W. Dam, and Herbert S. Lin, *Technology, Policy, Law and Ethics* (1st edn., National Academies Press 2009).

Program on Humanitarian Policy and Conflict Research at Harvard University, *Commentary on the HPCR Manual on the International Law Applicable to Air and Missile Warfare* (2010), available at www.ihlresearch.org/amw.

Marco Roscini, *Cyber Operations and the Use of Force in International Law* (1st edn. OUP 2014).

Marco Roscini, 'Evidentiary Issues in International Disputes Related to state Responsibility for Cyber Operations', (2015) *Texas International Law Journal*, Vol 50, Symposium Issue 2, 233-273.

Michael N. Schmitt (ed.), *Tallinn Manual on the International Law Applicable to Cyberwarfare* (Cambridge University Press 2013).

84 Schmitt (n 14), 736.

Michael N. Schmitt, 'Deconstructing Direct Participation in Hostilities: The Constitutive Elements', (2010) *International Law and Politics*, Vol. 42, 697-739.

Michael N. Schmitt, 'Direct Participation in Hostilities and 21st Century Armed Conflict' in H. Fischer (ed.) *Crisis Management and Humanitarian Protection: Festschrift für Dieter Fleck* (Berliner Wissenschafts-Verlag 2004), 505-529.

Wissenschaftliche Dienste des Bundestags, *Anwendbarkeit des humanitären Völkerrechts auf Computernetzwerkoperationen und digitale Kriegsführung* (Cyber Warfare), (2015) WD 2 – 3000 – 038/15.