GENEVA ACADEMY
Académie de droit international humanitaire et de droits humains
Academy of International Humanitarian Law and Human Rights

WORKING PAPERS

Protecting the global information space in times of armed conflict

ROBIN GEISS AND HENNING LAHMANN*

FEBRUARY 2021

*The authors wish to thank Dr Kubo Mačák for his helpful comments on an earlier draft of the paper.

THE GENEVA ACADEMY A JOINT CENTER OF
THE GRADUATE INSTITUTE GENEVA
INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES
UNIVERSITÉ DE GENÈVE
FACULTY OF LAW

TABLE OF CONTENTS

INTRODUCTION

The growing number of allegations of adversarial foreign influence operations over the past couple of years, carried out by a variety of international actors directed against democratic decision-making processes in other states have put the problem of information warfare high on the international agenda.[1] The interference in the 2016 U.S. and the 2017 French presidential elections as well as the 2016 Brexit referendum in the UK are only the most prominent examples. The phenomenon is certainly neither abating nor geographically limited: In late 2020, for instance, Somalia expelled Kenya's diplomatic staff after accusations of electoral meddling.[2] Since the beginning of 2020, an unprecedented surge of misinformation and disinformation surrounding the COVID-19 pandemic has added a new sense of urgency while at the same time expanding the scope of the legal questions. However, so far the ensuing debate among scholars and policy-makers has been focused on international human rights law and other questions of peacetime international law, such as whether and under which circumstances an (online) disinformation campaign targeting audiences abroad may amount to a violation of the target state's sovereignty, the principle of non-intervention, or even – in extreme cases – the prohibition of the use of force.[3] The legal implications of digital information warfare in the context of armed conflict, on the other hand, have so far received scarce attention. This brief working paper aims at filling this gap by exposing some of the legal issues arising in relation to mis- and disinformation tactics during armed conflict in order to serve as a starting point for further debate in this respect:

What, if any, limits exist concerning (digital) information operations in armed conflict? Does the humanitarian legal framework adequately capture the humanitarian protection needs that arise from these types of (military) conduct? Where and how to draw the line between effects and side-effects of digitalised information warfare that should remain either within or without the protective ambit of international humanitarian law (IHL)? What are, or what should be, the limits of disinformation campaigns, 'fake news', deep fakes and the systematic manipulation of a given information space in times of armed conflict? Does IHL, which is traditionally and primarily focused on preventing physical harms, sufficiently account for and is capable of mitigating potentially far-reaching consequences that such types of operations can have on societies? If not, should it?

While the laws of armed conflict have proven to be flexible enough to anticipate technological innovation in general and are applicable also to new means and methods of warfare, as thoroughly discussed in relation to the application of IHL to cyber warfare,[4] it is less obvious whether the protection they provide remains adequate in all instances in which novel forms of warfare are employed. And while it is certainly true that disinformation campaigns, ruses and other methods of deception and propaganda have always been part and parcel of warfare, recent technological developments, especially in the fields of cyber and artificial intelligence, are to be seen as a veritable gamechanger of (dis-)information

[1] This paper focuses on manipulation of the content of information. It does not or only peripherally deal with other issues in relation to the contemporary information ecosystem, such as hate speech or incitement to violence.

[2] See Latif Dahir, Abdi, "Somalia Severs Diplomatic Ties with Kenya", New York Times, 15 December 2020, https://www.nytimes.com/2020/12/15/world/africa/somalia-kenya.html.

[3] See only Milanovic, Marko, and Michael N. Schmitt, 'Cyber Attacks and Cyber (Mis)Information Operations During a Pandemic', Journal of National Security Law & Policy (2020).

[4] See only Laurent Gisel, Tilman Rodenhäuser and Knut Dörmann, 'Twenty Years On: International Humanitarian Law and the Protection of Civilians Against the Effects of Cyber Operations During Armed Conflict', *International Review of the Red Cross* (2020), 11-16; Michael N. Schmitt (ed.), Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (2017), 373 et seq.

warfare. Considering the scale, scope, and far-reaching effects of peacetime disinformation operations, and taking into account the constantly increasing level of military cyber capabilities, the traditional assumption that generally speaking all types of disinformation operations short of perfidy are permissible during armed conflict should be revisited. Thus, while the simulation of surrender with the intent to injure an enemy soldier undoubtedly amounts to prohibited perfidy, under IHL – leaving IHRL aside for a moment – it is far less clear that the widespread and deep manipulation of a target country's entire online information ecosystem – including news and social media, but even scholarship, expert opinions, or studies by policy analysts and pundits – is prohibited or limited by IHL. Propaganda as well as psychological and influence operations, including even operations directed at the civilian population[5], have been a common and widely accepted feature of warfare throughout the ages. What is more, Article 37(2) API entails an explicit – and for an IHL rule unusually – permissive provision confirming the permissibility of ruses of war, whereas only a specific and narrowly defined set of acts of deception, i.e., those amounting to perfidy, are explicitly prohibited as such. Last but not least, the Tallinn Manual lists 'psychological warfare activities' as an example of permissible ruses.[6] All of this taken together opens up a wide spectrum of permissible disinformation campaigns in times of armed conflict, and in combination with a long-standing practice of such operations, from the outset renders any attempt at discussing legal limits and prohibitive thresholds for such operations inherently difficult. This said, it is precisely for this reason and indeed the point of this paper to start a debate and to question whether the long-standing practice of psychological and influence operations, considering how powerful and damaging some of these operations have become in the wake of global digitalization, is still to be seen as a 'common feature of war' with basically only the prohibition of perfidy as a constraint.

After presenting a few brief scenarios of possible (military) information operations in situations of armed conflict to illustrate what is potentially at stake, the subsequent section defines some of the key concepts concerning the issue at hand. The main part examines whether and to what degree existing rules of IHL put limitations on the conduct of information warfare. A short look at international criminal law and international human rights law follows before the paper concludes with an outlook on potential paths to advance the debate.

MAPPING THE THREAT LANDSCAPE: RISKS TO THE INFORMATION SPACE IN CONTEMPORARY ARMED CONFLICT

Information operations in the context of armed conflicts can occur in vastly different contexts and can have a variety of different effects on the targeted societies and civilian populations, depending on the mode of conduct, namely the technologies employed, the scope, scale and sophistication of the operation or campaign, the target audience, and the aims pursued. In order to illustrate the matter, a set of hypothetical scenarios – loosely based on past events – follows below.

SCENARIO A – SOCIAL MEDIA-ENABLED FOREIGN ELECTORAL INTERFERENCE

The governmental armed forces of State A are involved in a protracted, low-intensity non-

[5] Mike Schmitt EJIL Blog French Position – explicit - https://www.ejiltalk.org/france-speaks-out-on-ihl-and-cyber-operations-part-ii/

[6] Commentary to rule 123, p. 495.

international armed conflict with Insurgent Group G, which controls parts of the territory of State A. In the months prior to a general election in State A, the military cyber unit of neighbouring State B – which has been supporting Insurgent Group G with weapons, logistics, and covert special forces operations over the course of the conflict – sets up a concerted disinformation campaign on social media in close coordination with domestic groups belonging to G. Employing tools such as fake accounts, bots, and micro-targeting algorithms, the operation disseminates misleading and false political content to State A's electorate in order to discredit the incumbent and boost support for her contender, who publicly supports the main demands of Insurgent Group G, including secession, and a close future alliance with State B. Despite having trailed in the polls for months, the contender surprisingly wins the election and assumes the presidency.

SCENARIO B – LARGE-SCALE DISTORTION OF THE MEDIA ECOSYSTEM

During a situation of sustained political tension between State A and State B, the military information operations unit of State B starts an open propaganda campaign, disseminated via social media, video streaming platforms, and state-owned TV channels, that attempts to undermine public support in State A for the policies of its government vis-à-vis State B by highlighting arguments that contradict the official justification of the government's positions. As the campaign does not seem to yield discernible results, the military of State B launches a limited number of missiles against the territory of State A while the military information operations unit spreads a video via social media – using fake accounts that appear to belong to ordinary citizens of State A – that ostensibly shows a high-ranking political leader admitting that the armed conflict was actually initiated by State A under false pretences. Shortly thereafter, the military of State B starts a large-scale cognitive warfare operation aiming at the distortion of the entire online media ecosystem of State A. The content on the websites of all of the most important public broadcasting services and the leading newspaper publishers is subtly, and at first virtually imperceptibly, falsified and manipulated, in line with the official position of State B. At various points, the leading news websites furthermore suffer from seemingly random DDoS attacks that render them inaccessible for considerable amounts of time. The military information operations unit even carefully rewrites the main points of already published expert opinions and academic studies dealing with political issues that are points of contention between the two countries. The combined 'epistemic assault' leads to a lasting corrosion of the media ecosystem of State A and results in widespread and sustained confusion among the civilian population. As the official language of State A is the *lingua franca* of much of the globalised markets, science and scholarship, and international diplomacy, the manipulation of the state's news media even has ripple effects across the globe. Although the original content can gradually be reinstalled and it eventually turns out that the video had been fabricated using 'deep fake' algorithms, support for the government and the war effort in State A drop significantly. Eventually, the military of State A is forced to retreat. The upheaval in the country proves to be lasting due to the loss of public trust in both the media and political structures, resulting in a sustained period of political instability that is further exploited by State B to achieve its own goals at the expense of State A.

SCENARIO C – MANIPULATION OF CIVILIAN BEHAVIOUR TO GAIN MILITARY ADVANTAGE

While a severe respiratory disease pandemic is spreading across the globe, State A and State B are engaged in an armed conflict that mainly revolves around disputed territory that is a province of State A but claimed by State B. The

information operations unit of the armed forces of State B gains access to private groups on a social media platform that are used and frequented mainly by members of the armed forces of State A. Pretending to be soldiers of State A, the unit disseminates the false information that ingesting methanol helps to prevent contracting the virus. Although the information is only shared within the closed groups, screenshots quickly spread all across the social network, which leads to the death of both members of the armed forces and civilians who drink pure methanol after having been exposed to the false information.

Further on, the information operations unit of State B disseminates via various social media platforms the false information that the contested territory has seen several large and severe outbreak clusters of the disease and that for that reason, the authorities of State A have imposed new health guidelines for the province, including a total lockdown for 14 days. The information leads to confusion and fear among the resident civilian population. While the government of State A tries to correct the disinformation and re-establish order, the armed forces of State B exploit the confusion and the lockdown to make extensive territorial gains.

SCENARIO D – COMPROMISING AND EXTORTING OF CIVILIAN INDIVIDUALS THROUGH INFORMATION WARFARE

During an armed conflict between State A and State B, the cyber operations unit of State A hacks into servers that store sensitive personal information about D, who is the CEO of a large defence contractor in State B. The unit subsequently starts to disseminate the information via social media platforms and to journalists working at major news outlets in State B; while most of the information is factually correct, the unit also subtly falsifies a number of documents and photographs to further compromise D. Finally, the cyber operations unit conveys the message to D that it will release the most intimate, embarrassing, and humiliating information unless D agrees to delay the further development of an advanced fighter jet by his company.

SCENARIO E – DISINFORMATION AS INCITEMENT TO VIOLENCE

State A has been ravaged by a protracted civil war that has mostly been fought along ethnic lines. The military, which is primarily composed of members belonging to the majority ethnic group, starts using a social media platform, which serves as the dominant means of communication and information in State A, to disseminate dehumanising disinformation about one of the minority ethnic groups which the government considers not to be part of the 'legitimate people of State A'. At least partly as a result of the sustained disinformation campaign, openly hostile attitudes towards the minority group among the majority population increase considerably. After the military suffers from some setbacks in its combat operations against various rebel groups, it begins to spread false rumours about certain members of the minority group having raped a woman belonging to the majority ethnicity. This false information, which spreads quickly and widely via the platform, leads to severe violence against the minority by civilian members of the majority population.

VARIANTS OF DISTORTING THE INFORMATION ECOSYSTEM IN WAR AND PEACE

As the brief scenarios show, the manipulation of specific pieces of information and the distortion of the digital information ecosystem in an entire country, a region, or even globally can take a variety of modes and manifestations. All of the above examples are, to a greater or lesser extent, based on real-world cases, although most of them did not occur in

the context of an ongoing international or non-international armed conflict. However, how such scenarios could play out as part of a military campaign is easily imaginable and it is only a question of time before such operations will occur during armed conflicts. Before commencing with an analysis of the legal implications of such operations within the framework of existing IHL, a couple of conceptual clarifications are in order.

INFORMATION

For the purpose of this working paper, 'information' can be defined as a set, allocation, or combination of data structured in such a way that it carries and conveys *meaning*. It can roughly be translated as the 'content' that is transmitted through media such as a newspaper, a TV or radio broadcast, a website, a social media platform such as Facebook, Twitter, VKontakte, WeChat, or Sina Weibo, but also via point-to-point communication such as an email or a text message. The distinction between 'information' and 'data' is crucial; information *can* be manipulated by an operation against the confidentiality, integrity, or availability of data, but that does not need to be the case. For instance, a news story disseminated via social media that conveys false or misleading information merely *creates* new data without necessarily altering any existing data. The distinction is not always made sufficiently clear in the literature, especially when (digital) information operations are treated as a mere sub-category of cyber operations, which can be misleading and risks to neglect the nuances that make information operations different, including, but not limited to, questions of causation. Although cyber and information operations will often be employed in combination, the mechanisms of impacting their targets are analytically distinct. The latter always depend on the actions of a *susceptible audience* to ultimately be successful.

INFORMATION SPACE

'Information space' is aptly described simply as 'a place [...] where information is available',[7] which can be a website, a YouTube channel, a podcast, an e-book, a journal, but also a classic library or a market square; the interconnection of an infinite number of such singular information spaces through the worldwide networks ('cyberspace') is what makes the combined ecosystem 'global'. At the same time, this does not preclude the existence of 'national' information spaces as distinct parts, for instance through the use of different languages or the dissemination of specific content. To the extent that this paper concerns 'the protection of the global information space in armed conflict', what is principally at stake is the information itself and the perception of such information by individuals and collectives, not so much the 'spaces' where it is processed and presented; the protection of the latter is realised by means of data and IT security, i.e. 'cybersecurity' in the proper sense.

MISINFORMATION, DISINFORMATION, AND MAL-INFORMATION

Although scholarly attention has grown exponentially since the sobering revelation of the extent of meddling in the 2016 U.S. presidential election, the discourse has at times suffered from a lack of clarity and definitional rigor regarding frequently employed notions such as 'fake news', 'disinformation', 'misinformation', 'propaganda', 'cognitive warfare', 'influence operations', and 'information operations'.

While the term 'fake news' is generally seen as misleading and should be avoided given its overuse in public discourse despite its inherent lack of clarity,[8] 'disinformation' is more

[7] See https://dictionary.cambridge.org/dictionary/english/information-space.

[8] Claire Wardle and Hossein Derakhshan, 'Information Disorder: Toward an Interdisciplinary Framework for

expedient even though the concept, too, suffers from an abundance of occasionally incoherent descriptions. It is useful to contrast 'disinformation' with 'misinformation': whereas the latter signifies information that is factually wrong yet not intentionally so, disinformation is 'deliberately false or misleading'.[9] The European Commission defines the concept as 'false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit'.[10] This intended harm does not necessarily manifest in the inaccuracy of the piece of information itself (as would be the case with hate speech or incitement to violence) but in its context, application, and purpose. In this sense, even otherwise factually correct information can be employed in a misleading way and thus *as disinformation*, for example in cases where the recipient of the information is deceived as to the identity of the speaker. Some actors have proven to be especially apt at posting content on social media in the guise of a citizen of the target audience's country,[11] as could be witnessed ahead of the 2016 election and again in 2020. To these two categories one may add 'mal-information', a concept that describes the 'spreading of true information, but with the intent to cause harm'.[12]

PROPAGANDA

Conceptually distinct from the notion of 'disinformation' is the term 'propaganda', which is in some ways older and originally had a neutral connotation. In its more recent discursive application, it is most appropriately described as a deliberate attempt to persuade a target audience, often in the form of a coordinated information campaign. Frequently, although not necessarily, the persuasion is achieved by means of manipulation or deception.[13] Utilising disinformation as described above may be a part of such efforts, but it is not by definition an inherent element of propaganda. In principle, the objectives can just as well be achieved by disseminating factually correct information that is merely framed in a way that has a manipulative effect on the target audience. Such a communicative act often takes the form of putting an *alternative* narrative concerning a current or historical event in competition with the official or established one. In this way, manipulative information does not require an actual falsehood or a deception of the speaker's identity. Depending on the method of persuasion, 'propaganda' is sometimes further classified as either 'white' (accurate information with a leading narrative framing), 'grey' (a combination of accurate and false information), and 'black' (inaccurate information and/or deception of speaker identity).[14]

INFORMATION OPERATIONS/INFLUENCE OPERATIONS

Closely related to the term 'propaganda' is the notion of 'information operation'. In 2017, the social media company Facebook described 'information operations' as 'actions taken by organised actors (governments or non-state actors) to distort domestic or foreign political sentiment, most frequently to achieve a strategic and/or geopolitical outcome. These operations can use a combination of methods, such as false news, disinformation, or networks of fake accounts aimed at manipulating public

Research and Policy Making', Council of Europe Report DGI(2017)09, 27 September 2017, at 15.

[9] Caroline Jack, 'Lexicon of Lies: Terms for Problematic Information', Data & Society Research Institute, 2017, at 2-3, https://datasociety.net/pubs/oh/DataAndSocietyLexiconofLies.pdf.

[10] European Commission, 'A Multi-dimensional Approach to Disinformation', 30 April 2018, at 10.

[11] See e.g. Scott Shane, 'The Fake Americans Russia Created to Influence the Election', The New York Times, 7 September 2017.

[12] See Cyberlaw Toolkit, Glossary, https://cyberlaw.ccdcoe.org/wiki/Glossary.

[13] Caroline Jack, at 6-7.

[14] Caroline Jack, at 7.

opinion'.[15] Although not strictly congruent, there is thus a considerable conceptual overlap between the notions of 'propaganda' and 'information operations'.

'Influence operations' is employed most frequently in the more limited context of military operations, although the term does not seem to signify conduct that is as such much different from 'information operations' as defined above. The concept has been described as 'a method by which a military actor aims to affect the cognitive aspects rather than the physical aspects of individuals. This is primarily done using information and communication, rather than physical force, to compel groups or individuals to behave or think in ways that are conducive to the aims of the actor'.[16]

INFORMATION WARFARE/COGNITIVE WARFARE

Finally, once information is used strategically and with adversarial aims by or on behalf of a state which is in a state of conflict with another state (whether or not armed), further concepts such as 'information' or 'cognitive warfare' are in use in the literature. The Russian Ministry of Defence defines 'information war' as 'a struggle between two or more states ... to destabilise a society and a state through massive psychological conditioning of the population, and also to pressure a state to make decisions that are in the interest of the opponent'.[17] Such conduct falls into the broader, emergent strategic category of 'hybrid warfare'.[18]

PROTECTING INFORMATION SPACES UNDER EXISTING LEGAL FRAMEWORKS

INTERNATIONAL HUMANITARIAN LAW

In the following, it will be examined whether and to what extent existing IHL offers protections against adversarial information operations and other forms of cognitive warfare that target the civilian population in situations of armed conflict. For the purpose of legal analysis, a distinction between the specific elements of such operations has been suggested, as different rules and legal consequences might attach. These identifiable elements are, at least: (1) the content of the communicative act; (2) the mode of disseminating the information; (3) the target audience; and (4) the (actual or foreseeable) consequences of the communicative act.[19]

The pertinent legal frameworks of the laws of armed conflict address communication and information activities only tenuously and non-systematically. This is primarily a consequence of IHL's traditional focus on the physical effects of armed conflicts. Thus, for instance, while Article 79 AP I clearly states that journalists 'shall be considered as civilians' and 'be protected as such under the Conventions and this Protocol', it has been pointed out that the scope of this specific protection only covers the individual journalists as natural persons, but not (at least not directly) 'their journalistic activities or products, such as content posted on

[15] Jan Weedon, William Nuland and Alex Stamos, 'Information Operations and Facebook', 27 April 2017, at 4, https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf.

[16] Winther, Pontus, 'Military Influence Operations & IHL: Implications of New Technologies,' Humanitarian Law & Policy, 27 October 2017, https://blogs.icrc.org/law-and-policy/2017/10/27/military-influence-operations-ihl-implications-new-technologies/.

[17] See Martin Russell, 'Russia's Information War: Propaganda or Counter-Propaganda', European Parliamentary Research Service, 3 October 2016, at 2.

[18] See Patrick J Cullen and Erik Reichborn-Kjennerud, 'Understanding Hybrid Warfare', January 2017, at 8.

[19] See Winther, Pontus, 'Military Influence Operations & IHL: Implications of New Technologies,' Humanitarian Law & Policy, 27 October 2017, https://blogs.icrc.org/law-and-policy/2017/10/27/military-influence-operations-ihl-implications-new-technologies/.

a website'.[20] When it comes to questions regarding the content of information more broadly, the Tallinn Manual submits that the general rule is that 'psychological operations such as dropping leaflets or making propaganda broadcasts are not prohibited even if civilians are the intended audience'.[21] In line with this, it has been suggested that 'through the longstanding, general, and unopposed practice of States, a permissive norm of customary law has emerged, which specifically permits' such operations 'as long as [they] do not violate any other applicable rule of IHL'.[22] For example, the German law of armed conflict manual states that '[i]t is permissible to exert political and military influence by spreading – *even false* – information to undermine the adversary's will to resist and to influence their military discipline (e.g. calling on them to defect, to surrender or to mutiny)'.[23]

At the same time, there are a number of specific rules in existing IHL that impose limits on certain forms of information operations. As will be shown below, principal among these rules are the prohibition of perfidy, the prohibition to terrorize the civilian population as well as the prohibition to encourage violations of IHL and the obligation to treat civilians and persons hors de combat humanely. What is more, information operations that qualify as military operations and especially information operations that amount to an attack in the sense of IHL, are subject to additional legal constraints.

The problem in all of this, however, is that many of these rules entail limiting criteria or thresholds that sit oddly with 21st century digital disinformation campaigns. The relevant rules are anchored, understood, and interpreted in light of 20th century warfare practices. Typically, these rules are linked, in one way or another, to violent activity. Their rationale is to protect the integrity of IHL (perfidy), to limit violence and its most drastic psychological effects (prohibition of encouragement of IHL violations, prohibition of terrorizing civilians), or are focused on the protection of individuals (human dignity, humane treatment). These protection rationales undoubtedly continue to be relevant and these rules impose important limits for certain types of information campaigns in times of armed conflict. However, they are not aimed at protecting national or even the global 'civilian' information space as such. This is particularly relevant when discussing military information operations, the aim of which is to degrade information spaces during armed conflict and to cause instability, confusion, and loss of trust in a country's public institutions, media and democratic decision-making processes (Scenario B above). Of course, in keeping with IHL's overarching rationale to mitigate the worst – but not all – humanitarian impacts of war, it may well be argued that such effects should remain outside the protective realm of IHL even under the conditions of 21st century warfare. And clearly, noting that the first victim of war is the truth, overly restrictive limits on information operations during armed conflict would be utterly unrealistic. At the same time, the nature, scope, and impact of manipulative information operations occurring in peacetime and their long-lasting divisive and corrosive effects on public trust and societal stability require that more attention be given to these types of operations during armed conflict. Does IHL impose any limits on information operations that wreak havoc on a country's public information environment and that, while not aiming to terrorize, incite violence or to expose targeted individuals, aim to systematically undermine public trust and to spread large-scale confusion among the civilian population, as in Scenario B above?

DIGITAL PERFIDY AND RUSES OF WAR

[20] Tallinn Manual 2.0, rule 139, para. 3.

[21] Tallinn Manual 2.0, rule 93, para. 5.

[22] See "Scenario 12: Cyber Operations against Computer Data." In International Cyber Law in Practice: Interactive Toolkit, 2020, https://cyberlaw.ccdcoe.org/wiki/Scenario_12:_Cyber_operations_against_computer_data.

[23] Bundesministerium der Verteidigung, Law of Armed Conflict: Manual, May 2013, para. 487 (emphasis added).

For one, whereas generally speaking an information operation would be lawful if it were to be qualified as a permissible ruse, it would violate IHL if amounting to a (prohibited) perfidious act. 'Perfidy', in accordance with Article 37(1) AP I, is an act that invites 'the confidence of an adversary to lead him to believe that he is entitled to, or is obliged to accord, protection under the rules of international law applicable in armed conflict, with the intent to betray that confidence'. As is quite obvious, the scope of this prohibition – especially when considered against the backdrop of modern disinformation practices as described in the scenarios above – is relatively narrow. It has been emphasised that 'the perfidious act must be the proximate cause' of the death, injury or capture of a person belonging to the adversary party.[24] This will only ever be relevant in relation to very specific information operations that directly aim at such (physical) consequences with a particular mode of deception. Ruses of war, on the other hand – understood as 'acts intended to mislead the enemy or to induce enemy forces to act recklessly'[25] have a broader scope of application that generally includes psychological warfare activities. Jensen and Crockett present the example of 'a deep-faked video including inaccurate intelligence information [which] might significantly impact the conduct of military operations'.[26] Such deception of the adversary by way of a communicative act may however not be in conflict with any other applicable rule of the laws of armed conflict.

Notably, however, the examples typically provided for permissible ruses of war refer to instances in which new information – in whichever form – is distributed, rather than existing and trustworthy sources of information (e.g. a country's online news environment) are being manipulated or falsified. Thus, when talking about a permissive norm of customary law[27] it might be necessary to draw further distinctions between different types of information operations. What is more, like in the German law of armed conflict manual cited above, which speaks of 'the adversary's will to resist' as well as 'military discipline', there is often a reference to an overarching military purpose of the information operation without it being clear whether such a limitation is considered to be somehow prescribed by IHL or whether it is rather to be seen as simply reflecting the typical context in which such operations are likely to occur. It is telling that the 1987 Commentary on Additional Protocol I, defines a ruse of war as consisting 'either of inducing an *adversary* to make a mistake [...], or of inducing *him* to commit an imprudent act' and therefore appears to understand ruses of war as practices that have at least a nexus to military operations.[28] The Commentary lists 'simulating the noise of an advancing column', 'creation of fictitious positions', 'circulating misleading messages' and 'simulated attacks' as examples of ruses of war.[29] On the basis of this definition and the examples of ruses provided above, it is not clear that corroding a civilian information space with the aim to spread confusion and uncertainty among the civilian population and without any direct link to combat activity – e.g. by manipulating content in all major online newspapers in a given country – should automatically qualify as a permissible ruse of war.

PERSONALITY RIGHTS

[24] Tallinn Manual 2.0, rule 122, para. 5.

[25] Tallinn Manual 2.0, rule 123, para. 2.

[26] Jensen, Eric Talbot, and Summer Crockett, 'Deepfakes' and the Law of Armed Conflict: Are They Legal?,' Articles of War, 19 August 2020, https://lieber.westpoint.edu/deepfakes/.

[27] See above n 22.

[28] Yves Sandoz, Christophe Swinarski and Bruno Zimmermann (eds), *Commentary on the Additional Protocols*, ICRC, Geneva, 1987, para. 1515 (emphasis added).

[29] Ibid, para. 1516.

The obligation of humane treatment might constitute one of the rules that prohibit certain types of information operations in situations of armed conflict. Pursuant to Article 27 GC IV, '[p]rotected persons are entitled, in all circumstances, to respect for their persons, their honour, their family rights, their religious convictions and practices, and their manners and customs. They shall at all times be humanely treated, and shall be protected especially against all acts of violence or threats thereof and against insults and public curiosity'. The ICRC has submitted that such public exposure is prohibited even when it 'is not accompanied by insulting remarks or actions' as it is 'humiliating in itself'.[30] Crucially, it has clarified that '[i]n modern conflicts, the prohibition also covers … the disclosure of photographic and video images, recordings of interrogations or private conversations or personal correspondence or any other private data, irrespective of which public communication channel is used, including the internet'.[31]

The 1958 commentary to the Fourth Geneva Convention calls the obligation of humane treatment the 'leitmotiv' of all four Conventions.[32] For this reason, '[t]he word 'treatment' must be understood in its most general sense as applying to all aspects of man's life'.[33] Rule 87 of the ICRC Customary Law Study stipulates a general obligation to treat civilians and persons hors de combat humanely under customary international law. What is more, in the context of non-international armed conflicts common Article 3 GC I-IV prohibits outrages upon personal dignity, in particular humiliating and degrading treatment. The ICRC's 2016 Commentary lists, *inter alia*, 'forced public nudity' and 'enduring the constant fear of being subjected to physical, mental or sexual violence, as relevant acts violating this prohibition (para. 672).[34] Therefore, information operations targeting a civilian and amounting to a violation of that person's personal dignity, such as the operation in Scenario D that aims at humiliating the CEO in order to blackmail him, would be in violation of the customary law obligation to treat civilians humanely.

INCITEMENT OF VIOLENCE

Pursuant to common Article 1 of the Geneva Conventions as well as Article 1(1) AP I, parties to an armed conflict are under an obligation to respect and ensure respect for the rules of IHL 'in all circumstances'. While some aspects regarding the interpretation of common Article 1 GC I-IV remain controversial, it is widely accepted that common Article 1 entails a prohibition to encourage violations of IHL. According to the ICRC Commentary, the rationale of this negative obligation is that '[i]t would be contradictory if common Article 1 obliged the High Contracting Parties to 'respect and ensure respect' by their own armed forces while allowing them to contribute to violations by other Parties to a conflict'.[35] This implies that a state would violate this rule in a situation of armed conflict if it disseminated information that induced combatants *or* civilians to attack and harm other civilians, for instance in inter-ethnic violence in the course of a civil war.[36] Despite the fact that some existing law of war manuals of armed forces, for example the German Bundeswehr's 'Handbuch humanitäres Völkerrecht in bewaffneten Konflikten', employ the terminology of 'instigating'

[30] Commentary of 2020 to Convention (III) relative to the Treatment of Prisoners of War, para. 1624.

[31] Id.

[32] Commentary of 1958 to Convention (IV) relative to the Protection of Civilian Persons in Time of War, p. 204.

[33] Id.

[34] International Committee of the Red Cross, Commentary of 2016, Article 3: Conflicts Not of an International Character, para. 672.

[35] Id., Article 1: Respect for the Convention, para. 158.

[36] See 'Scenario 19: Hate Speech', in Kubo Mačák, Tomáš Minárik and Taťána Jančárková (eds.), Cyber Law Toolkit (2019), revision as of October 1, 2020, para. L16, available at https://cyberlaw.ccdcoe.org/wiki/Scenario_19:_Hate_speech.

('Aufforderung'),[37] it can hardly make a difference whether the encouragement to violate IHL is made explicitly or implicitly. Thus, it is argued that the inducement can be carried out by way of disseminating inciting disinformation via social media as described in Scenario E, which is modelled after recent events in Myanmar.[38] There are therefore good reasons to conclude that such violence inciting types of disinformation in armed conflict would amount to a violation of existing IHL.

TERRORIZING

The prohibition against terrorising civilians might also provide protection against certain adversarial information operations in armed conflict.[39] According to Article 51(2) AP I, '[a]cts or threats of violence the primary purpose of which is to spread terror among the civilian population are prohibited'. This rule is furthermore accepted as part of customary IHL, applying to all kinds of armed conflicts.[40] However, two aspects of this rule considerably limit its scope vis-à-vis this type of military conduct. For one, the communicative act in question must either amount to an attack within the meaning of IHL or a threat thereof.[41] Whether an information operation may constitute an attack in and of itself at all will be discussed below; either way, it seems indisputable that typically most such conduct will not reach this threshold. Thus, even if disseminated disinformation spreads fear and terror among targeted civilians, the operation will not automatically come within the protective ambit of Article 51(2) AP I if it does not, at the same time, constitute or threaten an act of violence. A 'threat' is a purposely directed speech act 'that suggests to the addressee the future occurrence of a negative treatment or event'.[42] The mere *exploitation* of a state of fear and terror or the spreading of fear for general destabilization as in Scenario C, whether related to the aim of gaining a military advantage or not, will therefore typically not suffice to trigger the prohibition in the absence of an actual or threatened act of violence. Furthermore, it must be the *primary purpose* of the act or threat of violence to spread terror. This implies that in situations where other motives and objectives take precedence, the prohibition (as it currently stands) is not applicable even if the result of an information operation is extreme fear among the civilian population on the receiving end.[43] In light of the far-reaching and terrorizing effects digital information warfare campaigns can have in the 21st century, it should be reconsidered whether such operations, whenever it is their (primary) purpose to spread terror among the civilian population, should not be explicitly prohibited regardless of whether or not they can be qualified as an act of violence.

'MILITARY INFORMATION OPERATIONS': CONSTANT CARE TO SPARE THE CIVILIAN POPULATION

Furthermore, adversarial information operations in armed conflict might violate the

[37] Bundesministerium der Verteidigung, Law of Armed Conflict: Manual, May 2013, para. 487.

[38] See Mozur, Paul. 'A Genocide Incited on Facebook, with Posts from Myanmar's Military.' The New York Times, 15 October 2018 2018.

[39] See e.g. Ghia, Unnati, 'International Humanitarian Law in a Post-Truth World,' Cambridge International Law Journal Online, 17 December 2018, http://cilj.co.uk/2018/12/17/international-humanitarian-law-in-a-post-truth-world/; Jensen, Eric Talbot, and Summer Crockett, '"Deepfakes" and the Law of Armed Conflict: Are They Legal?,' Articles of War, 19 August 2020, https://lieber.westpoint.edu/deepfakes/; Winther, p. 147 et seq.; 'Scenario 19: Hate Speech' (n 36), para. L15.

[40] See ICRC, IHL Database: Customary IHL, Rule 2, https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule2.

[41] See Tallinn Manual 2.0, rule 98, para. 3.

[42] Winther, p. 148.

[43] Id., p. 152; but see International Committee of the Red Cross, Commentary of 1987 to Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflict (Protocol I), 8 June 1977, Article 51: Protection of the Civilian Population, para. 1940, which leaves open the possibility of a broader interpretation.

obligation of constant care as stipulated by Article 57(1) AP I: 'In the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects.'[44] The ILA's Study Group on the conduct of hostilities agreed that 'the obligation to take constant care to spare the civilian population applies to the entire range of military operations and not only to attacks in the sense of Art. 49 AP I'.[45] Against this backdrop, at least those communicative acts by armed forces that aim at furthering military objectives could be considered 'military operations' within the ambit of the provision, in line with the legal position put forward in certain military manuals such as the U.S. Department of Defense law of war manual, which deals with military operations and includes a section on 'propaganda'.[46] This broader reading of the notion of military operations does not align with traditional interpretations of the term that, in keeping with 20th century warfare practices, understood it to refer to physical military operations (such as manoeuvres or troop movements). But the term's natural meaning does not preclude the possibility to interpret it in a way as to include communicative acts such as military information operations affecting the civilian population. In view of the object and purpose of the precautions regime entailed in Article 57 API, namely to mitigate impact on the civilian population as much as possible, a more expansive reading seems defensible.[47]

At the same time, even if we accept the applicability of the obligation to take constant care to military information operations in principle, it is questionable how far-reaching this protection really is in view of the possibilities of contemporary digital technologies to deeply affect a target population in a variety of ways. Again, given that IHL is traditionally focused on the violent physical effects of warfare, the question is whether the existing rules still suffice. Jensen and Crockett suggest that the use of deep fake video technology to deceive the civilian population ahead of an attack with kinetic force with the result that the number of incidental civilian casualties rises would violate the obligation.[48] But in situations that are not followed by such destructive events, as in information operations that target democratic decision-making processes or promote a general sense of uncertainty and a loss of trust in media sources or a national information space as a whole (see Scenario B in particular), the protective reach of the rule is much less obvious. After all, even if it is accepted that the notion of military operations can be interpreted broadly to include certain types of military information operations, the question remains what 'sparing the civilian population' means and whether the interpretation can be expanded beyond violent effects in a more traditional sense. While there is no conceptual barrier to such an interpretation, there is hardly any state practice to support it. Opening up the interpretation as to which effects the notion of 'sparing the civilian population' might entail beyond violent effects immediately raises difficult line-drawing and definitional questions. After all, an obligation to avoid all detrimental impacts on the civilian population in times of armed conflict, even considering the relative due diligence nature of the constant care obligation, would be unrealistic and would go too far, certainly in the eyes of most states. Here is not the place to flesh out these issues in full, also considering that by and large the obligation to exercise constant care to spare the civilian population has generally remained somewhat underexplored. For purposes of the present paper, it suffices to conclude that while Article

[44] See also Art. 13(I) AP II and Customary Rule 15.

[45] See International Law Association Study Group, The Conduct of Hostilities and International Humanitarian Law: Challenges of 21st Century Warfare (2017), p. 42 et seq.

[46] See Winther, p. 131.

[47] Likewise Tallinn Manual 2.0, rule 92, para. 2.

[48] Jensen, Eric Talbot, and Summer Crockett, "Deepfakes" and the Law of Armed Conflict: Are They Legal?,' Articles of War, 19 August 2020, https://lieber.westpoint.edu/deepfakes/.

57(1) API and its customary law pendant may impose limits also on military information operations, at the present juncture the exact protective reach of these provisions vis-à-vis digital disinformation campaigns is unclear.

INFORMATION OPERATIONS REACHING THE THRESHOLD OF AN ATTACK

As hinted at above, the last aspect to be considered is the question whether certain information operations may even qualify as 'attacks' within the meaning of IHL, making them directly responsive to the rules on targeting, such as the principle of distinction, the principle of proportionality, and the principle of precautions in attack. In this context, it is noteworthy that most recently, in the context of health-related misinformation campaigns in the course of the COVID-19 pandemic, Milanovic and Schmitt argued that '[d]epending on the scale of the sickness or death caused and the directness of the causal connection, a cyber misinformation operation even could rise to the level of a use of force'.[49] Whereas this contention concerns the *jus ad bellum* rather than the *jus in bello* under scrutiny here, the argument's rationale might be suitable to being applied to the question at hand. As described previously, the Tallinn Manual 2.0 defines a 'cyber attack' as 'a cyber operation, whether offensive or defensive, that is reasonably expected to cause injury or death to persons or damage or destruction to objects'.[50] While it has been argued above that information operations are *per se* analytically distinct from cyber operations, even if conducted by digital means, the same consideration should pertain to this type of conduct.

Therefore, just like other types of military violence, if the causal nexus between an instance of disinformation and physical harm is sufficiently strong so as to render such operation an attack, it 'must respect the distinction, precaution, and proportionality triad'.[51] If this contention is accepted in principle, one might be inclined to make the argument that Scenario C involves an 'attack' by means of an information operation as the false information led members of the armed forces and civilians to ingest harmful methanol. To be sure, causation is of course the decisive issue. Whether or not an 'attack' occurred in this scenario hinges on the question of whether the causal relationship between the piece of information and the death of the persons is sufficiently direct for the operation to be considered an 'attack'. After all, as opposed to a cyber operation against an IT system that triggers a physical chain of events that leads to damage, an instance of disinformation requires the targeted audience to act upon the received information and *because of that* inflict harm on itself. This is in any case an entirely different type of causal connection, and it is not inherently obvious that this type of 'attack' was meant to fall within the ambit of existing IHL. In the context of international criminal law in regard to 'instigation' as a speech act that mentally induces the target audience to act in a harmful manner – which in this sense is similar to disinformation in its causal mechanics – the International Criminal Tribunal for the Former Yugoslavia (ICTY) and the International Criminal Tribunal for Ruanda (ICTR) have held that while it is 'not necessary to demonstrate that the crime would not have occurred without the accused involvement',[52] the (instigating) speech act needs to have been a 'substantially

[49] Milanovic, Marko, and Michael N. Schmitt. 'Cyber Attacks and Cyber (Mis)Information Operations During a Pandemic.' Journal of National Security Law & Policy (2020), p. 19.

[50] Tallinn Manual 2.0, rule 92.

[51] Choudhary, Vishakha, 'The Truth under Siege: Does International Humanitarian Law Respond Adequately to Information Warfare?,' International Law Under Construction – Shaping Sustainable Societies, 21 March 2019, https://grojil.org/2019/03/21/the-truth-under-siege-does-international-humanitarian-law-respond-adequately-to-information-warfare/.

[52] ICTY, Judgment, Kvočka (IT-98-30/I-T), Trials Chamber, 2 November 2001, para. 252.

contributing' factor for the crime to occur.[53] Analogously, one may perhaps ask whether the piece of disinformation *substantially contributed* to the harmful event, in this case the ingestion of the methanol. To be sure, this analogy requires that the standard of causality applied by the Tribunals in the context of 'instigation' is appropriate for the context under scrutiny, i.e. the necessary causal proximity between the piece of disinformation and the harmful event (ingestion of methanol) for the conduct to qualify as an 'attack' within the meaning of IHL. This question does not seem to have been addressed in the literature or in state practice to date, and a different standard might ultimately be considered more suitable. At any rate, the absence of any engagement with the particularities of causation again shows that the modes of military conduct analysed in this paper fall outside the ambit of what traditionally has been considered to be subject to the law of armed conflict.

If one supports the conclusion that the dissemination of disinformation might qualify as an 'attack', it must be asked whether the operation was in compliance with the rules pertaining to the conduct of hostilities. Given that the disinformation was targeted at members of the adversarial armed forces, the principle of distinction was arguably observed. At the same time, it is questionable whether the same holds true as regards proportionality and precautions in attack in view of the fact that it was likely reasonably foreseeable that the harmful disinformation would not stay confined to the soldiers' closed groups on social media but instead further spread to civilian audiences as well. Information is by definition difficult to contain once it has been published.[54]

With reference to the Tallinn Manual 2.0, it has furthermore been suggested that an information operation might also amount to an 'attack' within the meaning of IHL if it merely causes the psychological condition of 'severe mental suffering', which supposedly follows from the phrasing of the already mentioned Article 51(2) AP I, prohibiting acts or threats of violence the primary purpose of which is to spread terror among the civilian population.[55] Certainly, there is no reason to exclude mental injury from the protective ambit of IHL as a matter of principle. The problem, however, is that the degree of mental suffering is difficult to establish, given that, for instance, '[i]nconvenience, irritation, stress, [and] fear are outside of the scope' of the proportionality principle,[56] it should follow that at least not *every* psychological reaction to an information operation can be sufficient to render the conduct an attack. In order to render such an expansive interpretation of the protective rules of IHL workable, one would have to find clear, reliable, and detectable criteria to enable the assessment of mental injury caused by an adversarial information operation. Either way, it is argued that many conceivable operations, as demonstrated by above scenarios, will not lead to a sufficient degree of distress. In this context, it may be suggested that this calculation may shift towards the assumption of 'severe mental suffering' if a large-scale information operation – as for example in Scenario B above – leads to widespread confusion and sustained insecurity among the civilian population of the target state. However, even in such a scenario, the rule's ambit would still be concerned with the mental well-being of (a number of) individual civilians, but not with the *integrity of the targeted information space* as such. Again, it may thus be asked whether existing IHL remains sufficient to adequately protect civilian societies and their digital information spaces against the perils of novel modalities of modern warfare.

[53] ICTR, Judgment, Ndindabahizi (ICTR-2001-71-I), Trials Chamber, 15 July 2004, para. 463; ICTY, Judgment, Kordić and Cerkez (IT-95-14/II-A), Appeals Chamber, 17 December 2004, para. 27; ICTY, Judgment, Orić (IT-03-68-T), Trials Chamber, 30 June 2006, para. 274; ICTR, Judgment, *Nahimana et al.* (ICTR-99-52-A), Appeals Chamber, 28 November 2007, para. 501.

[54] See Choudhary (n 51).

[55] See Tallinn Manual 2.0, rule 92, para. 8.

[56] Tallinn Manual 2.0, rule 113, para. 5.

INTERNATIONAL CRIMINAL LAW

In the context of the use of information operations in situations of armed conflict, it is worth mentioning briefly that some forms of disinformation may not merely constitute breaches of IHL but may also rise to the level of an international crime. For example, disinformation about protected individuals or groups with the aim of instigating members of the armed forces or civilians to attack them can be qualified as inducing a war crime or another crime within the jurisdiction of the International Criminal Court: Article 25(3)(b) of the Rome Statute stipulates that 'a person shall be criminally responsible and liable for punishment for a crime within the jurisdiction of the Court if that person … induces the commission of such a crime which in fact occurs or is attempted'.[57] Recently, the UN Human Rights Council presented a detailed fact-finding report on the situation of the Rohingya in Myanmar that laid out the ways in which dehumanising disinformation can be weaponised in situations of inter-ethnic tensions.[58]

Relatedly, the Rome Statute furthermore provides in Article 25(3)(e) that 'a person shall be criminally responsible and liable for punishment for a crime within the jurisdiction of the Court if that person … directly and publicly incites others to commit genocide'. Incitement, too, is a mode of criminality that can – and often will – be committed by way of disseminating hateful disinformation about a targeted group. Note that as opposed to instigating or inducing the commitment of a crime, incitement does not require the genocide to actually have occurred; for criminal liability to be established, it is sufficient to show that the inciting speech act created the risk of genocidal acts to be carried out by the recipients.[59]

INTERNATIONAL HUMAN RIGHTS LAW

Adversarial information operations are obviously also capable of implicating the human rights of targeted civilian populations. A piece of disinformation disseminated by a state via social media that urges people to ingest methanol in order to avoid contracting a deadly virus *prima facie* violates the right to bodily integrity and the right to life, as guaranteed by virtually all existing human rights treaties such as the International Covenant on Civil and Political Rights (ICCPR) or the European Convention on Human Rights (ECHR).[60] A state-run disinformation campaign that pursues the purpose of interfering in the democratic decision-making process in another state might be considered a violation of the right to vote in elections that guarantee 'the free expression of the will of the electors' (Article 25(b) ICCPR) and of the collective right to self-determination, which is enshrined in Article 1(1) ICCPR as well as Article 1(1) of the International Covenant on Economic, Social and Cultural Rights (ICESCR).[61] More generally, it may even be worth inquiring whether and under which circumstances state-led adversarial disinformation from abroad interferes with a person's right to information pursuant to Article 19(2) ICCPR.

[57] See on this only Coco, Antonio. 'Instigation.' In Modes of Liability in International Criminal Law, edited by Jérôme de Hemptinne, Robert Roth, Elies van Sliedregt, Marjolein Cupido, Manuel J. Ventura and Lachezar Yanev, 257 (2019).

[58] UN Human Rights Council, Report on the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar, UN Doc A/HRC/39/CRP.2 (17 September 2018).

[59] Jens David Ohlin, 'Incitement and Conspiracy to Commit Genocide', in Paola Gaeta (ed), The UN Genocide Convention: A Commentary, 2009, 207, 212.

[60] See Milanovic, Marko, and Michael N. Schmitt. 'Cyber Attacks and Cyber (Mis)Information Operations During a Pandemic.' Journal of National Security Law & Policy (2020), p. 17-19.

[61] See on this Ohlin, Jens David. Election Interference: International Law and the Future of Democracy (2020); Tsagourias, Nicholas, 'Electoral Cyber Interference, Self-Determination and the Principle of Non-Intervention in Cyberspace,' EJIL: Talk! , 26 August 2019, https://www.ejiltalk.org/electoral-cyber-interference-self-determination-and-the-principle-of-non-intervention-in-cyberspace/.

However, the application of these and other human rights is contingent on two conditions: first, it needs to be established whether and to what extent states' human rights obligations apply extraterritorially, in view of the fact that the ICCPR, for example, stipulates that '[e]ach State Party to the present Covenant undertakes to respect and to ensure to all individuals *within its territory and subject to its jurisdiction* the rights recognized in the present Covenant' (Article 2(1) ICCPR).[62] Some authors have recently reemphasised that there are persuasive reasons to assume that states have an obligation not to infringe upon the rights of individuals located in other states given that the digital transformation as well as recent developments of weapons technologies have vastly increased the possibilities of states to endanger and compromise the enjoyment of human rights of persons abroad who otherwise possess no link to the acting state.[63]

Second, information operations and other forms of hybrid warfare add renewed urgency to the question of the relationship between the application of the laws of armed conflict (IHL) and international human rights law in situations of armed conflict. If the current state of the debate is that the rules of IHL should prevail as *lex specialis* wherever they deal more specifically with a subject matter also tackled by human rights – such as the right to life in targeting decisions or the right to personal liberty in decisions on military detention – but leave room for the application of human rights law in relation to other issues that are not explicitly addressed in existing IHL, one may conclude that novel forms of warfare such as the ones presented in this paper allow for a broader consideration of the human rights implications of adversarial military operations. After all, at least election interference or the coercion of individual civilians by way of an information operation is nothing that the Geneva Conventions or their Additional Protocols envisaged – with the possible, albeit in any case limited, exception of the law of military occupation as laid down in GC IV. Of course, a possible and potentially rather sweeping counter-argument against a stronger reliance on human rights protections regarding information operations during armed conflict, could be IHL's explicit recognition of permissible ruses of war. After all, if 'ruses of war are not prohibited', as stated by Article 37(2) AP I, IHL could potentially be invoked as the lex specialis in times of armed conflict whenever an information operation qualifies as a ruse of war. The same provision, however, clarifies that permissible ruses are limited to operations 'which infringe no rule of international law applicable in armed conflict'. This forestalls any sweeping invocations of the lex specialis argument and leaves considerable room for human rights law in the assessment of wartime information operations.

CONCLUSION: THE LIMITS OF EXISTING LAW AND OPTIONS FOR ADVANCING THE DEBATE

A central object and purposes of IHL is the protection of civilian populations against the consequences of armed conflict. IHL's anchoring in 20st century kinetic warfare and its traditional focus on the *physical impact* of military operations still pervades contemporary understandings and interpretations of the humanitarian legal framework. The extent of this 'physical anchoring' marks the linchpin in current debates about accommodating and mitigating the far-reaching intangible harms (potentially) inflicted by 21st century modes of warfare.

Shifts in the nature of conflict have seen an emergence of new modes of hybrid warfare

[62] Also see Article 1 ECHR.

[63] See in the context of disinformation and cyber operations Milanovic, Marko, and Michael N. Schmitt. 'Cyber Attacks and Cyber (Mis)Information Operations During a Pandemic.' Journal of National Security Law & Policy (2020), p. 12 et seq.

combining the employment of traditional kinetic force, cyber operations, and disinformation campaigns to destabilise or gradually demoralise the adversary – as can be witnessed, for example, in Ukraine since 2014. Digital technologies allow for information operations that can deeply affect targeted civilian populations and public structures in ways that were hitherto inconceivable.

On the other hand, it is still very much an open question whether the adverse (intangible) consequences on modern interconnected societies and information spaces are *humanitarian concerns* in the sense that contemporary IHL should be the legal regime addressing them. Are the potential harms laid out in this paper in fact reflective of protective gaps that humanitarian law should fill? If so, should such protection be achieved on the basis of existing rules and via links to traditional forms of violence or physical or mental impacts on individuals? Or are systemic values such as 'the integrity of national or global information spaces' or 'public trust' increasingly to be seen as 21st century humanitarian values that IHL should protect as such – at least against the worst types of impact when disinformation campaigns are designed to systematically corrupt and corrode informational spaces nation-wide?

There are essentially two paths available to move forward from here: One is to accept such adverse consequences as in principle within the ambit of the *raison d'être* of international humanitarian law, which would imply the need for a more progressive re-interpretation and (potentially) development of the existing body of the laws of armed conflict.

The other one is to consider threats from contemporary information operations beyond the (deliberately limited) reach of IHL given that it is the principal task of these rules to provide fundamental protection (rather than full-scale protection) against the worst (and not all) perils of war. In that case, other rules would have to step in lest civil societies were left without clear legal protection against some of the most consequential forms of modern conflict, as exemplified in Scenario B. The long-running but as yet unsettled questions of the extraterritorial ('virtual') and substantive reach of international human rights law in situations of armed conflict however suggest that states remain reluctant to proceed with the second option.

As far as information operations are concerned, however, states so far do not seem to be prepared to treat their consequences as humanitarian concerns either. In part, this may be due to the difficult line-drawing and definitional questions inherent in any attempt at broadening classic IHL understandings to include intangible impacts that for the time might be seen to militate against any such ostensibly 'radical' extensions. In fact, despite growing engagement within the community of international legal scholars, there is a palpable reluctance to address the issue within the framework of international law at all. While there is an increasing trend among states to publicly position themselves in regard to the application of international law to cyber operations, the same cannot be said about the growing phenomenon of adversarial conduct against a target state's information ecosystem, i.e., operations that are carried out solely on the *content layer* of network infrastructures without affecting the physical or logical layers as well. Regarding the legal implications of such operations, states have so far by and large remained silent and abstained from any nuanced categorizations.[64] In line with this reluctance to employ the language of international law, states and regional organisations have so far preferred an approach that focuses on monitoring adversarial campaigns by other states and counter-information to correct distorting or false media narratives.[65]

[64] See Lahmann, Henning, "Information Operations and the Question of Illegitimate Interference Under International Law" (2020) 53 Israel Law Review 189, 209-217.

[65] See as an example the European Union's "EU vs. Disinfo" initiative, https://euvsdisinfo.eu/.

The present paper has shown that digital communications technologies open up entirely new possibilities to affect the adversary, societies and the civilian population of a given area, state, region or even globally in situations of armed conflict. The foregoing analysis of existing legal framework allows for the following conclusions:

(1) Certain kinds of adversarial information operations in a situation of armed conflict and their consequences are covered by existing rules of IHL, in particular in regard to incitement, de-humanization of the adversary, and the terrorisation of a civilian population.

(2) The legal concepts of 'constant care' and 'attack' allow, in principle, for an expansive interpretation that encompasses certain modes of information operations and resulting harms, as exemplified in Scenario C. Such expansive understanding is contingent on corresponding mutual consent of the relevant international actors and should be supported.

(3) Adversarial conduct during armed conflict against the information space of a belligerent party beyond these relatively narrowly circumscribed scenarios finds only scarce (clear) limitations under existing legal frameworks. To a certain extent, this is an expression of IHL's unusually explicit permissive stance on ruses of war and a widespread sentiment that information operations (against the adversary) must remain legal. At the same time, recent developments suggest a significant shift towards more pervasive *epistemic attacks* that may lead to a large-scale corrosion of public information spaces without discernible military necessity. With the ever-increasing digitalization of societies across the globe, the adverse impact of such conduct might be too sustained and too grave to remain unaddressed by IHL. Given the further observation that in information warfare the lines between times of war and times of peace become increasingly blurred, there even appears to be an emerging need – and room – for a broader rule against systematic and highly corrosive military information operations against civilian information spaces that is not limited to situations of armed conflict but spans the entire spectrum of peace and war.

To be sure, we are under no illusion about the prospects of such a rule materialising any time soon. In our view, all of this first and foremost calls for a policy debate about humanitarian values on the future digital battlefield. If anything, we need to move on from the current widespread instinctive perception that any type of information operation not amounting to prohibited perfidy would automatically be permissible during armed conflict. In view of the possibilities and adverse impacts of digital information warfare in the 21st century, and for the sake of protecting civilian societies in the digital era, such an attitude can no longer reasonably be upheld. In our view, avoiding or mitigating the worst and most disruptive impacts digital information warfare can have on civilian populations and societies, should be considered a central humanitarian objective in 21st century warfare.

The Geneva Academy of International Humanitarian Law and Human Rights

The Geneva Academy provides post-graduate education, conducts academic legal research and policy studies, and organizes training courses and expert meetings. We concentrate on branches of international law that relate to situations of armed conflict, protracted violence, and protection of human rights.

Disruptive Military Technologies

New (military) technologies are set to revolutionize the ways wars are fought. This research project aims at staying abreast of the various military technology trends; promoting legal and policy debate on new military technologies; and furthering the understanding of the convergent effects of different technological trends shaping the digital battlefield of the future.

The Geneva Academy of International Humanitarian Law and Human Rights

Villa Moynier
Rue de Lausanne 120B
CP 1063 - 1211 Geneva 1 - Switzerland
Phone: +41 (22) 908 44 83
Email: info@geneva-academy.ch
www.geneva-academy.ch