International Review of the Red Cross (2020), **102** (914), 659–682.
Emerging Voices
doi:10.1017/S1816383121000473

IRRC

Liar's war: Protecting civilians from disinformation during armed conflict

Eian Katz
Eian Katz is a Legal Adviser at the American Bar Association's Center for Human Rights and previously served as Counsel at the Public International Law and Policy Group. He holds a JD from the University of Chicago and a BA from Yale University.

Abstract

Disinformation in armed conflict may pose several distinctive forms of harm to civilians: exposure to retaliatory violence, distortion of information vital to securing human needs, and severe mental suffering. The gravity of these harms, along with the modern nature of wartime disinformation, is out of keeping with the traditional classification of disinformation in international humanitarian law (IHL) as a permissible ruse of war. A patchwork set of protections drawn from IHL, international human rights law and international criminal law may be used to limit disinformation operations during armed conflict, but numerous gaps and ambiguities undermine the force of this legal framework, calling for further scholarly attention and clarification.

Keywords: disinformation, fake news, influence operations, international humanitarian law, international human rights law, international criminal law, ruse of war, armed conflict.

Introduction

According to the World Risk Poll, "fake news" topped the list of concerns for internet users in 2020.[1] This finding should come as no surprise to those who have followed with growing consternation the proliferation of false or misleading

© The Author(s), 2021. Published by Cambridge University Press on behalf of the ICRC

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

information dirtying the arenas of politics,[2] health,[3] finance[4] and other sectors. Less noticed amid the handwringing have been the unique effects of fake news on civilians in armed conflict settings. Whereas belligerents have long recognized the military value of disinformation for deceiving their enemies, in the age of hybrid warfare it is increasingly produced for a civilian audience. Positioning civilians as the subject or object of disinformation or propaganda may fan the flames of violence and may further lead to three distinct and cognizable forms of civilian harm.

First, fabrications that vilify disfavoured individuals or groups may serve to encourage and legitimate foreseeable acts of violence perpetrated against them by third parties. Perhaps the most infamous such episode occurred in the prelude to the Rwandan genocide, when radio stations owned by the State or by high-ranking State officials inflamed popular sentiment against the Tutsis through fake news reports.[5] In other conflicts, organized smear campaigns have likewise endangered minority populations or humanitarian workers.[6]

A second unsettling application of disinformation that conflict parties have recently embraced is in distorting information vital to securing human needs. Through widespread fearmongering and deliberate obfuscation, hybrid actors aim to sow dissent, undermine the social order, aggravate crises and discredit enemy institutions, including civilian ones. Throughout the COVID-19 pandemic, for instance, various conflict actors have spread disinformation about the virus, underreported case counts or boosted conspiracy theories for political and economic gain.[7] Disinformation in armed conflict may likewise be directed against humanitarian organizations providing relief to beleaguered civilians,

1 World Risk Poll, "Fake News Is the Number One Worry for Internet Users Worldwide", Lloyd's Register Foundation, 2020, available at: https://wrp.lrfoundation.org.uk/explore-the-poll/fake-news-is-the-number-one-worry-for-internet-users-worldwide/ (all internet references were accessed in September 2021).

2 See, for example, Danielle Abril, "Facebook Reveals that Massive Amounts of Misinformation Flooded Its Service during the Election", *Fortune*, 19 November 2020, available at: https://fortune.com/2020/11/19/facebook-misinformation-labeled-180-million-posts-2020-election-hate-speech-prevalence/.

3 See, for example, Rebecca Heilweil, "Facebook Is Finally Cracking Down Hard on Anti-Vaccine Content. It Is Facing an Uphill Battle", *Vox*, 9 March 2021, available at: www.vox.com/recode/22319681/vaccine-misinformation-facebook-instagram-spreading.

4 See, for example, Michael Siering, Jan Muntermann and Miha Grčar, "Design Principles for Robust Fraud Detection: The Case of Stock Market Manipulations", *Journal of the Association for Information Systems*, Vol. 22, No. 1, 2021.

5 Montreal Institute for Genocide and Human Rights Studies, "Rwanda Radio Transcripts: The Role of Radio", available at: www.concordia.ca/research/migs/resources/rwanda-radio-transcripts.html.

6 See, for example, *Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar*, UN Doc. A/HRC/39/CRP.2, 17 September 2018, para. 1324; Louisa Loveluck, "Russian Disinformation Campaign Targets Syria's Beleaguered Rescue Workers", *Washington Post*, 18 December 2018, available at: https://tinyurl.com/zk9x8cjt.

7 See, for example, Sonny S. Patel, Omar E. Moncayo, Kristina M. Conroy, Doug Jordan and Timothy B. Erickson, "The Landscape of Disinformation on Health Crisis Communications during the COVID-19 Pandemic in Ukraine: Hybrid Warfare Tactics, Fake Media News, and Review of Evidence", *Journal of Science Communication*, Vol. 19, No. 5, 2020; Evan Lawson and Natia Seskuria, "Disinformation Should Be the Final Straw in the Targeting of Medical Facilities", *The Russia File*, 17 August 2020, available at: www.wilsoncenter.org/blog-post/disinformation-should-be-final-straw-targeting-medical-facilities; Human Rights Watch, "Yemen: Houthis Risk Civilians' Health in Covid-19", 1 June 2021, available at: www.hrw.org/news/2021/06/01/yemen-houthis-risk-civilians-health-covid-19.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

jeopardizing their operations and sullying their reputations.[8] These rampant misrepresentations can place civilians in harm's way by disrupting access to services or by leading them to act in ways counter to their own interests.

Finally, disinformation may directly harm the mental health of civilians by arousing extreme fear, grief or other painful emotions or unsound mental states. Disinformation geared toward civilian audiences might lead them to develop paranoia or conspiratorial thinking, doubt in their continued ability to satisfy their human needs, a belief that friends or relatives have been or will be harmed, or a reasonable apprehension of death or bodily injury. These harms are less amenable to documentation and proof, but may be no less traumatic, damaging and enduring.

Often, disinformation produced during armed conflict differs little in content from that produced in other contexts. Even so, there are salient practical and legal distinctions. Practically, conflict-affected persons may be especially vulnerable to the ill effects of disinformation due to their desperate living conditions, elevated exposure to violence and decreased access to trustworthy sources of information. Compounding the potential for harm is the fact that conflict environments have proven especially fertile incubators for disinformation, as evidenced by its recent flourishing in Israel and Palestine, Libya, South Sudan, Syria, Ukraine and Yemen.[9] Military and civilian actors alike have made gainful use of disinformation in order to control the narrative regarding the conflict. The deep civil fragmentation characteristic of conflict settings amplifies the effect of such disinformation by creating echo chambers that heighten confirmation bias and accelerate the uncritical sharing of specious reports.[10] Moreover, the collapse of State institutions amid conflict weakens content regulation networks and limits the availability of judicial remedies for confronting libel,[11] leaving few restraints on the spread of disinformation.

8 Rachel Xu, "You Can't Handle the Truth: Misinformation and Humanitarian Action", *Humanitarian Law and Policy Blog*, 15 January 2021, available at: https://blogs.icrc.org/law-and-policy/2021/01/15/misinformation-humanitarian/.

9 Sheera Frenkel, "Lies on Social Media Inflame Israeli-Palestinian Conflict", *New York Times*, 14 May 2021, available at: www.nytimes.com/2021/05/14/technology/israel-palestine-misinformation-lies-social-media.html; "Fake News War: In Libya, Battles Also Rage on Social Media", *France 24*, 18 April 2019, available at: www.france24.com/en/20190418-fake-news-war-libya-battles-also-rage-social-media; Mel Bunce, "Humanitarian Communication in a Post-Truth World", *Journal of Humanitarian Affairs*, Vol. 1, No. 1, 2019; Fatima K. Abu Salem, Roaa Al Feel, Shady Elbassuoni, Mohamad Jaber and May Farah, "FAKES: A Fake News Dataset around the Syrian War", American University of Beirut, 2019, available at: https://ojs.aaai.org/index.php/ICWSM/article/view/3254/3122; Margarita Jaitner, *Russian Information Warfare: Lessons from Ukraine*, NATO Cooperative Cyber Defence Centre of Excellence, 2020, available at: https://ccdcoe.org/uploads/2018/10/Ch10_CyberWarinPerspective_Jaitner.pdf; Can Kasapoglu and Mariam Fekry, *Iran's Proxy War in Yemen: The Information Warfare Landscape*, NATO Strategic Communications Centre of Excellence, January 2020, available at: https://stratcomcoe.org/pdfjs/?file=/cuploads/pfiles/web_stratcom_coe_iran_proxy_war_against_yemen_13-02-2020.pdf?zoom=page-fit.

10 Filippo Menczer, "Fake Online News Spread through Social Echo Chambers", *Scientific American*, 28 November 2016, available at: www.scientificamerican.com/article/fake-online-news-spreads-through-social-echo-chambers/.

11 Cf. Peter Roudik *et al.*, *Initiatives to Counter Fake News: Comparative Summary*, US Library of Congress, April 2019, available at: https://digitalcommons.unl.edu/scholcom/179/.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

Legally, disinformation during armed conflict is unique in being governed by international humanitarian law (IHL), which has long characterized it as an allowable "ruse of war" (provided that it complies with other applicable rules).[12] But, it shall be argued, this permissive attitude is the product of a bygone era in which deceptive practices could be safely supposed to yield tactical or political gains without significantly impacting the condition of non-combatants. The contemporary trend of disinformation being deployed in ways that threaten civilian populations challenges that underlying assumption and begs for a re-evaluation of its treatment in IHL.

The primary enterprise of this article, therefore, is to pioneer a new legal approach to disinformation in armed conflict by mining the corpus of international law for rules that might constrain it. The first part of the article begins with a review of the traditional IHL stance towards disinformation and contends that this stance is no longer in keeping with disinformation's modern manifestations. After discussing two preliminary issues in the second part, the remainder of the article proceeds to search for other applicable protections with respect to the three forms of harm described above: exposure to foreseeable acts of violence, distortion of information concerning vital human needs, and mental suffering. In most cases, the applicability of protections under IHL, international human rights law (IHRL) or international criminal law is speculative and incomplete at best, calling for greater clarification in the law and focused attention on addressing this problem.

The scope of this article is confined by two principal limits. First, it is concerned only with the publication of false information, not advocacy of violence. It therefore does not delve deeply into the law of incitement, which is the subject of substantial existing case law and scholarship. Incitement is only examined herein insofar as it might arguably describe incendiary falsehoods published with the intent to provoke a violent reaction, even absent an explicit call for violence. Second, this article only deals with disinformation from a *jus in bello* perspective. Other qualified commentators have opined on disinformation as a use of force or as a violation of sovereignty,[13] and have examined its ordinary treatment in IHRL.[14]

A word on terminology: "disinformation" and "misinformation" both refer to false information but are separated by an element of intentionality in the former:

12 Protocol Additional (I) to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, 1125 UNTS 3, 8 June 1977 (entered into force 7 December 1978) (AP I), Art. 37(2); Michael N. Schmitt (ed.), *Tallinn Manual on the International Law Applicable to Cyber Warfare*, Cambridge University Press, Cambridge, 2013 (Tallinn Manual), Rule 61, para. 2.

13 Marko Milanovic and Michael N. Schmitt, "Cyber Attacks and Cyber (Mis)Information Operations during a Pandemic", *Journal of National Security Law and Policy*, Vol. 11, No. 1, 2020; Gary Corn, "Coronavirus Disinformation and the Need for States to Shore Up International Law", *Lawfare*, 2 April 2020, available at: www.lawfareblog.com/coronavirus-disinformation-and-need-states-shore-international-law.

14 Irene Khan, *Disinformation and Freedom of Opinion and Expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, UN Doc. A/HRC/47/25, 13 April 2021.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

disinformation is "deliberately misleading or biased".[15] "Fake news" is often a form of disinformation or misinformation but is poorly defined and therefore used sparingly here. Other related concepts, such as "propaganda",[16] "psychological operations",[17] "influence operations",[18] "information operations",[19] "information warfare"[20] and "cognitive warfare",[21] may also feature some combination of disinformation, misinformation, misleading information and accurate information. For the sake of clarity, this article avoids them.

Disinformation in armed conflict

The deployment of disinformation as a military tactic is hardly a modern innovation. Writing in the fifth century BC, Sun Tzu maintained that "all warfare is based on deception". Classical military history is replete with storied battlefield artifices, conjuring the legendary tales of the Trojan Horse, Alexander the Great's surprise crossing of the Hydaspes River, and Hannibal's ambush at Lake Trasimene. In the modern period, subterfuge and deceit remain critical components of military strategy, facilitated by increasingly sophisticated communications technologies.

IHL takes a remarkably lenient approach to such stratagems, with one exception. Combatants may not resort to perfidy, defined in Article 37(1) of Additional Protocol I to the Geneva Conventions (AP I) as "inviting the confidence of an adversary to lead him to believe that he is entitled to, or is

15 "Disinformation", *Dictionary.com*, available at: www.dictionary.com/browse/disinformation.

16 "Any form of adversary communication, especially of a biased or misleading nature, designed to influence the opinions, emotions, attitudes, or behavior of any group in order to benefit the sponsor, either directly or indirectly." US Department of Defense, *Joint Publication 3-13.2: Psychological Operations*, 27 January 2010, p. GL-7.

17 "Planned operations to convey selected information and indicators to foreign audiences to influence their emotions, motives, objective reasoning, and ultimately the behavior of foreign governments, organizations, groups, and individuals. The purpose of psychological operations is to induce or reinforce foreign attitudes and behavior favorable to the originator's objectives." *Ibid.*, p. GL-8.

18 "The coordinated, integrated, and synchronized application of national diplomatic, informational, military, economic, and other capabilities in peacetime, crisis, conflict, and postconflict to foster attitudes, behaviors, or decisions by foreign target audiences that further [partisan] interests and objectives." Eric V. Larson *et al.*, *Foundations of Effective Influence Operations: A Framework for Enhancing Army Capabilities*, RAND Arroyo Center, Santa Monica, CA, 2009, p. xii.

19 "Any coordinated or individual deployment of digital resources for cognitive purposes to change or reinforce attitudes or behaviours of the targeted audience." Dapo Akande *et al.*, "Oxford Statement on International Law Protections in Cyberspace: The Regulation of Information Operations and Activities", *Just Security*, 2 June 2021, available at: www.justsecurity.org/76742/oxford-statement-on-international-law-protections-in-cyberspace-the-regulation-of-information-operations-and-activities/.

20 "The application of destructive force on a large scale against information assets and systems, against the computers and networks that support the four critical infrastructures (the power grid, communications, financial, and transportation)." Brian C. Lewis, "Information Warfare", Federation of American Scientists, available at: https://fas.org/irp/eprint/snyder/infowarfare.htm.

21 "Manoeuvres in the cognitive domain to establish a predetermined perception among a target audience in order to gain advantage over another party." Paul Ottewell, "Defining the Cognitive Domain", *Over the Horizon*, 7 December 2020, available at: https://othjournal.com/2020/12/07/defining-the-cognitive-domain/.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

obliged to accord, protection under the rules of international law applicable in armed conflict, with intent to betray that confidence".[22] Belligerents thus may not, for example, disingenuously accept a truce or announce their capitulation, feign injury or illness, or claim civilian or other protected status.[23]

In the next paragraph of Article 37, however, AP I explicitly permits "ruses of war" – "acts which are intended to mislead an adversary or induce him to act recklessly but which infringe no rule of international law applicable in armed conflict and which are not perfidious".[24] While this particular provision only pertains to international armed conflicts (IACs), ruses are likewise licensed in non-international armed conflicts (NIACs) under customary international law.[25]

"Misinformation" is expressly classified in AP I as a type of permissible ruse in armed conflict.[26] Several other authoritative sources confirm this posture throughout IHL. The International Committee of the Red Cross (ICRC) Commentary on AP I categorizes "circulating misleading messages" as a form of ruse,[27] and numerous national military manuals likewise endorse the use of "misinformation", "disinformation", "false information", "psychological operations" or "bogus dispatches and newspapers".[28] The *Tallinn Manual on the International Law Applicable to Cyber Warfare* (Tallinn Manual), the leading treatise on the application of international law to cyberspace, also enumerates examples of acceptable ruses that include "psychological warfare activities" such as "dropping leaflets or making propaganda broadcasts".[29] In customary law, as conveyed by the ICRC, ruses are defined even more expansively than in AP I – as "acts intended to confuse the enemy"[30] – such that there can be little doubt that misinformation would be included. Following this logic, many scholars have presumed that IHL constrains disinformation in armed conflict only via the rule against perfidy and other background principles such as necessity, proportionality and distinction.[31]

But modern applications of misinformation – or, more accurately, disinformation – call into question its reflexive characterization in IHL as a ruse

22 AP I, Art. 37(1).
23 *Ibid.*, Art. 37(1).
24 *Ibid.*, Art. 37(2).
25 Jean-Marie Henckaerts and Louise Doswald-Beck (eds), *Customary International Humanitarian Law*, Vol. 1: *Rules*, Cambridge University Press, Cambridge, 2005 (ICRC Customary Law Study), Rule 57, available at: https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1.
26 AP I, Art. 37(2).
27 Yves Sandoz, Christophe Swinarski and Bruno Zimmerman (eds), *Commentary on the Additional Protocols*, ICRC, Geneva, 1987 (ICRC Commentary on APs), para. 1516.
28 Jean-Marie Henckaerts and Louise Doswald-Beck (eds), *Customary International Humanitarian Law*, Vol. 2: *Practice*, Cambridge University Press, Cambridge, 2005 (ICRC Customary Law Study (Practice)), Practice relating to Rule 57, available at: https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2.
29 Tallinn Manual, above note 12, Rule 61, para. 2; Rule 31, para. 5.
30 ICRC Customary Law Study, above note 25, Rule 57.
31 See Duncan B. Hollis, "Why States Need an International Law for Information Operations", *Lewis & Clark Law Review*, Vol. 11, No. 4, 2007; Joshua E. Kastenberg, "Tactical Level PSYOP and MILDEC Information Operations: How to Smartly and Lawfully Prime the Battlefield", *The Army Lawyer*, July 2007, pp. 67–70; Peter M. Smyczek, "Regulating the Battlefield of the Future: The Legal Limitations on the Conduct of Psychological Operations (PSYOP) under Public International Law", *Air Force Law Review*, Vol. 57, No. 1, 2005.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

for several reasons. First, while ruses are presented in source materials as being intended to mislead an "enemy" or "adversary", disinformation campaigns during armed conflict today are instead often oriented primarily towards the civilian population.[32] While this does not automatically render them illegal,[33] it highlights a glaring descriptive deficiency in the traditional IHL conception.

More problematic still for the classification of disinformation as a permissible ruse under IHL is the requirement that such acts "infringe no rule of international law applicable in armed conflict".[34] This proviso demands a fuller analysis of the relevant aspects of IHL, IHRL and international criminal law governing the spread of disinformation during armed conflict, an inquiry that this article will take up later on. As shall be shown, certain forms of disinformation propagated during armed conflict may in fact be in breach of international law.

Finally, the inaptness of the definition of disinformation points to larger issues related to the ability of IHL to adapt to technological developments. Modern strains of disinformation differ both in degree and in kind from their progenitors. Mass media facilitates the generation and circulation of disinformation at unprecedented levels, such that sorting fact from fiction becomes exceedingly difficult.[35] This phenomenon is sometimes referred to as an "infodemic".[36] Relatedly, technological advancements have made fabrications appear increasingly realistic.[37] Thus, whereas the ICRC once supposed a ruse to be "very often the only course open to a weak combatant",[38] disinformation today is often employed by highly sophisticated combatants and non-combatants using complex methods. For these reasons, it no longer makes sense to label disinformation as a "ruse of war" without fuller consideration of the circumstances.

Preliminary considerations

Accepting the contention that disinformation during armed conflict should not be afforded a blanket permit under international law as a ruse of war, it must be considered how else it is to be governed. The remainder of this article explores this topic. It finds that there is no single rule, but rather a patchwork set of protections cobbled together from different branches of international law that differ substantively with respect to the aforementioned three categories of harm: exposure to foreseeable acts of violence, distortion of information concerning vital human needs, and mental suffering. But the straitjacketed application of

32 It is possible that the term "enemy" might encompass non-combatants – IHL sources use "enemy civilians", "enemy nationals" and "enemy aliens" without necessarily referring to combatants. "Adversary" is more narrowly confined to a military context.
33 See Tallinn Manual, above note 12, Rule 31, para. 5.
34 AP I, Art. 37(2).
35 R. Xu, above note 8.
36 "Infodemic", *Dictionary.com*, available at: www.dictionary.com/browse/infodemic.
37 See Lisa M. Cohen, "The New Era of Disinformation Wars", *Völkerrechtsblog*, 30 November 2020, available at: https://voelkerrechtsblog.org/the-new-era-of-disinformation-wars/.
38 ICRC Commentary on APs, above note 27, para. 1514.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

these protections leaves considerable legal blind spots through which disinformation may continue to exact a civilian toll.

Before embarking upon this legal analysis, two threshold issues must be addressed. The first is whether the protections of IHRL are applicable in this context at all, and if so, with what limitations. The second is whether persons and objects ordinarily protected by IHL surrender their privileged status by participating in disinformation operations.

Applicability of IHRL

Several technical considerations complicate the continued applicability of human rights law during armed conflict. The first is its relationship to IHL. During armed conflict, IHRL and IHL apply concurrently.[39] In cases of overlap or conflict between them, however, IHL generally prevails as *lex specialis*.[40] Alternatively, when one corpus does not address a particular issue, the other governs by default. It is asserted here that the provisions of IHL accepting ruses of war do not apply to disinformation in its modern form.[41] The parts of the article that follow further demonstrate that, on the whole, there are few other IHL rules regulating such disinformation. This relative silence should admit of greater reliance on IHRL in this space.

The involvement of non-State actors or of States acting outside their own territories in producing disinformation during armed conflict introduces additional wrinkles to the analysis. IHRL is primarily addressed to sovereigns, and non-State actors are usually not counted among the subjects of human rights treaties.[42] Nonetheless, non-State actors are bound at minimum to abide by *jus cogens* norms,[43] such as the right to life.[44] Moreover, the United Nations (UN) considers non-State actors to be subject to IHRL when they exercise governmental functions or exert control over a territory or population.[45] But

39 UN Office of the High Commissioner for Human Rights (UN Human Rights), *International Legal Protection of Human Rights in Armed Conflict*, UN Doc. HR/PUB/11/01, 2011, pp. 55–58. See also International Court of Justice (ICJ), *Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory*, Advisory Opinion, 9 July 2004, para. 106.

40 UN Human Rights, above note 39, p. 58.

41 See the above section "Disinformation in Armed Conflict".

42 Notable exceptions include the Optional Protocol to the Convention on the Rights of the Child on the Involvement of Children in Armed Conflict, 39 ILM 1285, 25 May 2000, Art. 4(1), available at: www.refworld.org/docid/47fdfb180.html; African Union Convention for the Protection and Assistance of Internally Displaced Persons in Africa, 52 ILM 397, 23 October 2009, Art. 7, available at: www.refworld.org/docid/4ae572d82.html

43 *Report of the Independent International Commission of Inquiry on the Syrian Arab Republic*, UN Doc. A/HRC/19/69, 22 February 2012, para. 106.

44 W. Paul Gormley, "Book Review: *The Right to Life in International Law*", *Denver Journal of International Law & Policy*, Vol. 16, No. 1, 2020; Christof Heyns and Thomas Probert, "Securing the Right to Life: A Cornerstone of the Human Rights System", *EJIL: Talk!*, 11 May 2016, available at: https://tinyurl.com/pj63erfs.

45 Anyssa Bellal, *Human Rights Obligations of Armed Non-State Actors: An Exploration of the Practice of the UN Human Rights Council*, Geneva Academy of International Humanitarian Law and Human Rights, December 2016, p. 26.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

there is no clear legal source for these imputed obligations, leaving the applicability of IHRL to non-State actors an unresolved question.[46] While non-State groups that produce disinformation would therefore be required to respect the right to life everywhere as *jus cogens*, the effect of other human rights standards within their areas of control is a matter of debate.

States may also produce disinformation for consumption by civilians beyond their territorial limits. Whether IHRL limits this activity depends upon the theory of extraterritorial jurisdiction applied. International courts have mostly hewn to the "effective control" or "personal jurisdiction" models,[47] according to which it is unlikely that human rights law would apply in these circumstances. Some scholars have instead advocated for frameworks based upon the distinction between positive and negative rights[48] or a functional analysis.[49] These more flexible approaches would be more accommodating of extraterritorial jurisdiction in human rights law as pertains to wartime disinformation operations, but they are merely *de lege ferenda* at present.

Disinformation as direct participation in hostilities

Combatants are not the only actors who have weaponized information in armed conflict. Public officials, bureaucrats and civilian branches of government may be involved in the production and dissemination of disinformation. Private citizens and other non-State actors, whether acting in the service of the conflict actors or in their individual capacities, also have a prominent role, and in some cases have been shown to be the primary drivers of disinformation in conflict.[50] In several contemporary armed conflicts, the combatants have deputized "electronic armies", "troll farms" or "web brigades" to spew disinformation.[51]

46 ICRC, *International Humanitarian Law and the Challenges of Contemporary Armed Conflicts*, Geneva, 2019, p. 54, available at: www.icrc.org/sites/default/files/document/file_list/challenges-report_ihl-and-non-state-armed-groups.pdf.

47 See, for example, ICJ, *Armed Activities on the Territory of the Congo (Congo v. Uganda)*, Judgment, 19 December 2005, paras 179, 216–217; European Court of Human Rights, *Al-Skeini v. United Kingdom*, Appl. No. 55721/07, Judgment, July 7, 2011, para. 137.

48 Marko Milanovic, "Human Rights Treaties and Foreign Surveillance: Privacy in the Digital Age", *Harvard International Law Journal*, Vol. 56, No. 1, 2015.

49 Yuval Shany, "Taking Universality Seriously: A Functional Approach to Extraterritoriality in International Human Rights", *The Law & Ethics of Human Rights*, Vol. 7, No. 1, 2013. See also Maria L. Banda, "Regime Congruence: Rethinking the Scope of State Responsibility for Transboundary Environmental Harm", *Minnesota Law Review*, Vol. 105, No. 1, 2019 (positing a theory of extraterritorial application of human rights based on their "direct effects").

50 Yevgeniy Golovchencko, Mareike Hartmann and Rebecca Adler-Nissen, "State, Media and Civil Society in the Information Warfare over Ukraine: Citizen Curators of Digital Disinformation", *International Affairs*, Vol. 94, No. 5, 2018.

51 See, for example, Wolfram Lacher, "Drones, Deniability, and Disinformation: Warfare in Libya and the New International Disorder", *War on the Rocks*, 3 March 2020, available at: https://warontherocks.com/2020/03/drones-deniability-and-disinformation-warfare-in-libya-and-the-new-international-disorder/; Mona Alami, "Russia's Disinformation Campaign Has Changed How We See Syria", Atlantic Council, 4 September 2018, available at: www.atlanticcouncil.org/blogs/syriasource/russia-s-disinformation-campaign-has-changed-how-we-see-syria/; Andrei Soshnikov, "Inside a Pro-Russian Propaganda Machine in Ukraine", *BBC News*, 13 November 2017, available at: www.bbc.com/news/

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

Civilians and civilian objects – defined in the negative as, respectively, persons who are not members of the armed forces[52] and objects which are not military objectives[53] – are afforded a protected status under IHL, but it is not immutable. Civilian protections may be forfeited by direct participation in hostilities, which consists of three elements: (1) a minimum threshold of harm, including "death, injury, or destruction"; (2) "a direct causal link between the act and the harm"; and (3) the act must be undertaken in support of a party to the conflict (the so-called belligerent nexus).[54]

Disinformation harming civilian life or civilian objects during armed conflict may satisfy the first and third elements, but it is likely to fail on the causal element, save for mental injuries inflicted directly. According to the ICRC, direct causation means involvement in the actual conduct of hostilities, not merely contributing to the war effort or to war-sustaining activities.[55] Disrupting the supply of electricity, water or food, as disinformation concerning vital human needs might, does not qualify as direct participation in hostilities.[56] Nor does political propaganda,[57] which has analytical similarities to disinformation. On the other hand, publishers may become legitimate military objectives if they engage in incitement.[58] Depending on the context, disinformation may be more akin to propaganda, incitement, or neither.

Even if publishing disinformation were in some cases treated as direct participation in hostilities, it is difficult to imagine combat action being directed against individual social media users. Civilian objects engaged in disinformation, such as broadcasting stations or government bureaus, would make for more realistic targets. However, the protected status of civilian objects would only be revoked for the period of time during which they were producing disinformation, which may be very brief in each instance.[59] Nor is there a clear standard in IHL defining the duration of direct participation amid repeated occurrences,[60] such as by frequent publishers of disinformation. For these reasons, it would be exceedingly difficult to legally justify military operations against civilians or civilian objects directly participating in disinformation.

blogs-trending-41915295; "Online Trolls and Fake Accounts Poison Arab Social Media", *BBC News*, 31 August 2018, available at: www.bbc.com/news/technology-45372272.

52 AP I, Art. 50; ICRC Customary Law Study, above note 25, Rule 5.

53 AP I, Art. 52; ICRC Customary Law Study, above note 25, Rule 9.

54 ICRC, *Interpretive Guidance on the Notion of Direct Participation in Hostilities*, Geneva, 2009 (ICRC Interpretive Guidance), Part 2.B.V, available at: https://casebook.icrc.org/case-study/icrc-interpretive-guidance-notion-direct-participation-hostilities. See also Federal Republic of Germany, *On the Application of International Law in Cyberspace*, March 2021, p. 8, available at: https://tinyurl.com/4v5kcste.

55 ICRC Interpretive Guidance, above note 54, Part 2.B.V.2.a.

56 *Ibid.*, Part 2.b.V.2.a.

57 *Ibid.*, Part 2.B.V.1.b.

58 International Criminal Tribunal for the former Yugoslavia (ICTY), *Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign against the Federal Republic of Yugoslavia*, 8 June 2000, paras 55, 76.

59 Iphigenia Fisentzou, "Social Media in Armed Conflict", *Justis*, 15 April 2019, available at: www.justis.com/blurred-lines/#_ftnref20.

60 Tallinn Manual, above note 12, Rule 35, paras 10–11.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

Exposure to foreseeable acts of violence

This article now turns to its core project: evaluating the restrictions placed by international law upon disinformation in armed conflict with respect to the various types of civilian harm that it may cause. This part of the article directs its scrutiny toward the first type: disinformation about an individual or group that is liable to inspire others to react violently against them. Few legal rules restrain this behaviour. Despite causing harm of a magnitude accordant to an "attack", the inherent attenuation in agency makes it unlikely that this form of disinformation would be designated as such under IHL. Instead, disinformation that leads to foreseeable violations of the Geneva Conventions would be prohibited under common Article 1, but only with respect to perpetrators over whom the conflict actor "exercises authority". Nor are disinformation operations likely to be treated as incitement unless they are accompanied by advocacy of violence. That said, it is possible that the individuals responsible for them could be charged as criminal accessories.

Violence and the problem of indirect harm

Among the most fundamental rules in IHL is that civilians "shall not be the object of attack".[61] Were disinformation deemed an "attack", therefore, civilians and civilian objects would be legally shielded from foreseeable and non-incidental injuries, deaths, destruction or damage caused by it.[62] The conflict parties would likewise be barred from employing disinformation indiscriminately, without a specific military objective in mind.[63] Furthermore, civilians in armed conflict are generally protected from murder, torture or other violence to life or well-being.[64]

Attacks are defined in AP I as "acts of violence against the adversary, whether in offence or defence".[65] "Violence" does not only mean uses of physical force;[66] possible attacks are instead commonly evaluated with reference to their expected effects. For persons, these effects must include "injury or death" in order to be considered an attack.[67] As confirmed by the cases of Myanmar, Rwanda and Syria,[68] disinformation in armed conflict may very well cause these sorts of harms to occur, but the inability to attribute these harms to the publisher

61 AP I, Art. 51; Protocol Additional (II) to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non-International Armed Conflicts, 1125 UNTS 609, 8 June 1977 (entered into force 7 December 1978) (AP II), Art. 13.
62 Tallinn Manual, above note 12, Rule 32, paras 5–6.
63 AP I, Art. 51(4).
64 *Ibid.*, Art. 75; AP II, Art. 4; common Art. 3.
65 AP I, Art. 49(1).
66 Tallinn Manual, above note 12, Rule 30, para. 3; ICRC, *Commentary on the First Geneva Convention: Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field*, 2nd ed., Geneva, 2016 (ICRC Commentary on GC I), paras 255–256 (on common Article 2).
67 Tallinn Manual, above note 12, Rule 30.
68 See above notes 5–6.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

makes it unlikely that disinformation would be treated as an attack or other act of physical violence under IHL.

Disinformation defaming particular individuals or groups only leads to physical injury when others choose to act upon it. To say that the publisher has attacked the civilians in question would therefore be to attribute to it the violent response of third parties. There are certain circumstances in which harms may in fact be ascribed to the principal despite the involvement of intervening actors. The conduct of persons or groups acting under the instruction, direction or control of a State, for instance, legally accrues to the State.[69] But disinformation by itself does not ordinarily entail direction or compulsion[70] to commit an offending act; it simply supplies or reinforces the motive.

Another way to look at the relationship between disinformation and civilian harm would be through the lens of causation. As described above, in the context of direct participation in hostilities, the ICRC has construed direct causation to require that "the harm in question must be brought about in one causal step".[71] Merely providing the motive to effect harm, as disinformation might, is not enough.[72] Another view that has been put forward is the notion of an "indirect attack" committed by combatants who wilfully expose civilians to danger from the otherwise lawful action of their adversaries, such as by using them as human shields.[73] Disinformation may analogously place civilians at risk of violence exacted by third parties, though not the sort that would ever be legally authorized. It thus seems unlikely, for reasons related to attribution and causation, that disinformation would be labelled an unlawful attack or other physically violent act.

Common Article 1

In common Article 1, the parties pledge to "ensure respect" for the Geneva Conventions. The ICRC Commentary interprets this duty as applying to "the whole population over which a High Contracting Party exercises authority", including "private persons whose conduct is not attributable to the State".[74] Combatants must, therefore, "take measures to ensure respect for the Conventions by private persons".[75]

The obligation to "ensure respect" bears both positive and negative aspects. On the positive side of the equation, conflict parties must affirmatively act to prevent

69 International Law Commission, *Responsibility of States for Internationally Wrongful Acts*, 2001, Art. 8.
70 The operative standard of control is a matter of debate, but is not likely to encompass disinformation even when viewed in the most generous light. See Marko Milanovic, "Special Rules of Attribution of Conduct in International Law", *International Law Studies*, Vol. 96, 2020, pp. 317–324.
71 ICRC Interpretive Guidance, above note 54, Part 2.b.V.2.b.
72 *Ibid.*: "Therefore, individual conduct that merely builds up or maintains the capacity of a party to harm its adversary, or which otherwise only indirectly causes harm, is excluded."
73 Geoffrey S. Corn, "Beyond Human Shielding: Civilian Risk Exploitation and Indirect Civilian Targeting", *International Law Studies*, Vol. 96, No. 1, 2020.
74 ICRC Commentary on GC I, above note 66, para. 150.
75 *Ibid.*, para. 151.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

foreseeable violations.[76] This might plausibly be read to enjoin the publication of disinformation when it is foreseeable that it would lead to violations of IHL, regardless of the number of links in the causal chain. But common Article 1 is confined in scope to persons over which a party "exercises authority", likely requiring some form of control or jurisdiction. Thus, a conflict actor may not be responsible under this rule for disinformation produced within its domain that causes third parties beyond its area of control to violate IHL.

In the negative, common Article 1 is also held to provide that the parties are to refrain from "encourag[ing]" violations of the Conventions.[77] The ICRC derives this duty from the International Court of Justice's (ICJ) judgment in the *Nicaragua* case, wherein the United States distributed manuals openly calling on the rebels to commit acts of violence.[78] The facts of the *Nicaragua* case therefore suggest that the encouragement which the ICRC had in mind exhibits actual advocacy of wrongdoing, not mere disinformation.

Incitement

While human rights treaties do not directly forbid State signatories from engaging in incitement to violence, they are obligated to enact legislation outlawing it.[79] Many States have fulfilled this duty, which is closely tied to the right to life,[80] by passing statutes criminalizing incitement.[81]

Incitement may be distinguished from other expressive acts by the application of a six-factor test.[82] In the context of disinformation, the two most important factors are intent and likelihood. The element of intent differentiates actual "advocacy" of discrimination, hostility or violence from the "mere distribution or circulation of material".[83] Disinformation that takes the form of libellous accusations against individuals or groups likely does not reach this level, no matter how provocative, unless it specifically promotes a violent course of action. Even if it does, the publication would not be considered incitement unless there is also a "reasonable probability" that the allegations therein would directly

76 *Ibid.*, para. 164.
77 *Ibid.*, para. 158.
78 ICJ, *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America)*, Judgment, 26 Nov 1984, paras 118–119, 122.
79 International Convention on the Elimination of All Forms of Racial Discrimination, 660 UNTS 195, 21 December 1965, Art. 4; International Covenant on Civil and Political Rights, 999 UNTS 171, 16 December 1966 (ICCPR), Art. 20(2). While these provisions only address incitement to violent speech on the basis of a few protected categories, a preeminent advocacy group has argued that they should be read broadly to include "race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth, or other status", as in other articles of the ICCPR. Article 19, *Prohibiting Incitement to Discrimination, Hostility, or Violence*, London, December 2012, pp. 19–22.
80 Human Rights Committee, General Comment No. 36, "Article 6: Right to Life", UN Doc. CCPR/C/GC/36, para. 59, 3 September 2019.
81 ICRC Customary Law Study (Practice), above note 28, Practice relating to Rule 89.
82 *Report of the United Nations High Commissioner for Human Rights on the Expert Workshops on the Prohibition of Incitement to National, Racial, or Religious Hatred*, UN Doc. A/HRC/22/17/Add.4, 11 January 2013, para. 29.
83 *Ibid.*, para. 29(c).

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

lead to harm, regardless of whether or not such harm ultimately materializes.[84] These requirements make it rather unlikely that disinformation which does not itself commend violence, even if implicitly designed to spark outrage, would be considered incitement under IHRL.

In a small subset of cases, incitement may also be a criminal matter. Specifically, the Rome Statute of the International Criminal Court (ICC) criminalizes "direct" and "public" incitement to commit genocide.[85] Case law from the International Criminal Tribunal for the former Yugoslavia (ICTY) and International Criminal Tribunal for Rwanda (ICTR), whose statutes included similar provisions, reveals that "direct" means an actual invitation, whether explicit or implicit, to commit genocidal acts.[86] It is therefore again dubious that disinformation alone would ever be prosecuted as incitement to genocide.

Accessorial liability

Even if not violating international law themselves, those propagating disinformation likely to elicit a violent response may accrue vicarious or accessorial liability for the criminal acts of others. International law recognizes several different forms of accessorial liability, which are compiled in Article 25(3) of the Rome Statute. These include solicitation and inducement, aiding and abetting, and contributions to group criminality.

Solicitation and inducement

Liability for a criminal offence may extend to the person who "orders, solicits, or induces" its commission.[87] These acts each entail "prompting another person to commit a crime" and are referred to collectively as "instigating".[88] "Ordering" differs from the latter two modalities in that it implies a relationship of authority between the accessory and the principal,[89] which likely would not exist in a case of disinformation. Inducement involves an "exertion of influence by the accessory over the physical perpetrator", whether by rhetoric or action.[90] Solicitation is a lower standard, whereby the accessory merely "asks for" the commission of the crime.[91] In any case, the instigation must have a "direct effect" on the commission or attempted commission of the crime, such that the instigator is its

84 *Ibid.*, para. 29(f).
85 Rome Statute of the International Criminal Court, 2187 UNTS 90, 17 July 1998 (entered into force 1 July 2002) (Rome Statue), Art. 25(3)(e).
86 Michail Vagias, "The Territorial Jurisdiction of the ICC for Core Crimes Committed through the Internet", *Journal of Conflict and Security Law*, Vol. 21, No. 3, 2016.
87 Rome Statute, above note 85, Art. 25(3)(b).
88 ICC, *Prosecutor v. Gbagbo*, Case No. ICC-02/11-01/11, Decision on the Confirmation of Charges, 12 June 2014, para. 243.
89 ICC, *Prosecutor v. Bemba*, Case No. ICC-01/05-01/13, Judgment Pursuant to Article 74 of the Statute, 19 October 2016, para. 77.
90 *Ibid.*, para. 76.
91 *Ibid.*

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

"intellectual author".[92] The instigator must also have been aware that the resultant conduct would occur in the ordinary course of events.[93]

In case law, solicitation and inducement often involve recruitment or instruction to commit a specific act.[94] However, more general speech acts have also been treated at times as indicative of instigation, such as derogatory language toward a minority group,[95] campaign activities employing violent rhetoric,[96] and "public statements which indicate an intention to hold on to power at any cost, including by use of force against civilians".[97] Disinformation that tends to arouse popular animosities towards identifiable targets might be viewed similarly.

Aiding and abetting

A second mode of co-perpetration in international law is "aid[ing], abet[ting] or otherwise assist[ing]" in the commission or attempted commission of a crime.[98] This is associated with a lower degree of culpability than instigation because the principal is determined to commit the crime before the intervention of the accessory.[99] "Aid[ing]" and "otherwise assist[ing]" are given the same meaning: providing "practical or material assistance".[100] "Abet[ting]", by contrast, connotes "moral or psychological assistance", including "encouragement of or even sympathy for the commission of the particular offense".[101] While there is no "minimum threshold", the assistance must have "furthered, advanced or facilitated" the criminal conduct.[102] Finally, according to the Rome Statute, the perpetrator must have acted with "the purpose of facilitating the commission of [the] crime",[103] a stricter *mens rea* requirement than is found elsewhere in the Statute.

Liability for aiding and abetting has mostly been found for defendants who in some way actively take part in the criminal conduct or bolster the principal's capacity to commit the crime. Culpable behaviour may include providing logistical or communications support, knowingly transferring illicit payments, or participating in planning, recruitment or training.[104] Moral support may be expressed through physical presence or by failing to object when informed of the crime.[105] Based on these precedents, a disinformation campaign would aid and

92 *Ibid.*, para. 81.
93 *Ibid.*, para. 82. See also Rome Statute, above note 85, Art. 30(2)(b).
94 ICC, *Bemba*, above note 89, para. 854; ICC, *Gbagbo*, above note 88, para. 246.
95 ICC, *Prosecutor v. Ntaganda*, Case No. ICC-01/04-02/06, Decision Pursuant to Article 61(7)(a) and (b) of the Rome Statute on the Charges of the Prosecutor against Bosco Ntaganda, 9 June 2014, para. 155.
96 ICC, *Gbagbo*, above note 88, paras 117–119, 237.
97 *Ibid.*, paras 111–116, 237.
98 Rome Statute, above note 85, Art. 25(3)(c).
99 ICC, *Bemba*, above note 89, para. 86.
100 *Ibid.*, para. 88.
101 *Ibid.*, para. 89.
102 *Ibid.*, paras 93–94.
103 Rome Statute, above note 85, Art. 25(3)(c).
104 ICC, *Bemba*, above note 89, paras 864–893; ICC, *Prosecutor v. Blé Goudé*, Case No. ICC-02/11-02/11, Decision on the Confirmation of Charges, 11 December 2014, paras 168–171.
105 ICC, *Bemba*, above note 89, paras 867–868.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

abet offences only if it were specifically designed to embolden actors who had already been plotting criminal activity.

Contributions to group criminality

The final form of accessorial liability recognized by the Rome Statute is for acts which "[i]n any other way contribute[ ] to the commission or attempted commission of … a crime by a group of persons acting with common purpose".[106] This is a residual mode of secondary liability intended to capture culpable acts other than ordering, soliciting, inducing or aiding and abetting.[107] It is also directed specifically toward group criminality,[108] which might include the acts of States or organized armed groups involved in disinformation operations. In order for liability to attach, the contribution must be "significant",[109] intentional, and either in furtherance of the criminal purpose or with knowledge of the group's criminal intentions.[110]

Charges against media officials peddling disinformation amid conflict have been brought before the ICC under this theory in at least two cases. In Kenya, charges were confirmed against the head of a radio station accused of "broadcasting false news regarding alleged murder(s) of Kalenjin people in order to inflame the atmosphere".[111] In the Democratic Republic of the Congo, the executive secretary of an armed group conducted an international disinformation campaign through radio communications and press releases with the goals of skirting accountability and winning political and military concessions.[112] This charge was not confirmed for a number of evidentiary reasons.[113] These cases indicate that individuals responsible for disinformation during armed conflict may be held criminally liable if their dispatches significantly contribute to crimes perpetrated by the belligerents and are intended or known to be likely to do so.

Distortion of information concerning vital human needs

Disinformation concerning health care, food and water supplies, or other essential needs represents a second form of civilian harm that may be brought about by disinformation in armed conflict. From a human rights perspective, these

106 Rome Statute, above note 85, Art. 25(3)(d).
107 ICC, *Prosecutor v. Mbarushimana*, Case No. ICC-01/04-01/10, Decision on the Confirmation of Charges, 16 December 2011, para. 278.
108 *Ibid.*
109 *Ibid.*, paras 283–284 (setting forth a set of factors for the assessment of significance).
110 Rome Statute, above note 85, Art. 25(3)(d)(i–ii); ICC, *Prosecutor v. Ruto, Kosgey, and Sang*, Case No. ICC-01/09-01/11, Decision on the Confirmation of Charges Pursuant to Article 61(7)(a) and (b) of the Rome Statute, 23 January 2012, para. 351.
111 ICC, *Ruto*, above note 110, para. 53.
112 ICC, *Mbarushimana*, above note 107, para. 304.
113 *Ibid.*, paras 311–315.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

deliberate misrepresentations may violate the rights to health, food, clothing, shelter or water, or the putative right to humanitarian assistance. By discouraging civilians from availing themselves of institutions providing these necessities, such disinformation may also illegally "render [them] useless" and might furthermore be a type of collective punishment. A separate set of IHL protections would apply if the disinformation campaigns were construed as "military operations". Finally, infodemics that effectively deprive civilians of the information required to make responsible decisions related to their well-being could conceivably constitute a crime against humanity or even the war crime of starvation.

Fundamental guarantees and human rights

Under IHL, civilians in armed conflict are entitled to respect for person and humane treatment,[114] including medical care "to the fullest extent practicable and with the least possible delay".[115] At a minimum, respect for person includes all individual rights, and humane treatment is to be interpreted in its "broadest sense".[116] But the content of these "fundamental guarantees" is mostly supplied by other branches of international law, primarily, IHRL.[117]

Disinformation regarding critical services may run afoul of a number of core human rights provisions. During the COVID-19 pandemic, disinformation concerning medical care and medical providers has proliferated. This impinges upon the rights to life and health, both of which impose obligations upon States to prevent and control the spread of diseases.[118] In order to do so, States must ensure that health information is freely accessible to the public.[119] Any deliberate "misrepresentation of information vital to health protection or treatment" is a violation of the right to health.[120]

Misinformation concerning humanitarian operations during armed conflict, too, is commonplace.[121] These misrepresentations may have obvious bearing on the rights to health and life, along with the rights to food, clothing, housing and water.[122] In addition, some scholars have posited an independent,

114 AP I, Art. 75; AP II, Art. 4; common Art. 3.
115 AP I, Art. 10; AP II, Art. 7.
116 ICRC Commentary on APs, above note 27, para. 4521, 4523.
117 Jakob Kellenberger, "International Humanitarian Law and Other Legal Regimes: Interplay in Situations of Violence", statement to the 27th Annual Round Table on Current Problems of International Humanitarian Law, 6 September 2003, available at: www.icrc.org/en/doc/resources/documents/statement/5rfgaz.htm.
118 Human Rights Committee, above note 80, para. 26; Committee on Economic, Social and Cultural Rights (CESCR), General Comment No. 14, "The Right to the Highest Attainable Standard of Health (Art. 12)", UN Doc. E/C.12/2000/4, 11 August 2000, para. 16.
119 CESCR, above note 118, para. 12(b)(4).
120 *Ibid.*, para. 50.
121 Viviane Lucia Fluck, *Managing Misinformation in a Humanitarian Context*, Internews, 2019, Part 1, pp. 7–9; M. Bunce, above note 9, p. 49.
122 International Covenant on Economic, Social and Cultural Rights, 993 UNTS 3, 16 December 1966 (ICESCR), Art. 11(1); CESCR, General Comment No. 15, "The Right to Water (Arts. 11 and 12 of the Covenant)", UN Doc. E/C.12/2002/11, 20 January 2003, para. 3.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

albeit limited, right to receive humanitarian assistance.[123] Disinformation regarding access to life-saving goods and services would clearly interfere with this right as well.

Access to information pervades the discourse with respect to all of these rights. In addition to its instrumental role in advancing other rights, access to information is itself substantively protected by the International Covenant on Civil and Political Rights and includes the "freedom to seek, receive, and impart information and ideas of all kinds".[124] Information covered by this right includes records retained by public bodies and personal data stored in electronic data files.[125] But it is not facially apparent that this guarantee restrains the State from publishing inaccurate information or that it would require the State to validate information published by private actors.

One specific "fundamental guarantee" found in the Geneva Conventions and their Additional Protocols is the prohibition of collective punishment.[126] Collective punishment is considered to consist of "penalties of any kind" and is to be understood in its "widest sense".[127] This might reasonably include disinformation designed to harm civilians, though only when undertaken with retributive intent.[128]

Protections for civilian objects

Just like attacks against civilians, attacks against civilian objects are forbidden by IHL.[129] In this context, an "attack" means acts leading to "damage or destruction to objects".[130] "Damage" is not restricted to physical harm, but may also connote "impairing the functionality of an object".[131] Following this, it might be argued that certain forms of disinformation have the effect of functionally disabling objects delivering medical or humanitarian assistance in armed conflict by dissuading civilians from making use of them.[132] While most of the Tallinn Manual drafters considered a loss of functionality to occur only if it "requires the replacement of physical components" to restore operability,[133] some States have

123 Maria Giovanna Pietropaolo, "A Human Rights-Based Approach to Humanitarian Assistance", *Journal of International Humanitarian Legal Studies*, Vol. 7, No. 2, 2016; Yoram Dinstein, "The Right to Humanitarian Assistance", *Naval War College Review*, Vol. 54, No. 4, 2000.

124 ICCPR, Art. 19(2).

125 Human Rights Committee, General Comment No. 34, "Article 19: Freedoms of Opinion and Expression", UN Doc. CCPR/C/GC/34, 12 September 2011, para. 18.

126 Geneva Convention (III) relative to the Treatment of Prisoners of War of 12 August 1949, 75 UNTS 135 (entered into force 21 October 1950), Art. 33; AP I, Art. 75(2)(d); AP II, Art. 4(2)(b).

127 ICRC Commentary on APs, above note 27, paras 4535–4536.

128 Special Court for Sierra Leone (SCSL), *Prosecutor v. Brima, Kamara, and Kanu*, Case No. SCSL-2004-16-T, Judgment, 20 June 2007, para. 676; SCSL, *Prosecutor v. Fofana and Kondewa*, Case No. SCSL-04-14-A, Judgment, 28 May 2008, para. 224.

129 AP I, Art. 52; ICRC Customary Law Study, above note 25, Rule 7.

130 Tallinn Manual, above note 12, Rule 30.

131 ICRC, *International Humanitarian Law and the Challenges of Contemporary Armed Conflict*, Geneva, 2015, p. 41.

132 See M. Milanovic and M. N. Schmitt, above note 13, p. 269, equating physical disablement of public health services with "a misinformation campaign that fatally undermines public confidence" in them.

133 Tallinn Manual, above note 12, Rule 30, para. 10.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

taken a contrary position in their assessment of cyber operations.[134] France, for example, has stated that a cyber operation constitutes an attack if the targeted object is no longer able to serve its intended purpose, regardless of physical damage.[135] Disinformation might analogously corrode the utility of civilian objects by altering public perception towards them, albeit through a mechanism less direct than a cyber attack.

In both IACs and NIACs, it is also "prohibited to attack, destroy, remove, or render useless objects indispensable to the survival of the civilian population".[136] This provision, which appears alongside the prohibition against starvation as a method of warfare, was conceived as a response to "scorched earth" tactics.[137] It is therefore primarily applicable to objects vital for sustenance,[138] namely food and water installations.[139] While the ICRC Commentary on the Additional Protocols notes that the assortment of verbs ("attack, destroy, remove, or render useless") is intended to "cover all possibilities",[140] examples enumerated in both the Additional Protocols and their Commentary exclusively concern food and water.[141] In application, too, the scope of "indispensable objects" has been restricted to facilities critical to the food and water supply.[142] Disinformation that takes the form of discouraging civilians from partaking of such provisions, such as by unjustly questioning their safety, might be thought of as "render[ing] useless" the institutions delivering them, even if not qualifying as "attacks" against them.

Distinction and constant care in military operations

Instead of "attacks", some IHL provisions purport to protect civilians during "military operations".[143] Specifically, IAC combatants may direct their military operations "only against military objectives"[144] and must take "constant care" in their conduct "to spare the civilian population".[145] In both IACs and NIACs, civilians "enjoy general protection against the dangers arising from military

134 See, for example, Federal Republic of Germany, above note 54, p. 4.
135 French Ministry of Armed Forces, *Droit international appliqué aux opérations dans le cyberespace*, 19 September 2019, p. 13, available at: www.justsecurity.org/wp-content/uploads/2019/09/droit-internat-appliqu%C3%A9-aux-op%C3%A9rations-cyberespace-france.pdf.
136 AP I, Art. 54; AP II, Art. 14.
137 Knut Dörmann, *Elements of War Crimes under the Rome Statute of the International Criminal Court*, ICRC, Geneva, 2003, pp. 254–255.
138 Article 54 of AP I is limited in scope to acts taken "for the specific purpose of denying [indispensable objects] for their sustenance value". The protection for indispensable objects in Article 14 of AP II is linked textually to the prohibition of starvation by the word "therefore".
139 AP I, Art. 54 ("foodstuffs, crops, livestock, drinking water installations and supplies, and irrigation works"); AP II, Art. 14.
140 ICRC Commentary on APs, above note 27, paras 2011, 4801.
141 AP I, Art. 54; AP II, Art. 14; ICRC Commentary on APs, above note 27, para. 2100.
142 UN Fact-Finding Mission on the Gaza Conflict, *Human Rights in Palestine and Other Occupied Arab Territories*, UN Doc. A/HRC/48, 25 September 2009, paras 1318–1320.
143 AP I, Art. 48 (the ICRC Commentary instructs that "operations" in this article is to be read as "military operations": ICRC Commentary on APs, above note 27, para. 1875); AP II, Art. 13.
144 AP I, Art. 48.
145 *Ibid.*, Art. 57.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

operations".[146] This requires the combatants not only to abstain from attacks against civilians and civilian objects but also to minimize "incidental losses" and to take reasonable precautions.[147] Together, these rules would sharply restrict the use of disinformation in armed conflict to the extent that it may be considered a military operation – though it is not clear that it would be.

While "military operations" are surely broader in ambit than "attacks",[148] they are not defined consistently. In one section, the ICRC Commentary on the Additional Protocols distinguishes military operations from "ideological, political, or religious campaigns" based on the element of violence inherent to the former.[149] It further characterizes military operations as "all movements and acts related to hostilities that are undertaken by armed forces".[150] In another section, however, the Commentary defines military operations as "movements of attack or defence by the armed forces in action".[151]

This erratic guidance leaves considerable ambiguity as to whether wartime disinformation might qualify as a military operation. It is possible that disinformation could be an act related to "hostilities", which is interpreted broadly in IHL.[152] But it is unclear how "movements of attack" would differ from outright "attacks", or when disinformation would be adjudged "violent" rather than "ideological" or "political". Moreover, the "armed forces" may not be the sole entity engaged in disinformation operations. Even when they are involved, can they be said to be "in action" at that time? In short, many questions remain as to whether the IHL protections for civilians during "military operations" would apply to disinformation campaigns.

Criminal liability

Several of the criminal offences that might fit disinformation are premised on the occurrence of an attack. In both IACs and NIACs, the Rome Statute designates as war crimes "[i]ntentionally directing attacks" against civilians, civilian objects, humanitarian assistance personnel and objects, and hospitals.[153] Because war crimes are defined in reference to IHL, the meaning of "attacks" in these instances is the same as that given in AP I – i.e., "acts of violence against the adversary".[154] As discussed above, this would probably exclude disinformation owing to the unlikeliness of the criminal conduct being attributed to the publisher.

146 *Ibid.*, Art. 51(1); AP II, Art. 13(1).
147 ICRC Commentary on APs, above note 27, para. 4770.
148 International Law Association Study Group on the Conduct of Hostilities in the 21st Century (ILA Study Group), "The Conduct of Hostilities and International Humanitarian Law: Challenges of 21st Century Warfare", *International Law Studies*, Vol. 93, 2017.
149 ICRC Commentary on APs, above note 27, para. 1875.
150 *Ibid.*, para. 1875.
151 *Ibid.*, para. 4769.
152 *Ibid.*, para. 2070: "An act of hostility must be understood as any act arising from the conflict which has or can have a substantial detrimental effect on … protected objects."
153 Rome Statute, above note 85, Arts 8(2)(b)(i–iii), 8(2)(b)(ix), 8(2)(e)(i–iv).
154 AP I, Art. 49.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

For an act to qualify as a crime against humanity, it too must be committed in the course of a "widespread or systematic attack".[155] However, the usage of "attack" differs here and "need not constitute a military attack".[156] Instead, an attack in the context of crimes against humanity denotes "any mistreatment of the civilian population"[157] that "caus[es] physical or mental injury",[158] including by "exerting pressure on the population to act in a particular manner".[159] These broader glosses, drawn from the case law of international tribunals, might encompass disinformation that leads to civilian harm, including misrepresentations as to the quality or availability of essential goods and services. The disinformation in question would also have to be published in a "widespread" or "systematic" manner and "pursuant to a State or organizational policy",[160] but these conditions are common features of coordinated media operations.

To the extent that disinformation causes or contributes to a health or humanitarian crisis, it might amount to the crime against humanity of "other inhumane acts … intentionally causing great suffering, or serious injury to body or to mental or physical health".[161] Just as disinformation may interfere with the delivery of food, water or medical care, several international tribunals have found that similar deprivations in custodial settings constitute inhumane acts.[162] And if disinformation in an IAC were believed to have "wilfully imped[ed] relief supplies", it might also be charged as the war crime of "starvation of civilians as a method of warfare",[163] for which an attack is not a prerequisite.

Mental suffering

The final type of civilian harm owing to disinformation in armed conflict is mental suffering. Disinformation at any time can produce panic and distress, but all the more so during warfare. It is possible that these harms could be an unlawful attack against civilians or even a war crime or crime against humanity. It is

155 Rome Statute, above note 85, Art. 7(1).
156 ICC, *Elements of Crimes*, 2011, p. 5.
157 ICTR, *Prosecutor v. Nahimana*, Case No. ICTR-99-52-A, Judgment (Appeals Chamber), 28 November 2007, para. 916, citing ICTY, *Prosecutor v. Kunarac*, Case No. IT-96-23-T, Judgment (Appeals Chamber), 12 June 2002, para. 86.
158 ICTY, *Prosecutor v. Krajisnik*, Case No. IT-00-39-T, Judgment (Trial Chamber), 27 September 2006, para. 706.
159 ICTR, *Prosecutor v. Akayesu*, Case No. ICTR-96-4-T, Judgment (Trial Chamber), 2 September 1998, para. 581.
160 Rome Statute, above note 85, Arts 7(1), 7(2)(a).
161 *Ibid.*, Art. 7(1)(k).
162 Special Panels for Serious Crimes, *Prosecutor v. Leite*, Case No. 04a/2001, Judgment, 7 December 2002, paras 156–162; ICTY, *Prosecutor v. Prlić*, Case No. IT-04-74-T, Judgment (Trial Chamber), 29 May 2013, Vol. 3, paras 1059–1067; Extraordinary Chambers in the Courts of Cambodia, *Prosecutor v. Nuon and Khieu*, Case No. 002/19-09-2007/ECCC/TC, Judgment (Trial Chamber), 7 August 2014, paras 456–458.
163 Rome Statute, above note 85, Art. 8(2)(b)(xxv).

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

indisputable that they violate the right to health, which includes mental health, though only insofar as IHRL jurisdiction applies.

Attacks and spreading terror

As noted above, acts that result in "injury or death" are "attacks" under IHL.[164] Importantly, "injury" may include "severe mental suffering".[165] The meaning of "severe mental suffering" in IHL – including the level and types of harm contemplated and the evidentiary showing required – remains unsettled,[166] but criminal jurisprudence has provided a number of clues. International case law establishes that great suffering and serious mental harm "need not [be] permanent and irremediable", but must go "beyond temporary unhappiness, embarrassment, or humiliation … result[ing] in a grave and long-term disadvantage to a person's ability to lead a normal and constructive life".[167] Mental injuries might be caused by death threats[168] or by otherwise subjecting a person to "intense fear, terror, [or] intimidation",[169] and may be proven by medical records, victim statements or circumstantial evidence.[170] Some experts hold that post-traumatic stress disorder is sufficiently debilitating to be classified as severe mental suffering.[171] It is possible to imagine disinformation triggering similarly detrimental psychological reactions and therefore constituting a wrongful attack against civilians.

The mental harms brought about by disinformation might also evoke the IHL prohibition on "[a]cts or threats of violence the primary purpose of which is to spread terror among the civilian population".[172] The similarity of this provision to the prevailing definition of attack ("acts of violence against the adversary") is not coincidental – it is portrayed by the ICRC as merely an application of the antecedent prohibition on attacking civilians.[173] Indeed, it has in practice been primarily applied to incidents that unquestionably qualify as attacks, such as indiscriminate shelling or bombardment and physical assault or abuse.[174] To characterize disinformation as spreading terror, therefore, it must first be labelled an attack, or perhaps a threat of attack. Even if it is, however, it only conflicts with IHL if spreading terror is its primary purpose.[175] This may

164 Tallinn Manual, above note 12, Rule 30.
165 *Ibid*., para. 8.
166 ILA Study Group, above note 148, pp. 359–360.
167 ICTY, *Prosecutor v. Krstic*, Case No. IT-98-33-T, Judgment (Trial Chamber), 2 August 2001, para. 245.
168 ICJ, *Bosnia and Herzegovina v. Serbia and Montenegro*, Judgment, 26 February 2007, paras 290–291; ICTY, *Prosecutor v. Tolimir*, Case No. IT-05-88/2-A, Judgment, 8 April 2015, para. 206.
169 Quebec Superior Court, *R. v. Munyaneza*, Judgment, 22 May 2009, para. 89.
170 Nena Milaninia, "Understanding Serious Bodily or Mental Harm as an Act of Genocide", *Vanderbilt Journal of Transnational Law*, Vol. 51, No. 5, 2018.
171 ILA Study Group, above note 148, pp. 359–360.
172 AP I, Art. 51; AP II, Art. 13. See also ICRC Customary Law Study, above note 25, Rule 2.
173 ICRC Commentary on APs, above note 27, para. 1940 ("In the second sentence the Conference wished to indicate that the prohibition [on attacks directed against the civilian population] covers acts intended to spread terror"); Tallinn Manual, above note 12, Rule 36, para. 2.
174 ICRC Customary Law Study, above note 25, Rule 2.
175 ICRC Commentary on APs, above note 27, para. 1940.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

not be the case with disinformation, in which the principal objectives may also include punishing opponents, deflating morale or breeding confusion.

Criminal liability

Recalling the causation discussion above, it would likely be inappropriate to charge the perpetrators of a disinformation operation with offences involving killing or bodily harm. However, it may be possible to hold them to account for directly inflicting great suffering or serious mental injury upon civilians. Various forms of genocide, crimes against humanity and war crimes capture this unique form of harm. Because of the unique *mens rea* requirement ("intent to destroy, in whole or in part, a national, ethnical, racial, or religious group"), it is unlikely that disinformation would be prosecuted as the genocidal act of "[c]ausing serious bodily or mental harm to members of the group".[176] However, disinformation might reasonably be charged as the crime against humanity of "other inhumane acts … causing great suffering, or serious injury to body or to mental or physical health",[177] the IAC war crime of "wilfully causing great suffering, or serious injury to body or health",[178] or the NIAC war crime of "cruel treatment",[179] which is the infliction of "severe physical or mental pain or suffering."[180]

Right to health

Health in international law is understood in both its physical and mental dimensions. The International Covenant on Economic, Social, and Cultural Rights consecrates the "right of everyone to the enjoyment of the highest attainable standard of physical and mental health".[181] Disinformation that damages mental health is a betrayal of this guarantee, though it must be considered which of the conflict actors bears this obligation with respect to the victim population.

Conclusion

Disinformation in armed conflict poses an ever-growing threat to civilians, who may be exposed to retaliatory violence, denied accurate information critical to their health or sustenance, or caused mental anguish by it. These harms have been overlooked in international law, in part due to the historical characterization of "misinformation" as a permissible ruse of war. This article has argued for the

176 Rome Statute, above note 85, Art. 6(b).
177 *Ibid.*, Art. 7(1)(k).
178 *Ibid.*, Art. 8(2)(a)(iii).
179 *Ibid.*, Art. 8(2)(c)(i).
180 ICC, above note 156, Art. 8(2)(c)(i)–(3).
181 ICESCR, Art. 12.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press

abandonment of that approach in light of its inconsistency with the contemporary nature of disinformation.

Even if this proposition is accepted, attempts to apply other rules of international law to disinformation encounter numerous grey areas and juridical impediments. First, while disinformation can be produced from afar, the applicability of IHRL and common Article 1 are both confined, respectively, to areas under the "control" or "authority" of conflict parties. Second, because the harms wrought by disinformation are often caused by indirect means, it is not likely to be viewed as an attack or other act of violence[182] under IHL. Nor would it be a form of incitement, unless it also advocated violence or hostility. Third, while disinformation can inflict direct harms to mental health, these injuries are difficult to assess and are frequently neglected in IHL.[183] Fourth, disinformation may also frustrate the purpose of civilian objects, but not necessarily in a way that would amount to an attack. Finally, civilians and civilian objects directly participating in disinformation are unlikely to be deterred by the temporary stripping of their IHL protections, though they may be exposed to criminal liability.

The inadequacy of the legal response to disinformation in armed conflict, coupled with the gravity of the harms it may occasion, underscores the need for the articulation of clearer standards. In June 2021, the international community took an important first step with the release of the Oxford Statement on International Law Protections in Cyberspace,[184] which sets forth ten general principles, touching on sovereignty, incitement, human rights, criminal law, common Article 1, and general rules of IHL and international criminal law. This article has presented further details as to the specific application of these principles to disinformation during armed conflict, providing a theoretical foundation for further efforts in the field.

182 See AP I, Art. 75(2); AP II, Art. 4(2); common Art. 3(1).

183 Eliav Lieblich, "Beyond Life and Limb: Exploring Incidental Mental Harm under International Humanitarian Law", in Derek Jinks, Jackson Nyamuya Maogoto and Solon Solomon (eds), *Applying International Humanitarian Law in Judicial and Quasi-Judicial Bodies: International and National Aspects*, Asser Press, The Hague, 2014, p. 187.

184 D. Akande *et al.*, above note 19.

https://doi.org/10.1017/S1816383121000473 Published online by Cambridge University Press