

Perpetrators 2.0: Rethinking Violence and Violent Actors in the Digital Era?

Timothy Williams

Abstract: In recent years, digital technologies have become increasingly important in the perpetration of violence. Online spaces facilitate not only violence in solely digital spaces (e.g. hate speech, cyber-bullying, conspiracy theories) but are also constitutive of hybrid violence, when physical violence and digital violence intersect. Here, online spaces contribute to the recruitment of potential violent actors, incitement and motivation of participants, polarisation of discourses, and concrete logistics of implementing violence, among other things. As yet, however, there has been no evaluation of how the concepts of ‘perpetrator’ and ‘perpetration’ that were developed for physical violence can – or cannot – be mapped onto these new forms of digital and hybrid violence. It is important to consider the question of ‘digital perpetrators’ though, as this is a useful starting point for thinking through issues such as culpability and implication in digital space, as well as thinking about how to prevent such action in digital spaces and who should be held responsible for this. This article pursues several avenues of inquiry that are necessary to understand whether and how the concept of perpetrator can be useful in digital spaces. The article develops specific criteria for defining “Perpetrators” and “Perpetration” for physical violence (nature of harm; degree of participation) and reflects on how these can be applied to digital and hybrid violence, while also demarcating and specifying the idea of perpetrator vis-à-vis other types of actors in online spaces (bystanders; upstanders). The article is not an attempt to finally settle these issues, but sketches first ideas on how to think about perpetrators and perpetration in online spaces and identifies key issues for further discussion.

Keywords: perpetrators, digital violence, social media, hybrid violence, online spaces

Introduction

Violence research and public discussion are paying increased attention to violence in online spaces and how it contributes to physical violence in riots, genocides, and terrorist attacks around the world. Participants to violence have been mobi-

I would like to thank Kirsty Campbell and Max Wegener, as well as the organizers and participants of the conference ‘Reflections on the Boundaries of Perpetrator Studies: Looking at Violence, Politics, and Harm,’ held in Brussels in June 2024, and the guest editors and reviewers for their helpful feedback. Furthermore, I thank Aaron Emde and Tim Martick for their research support.

Journal of Perpetrator Research 8.2 (2026), 77–102

DOI: 10.21039/jpr.8.2.183 © 2026 by the Author



This article is published under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

lised through social media such as Facebook to participate in genocide against the Rohingya in Myanmar and the Yazidi in Iraq, spreading inflammatory legitimations of the violence widely. The surveillance of social media has been key to the violence against Uyghur in China. We have seen how Facebook, WhatsApp, and X (formerly Twitter) were instrumental in coordinating and motivating participants in riots from Delhi in India, to Gauteng in South Africa, and the Capitol Hill in Washington D.C., USA.

And yet, while the term perpetrator is easily and routinely applied to participants of these instances of physical violence, there is a noticeable avoidance of the term perpetrator in discussing online dynamics. For example, if one searches for the terms ‘perpetrator’ and ‘terrorism’ on the academic database Google Scholar one finds around 90,000 results. For ‘perpetrator’ and ‘war’ this number is even around 192,000 (on JSTOR the search resulted in 6994 and 28,408 hits, respectively). However, when searching for ‘perpetrator’ and ‘digital violence’, one finds only 1,080 hits on Google Scholar, and only twenty on JSTOR.¹ Is the implication of this relative dearth of hits that digital violence ‘just occurs’ and its agents are irrelevant, or that it is not deemed problematic enough to justify the use of a label with such a negative connotation as ‘perpetrator’? In this article, I would like to suggest that a debate on the nature of perpetration in digital spaces is important for research to be able to subsequently address issues such as culpability for violent action undertaken in digital spaces. It is likewise important when it comes to prevention attempts. This article seeks to address these issues and asks how we can conceptualise perpetrators and their acts of perpetration in the context of digital violence.

In this article, I pinpoint two defining criteria of perpetrators from previous research, the nature of harm and degree of participation, and apply these to online spaces. I demonstrate that it is considerably more difficult to define the nature of harm and the degree of participation for digital violence than it is in offline spaces, resulting in either very narrow or very broad understandings of perpetration. I also show how Michael Rothberg’s idea of the ‘implicated subject’ can be useful in interrogating some subject positions beyond a narrow approach to perpetration.²

1 Results from a simple search of the key terms in inverted commas (e.g. one search command was: “Perpetrator” + “digital violence”) on 25 February 2026 on <<https://scholar.google.com/>> and <<https://www.jstor.org/>>.

2 Michael Rothberg, *The Implicated Subject: Beyond Victims and Perpetrators* (Stanford University Press, 2019).

To make these arguments I first locate this article within the broader state of the art on violence in social media and online spaces. Subsequently, the article delineates previous approaches to how perpetrators have been conceptualised, outlining various typologies, definitions, and approaches to identify two core factors in understanding who is a perpetrator: the nature of harm and degree of participation. These are then applied to digital spaces to think through the concept of digital perpetrators, also differentiating them from understandings of bystanders and upstanders in online spaces.

Violence in the Digital Era

While there are myriad discussions on violence in digital spaces, few works define it cogently. For this article, I draw on a broad definition of digital violence as ‘the use of words, pictures, computer code or instructions within a digital realm to cause psychological or physical harm’ to other individuals or groups.³ While cybercrimes, hacking, and other forms of violent action in digital spaces are certainly important particularly for understanding war and organised crime, most research on political violence in the digital era focuses on social media.

Social media are prominent in political discourse and enjoy common usage throughout the population. They rest on bottom-up interactions generated between users that necessitate little or no specialised knowledge (albeit users’ reach may be enhanced if they produce more professional content), allowing broad swathes of the population to participate in discussions, social relations, as well as political action including violence. Social media have low barriers for entry,⁴ platforms enable discussion and engagement in relatively open and loose formations, and they foster translocal and even globalised networks of communication that impact behaviors, attitudes, and identities.⁵ While digital spaces are in principle open to all, some more vulnerable groups do not have access to the required technological infrastructure and power relations in digital spaces often mirror intersectional discrimination and marginal-

3 Max Kilger, ‘Interventions, Policies, and Future Research Directions in Cybercrime’, in *The Wiley Handbook on the Psychology of Violence*, ed. by C. A. Cuevas and C. M. Rennison (Wiley, 2016), pp. 604–22 (p. 606).

4 Ekaterina Zhuravskaya, Maria Petrova, and Ruben Enikolopov, ‘Political Effects of the Internet and Social Media’, *Annual Review of Economics*, 12.1 (2020), pp. 415–38 (p. 416).

5 Marco Bastos, *Spatializing Social Media: Social Networks Online and Offline* (Routledge, 2022).

isation patterns found in offline spaces.⁶ Most users are concentrated on a few large platforms such as X, Facebook, Instagram, TikTok, Snapchat, and YouTube;⁷ smaller platforms complement these and gather specific groups of people, for example image boards such as 8kun or 4chan, or specific threads on Reddit have been prominent in facilitating discussions and radicalisation processes within far-right communities.⁸

Despite initial optimism about the positive impact and democratic potential of the internet, and social media's connectivity, digital spaces are now more broadly recognised as contributing to social polarisation and inequality, disinformation, and hate speech.⁹ In digital spaces there is increased competition for status and legitimacy between groups, contributing to polarisation and the denigration of other groups.¹⁰ These interactions between groups are structured significantly by the social media platforms' architectural design and algorithm structure,¹¹ creating 'filter bubble' and 'echo chamber' effects when individuals are displayed systematically different information and news, including misinformation and fake news, which can reinforce pre-existing political attitudes and foster polarisation.¹² Also, people who hate their political opponents most are also most likely to share fake news in an attempt to denigrate their opponents,¹³ amplifying such voices. Furthermore, as social media companies prioritise increasing user engagement duration to grow advertising revenue, conflictual content that engages users is profitable.

6 Jen Schradie, 'The Digital Activism Gap: How Class and Costs Shape Online Collective Action', *Social Problems*, 65.1 (2018), pp. 51-74, art. 1.

7 Meltwater, *Digital 2024: Global Overview Report* (Meltwater, 2024).

8 Maura Conway, 'Determining the Role of the Internet in Violent Extremism and Terrorism', *Studies in Conflict & Terrorism*, 40.1 (2017), pp. 77-98, art. 1 (p. 85).

9 Christopher A. Bail and others, 'Exposure to Opposing Views on Social Media Can Increase Political Polarization', *Proceedings of the National Academy of Sciences*, 115.37 (2018), pp. 9216-21; James A. Piazza, 'Fake News: The Effects of Social Media Disinformation on Domestic Terrorism', *Dynamics of Asymmetric Conflict*, published online 2021; Cass R. Sunstein, *#Republic: Divided Democracy in the Age of Social Media* (Princeton University Press, 2017).

10 Emily Harmer and Karen Lumsden, 'Online Othering: An Introduction', in *Online Othering*, ed. by Karen Lumsden and Emily Harmer (Springer International Publishing, 2019), pp. 1-33; Bastos, p. 8.

11 Christian Fuchs, *Social Media: A Critical Introduction*, 3rd edn (SAGE, 2021); Anita R. Gohdes, 'Studying the Internet and Violent Conflict', *Conflict Management and Peace Science*, 35.1 (2018), pp. 89-106.

12 Bastos; Piazza.

13 Mathias Osmundsen and others, 'Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter', *American Political Science Review*, 115.3 (2021), pp. 999-1015.

How do social media platforms impact the likelihood of digital violence? Here we see that violence can occur solely in digital spaces, particularly in the form of discriminatory discourses, cyber-bullying, trolling, and hate speech, which can map intersectional discrimination patterns onto digital spaces, but also develop new forms.¹⁴ Furthermore, discourses in social media can reinforce or change social identities, influence perceptions of other groups and legitimise radicalising attributions and discourses.

While social media constitute connections in digital spaces, these can, of course, overlap and intersect with offline spaces: dynamics in social media are impacted by, and in turn influence, political action and physical violence in offline spaces, and also impact the development of social identities.¹⁵ These 'hybrid' spaces demonstrate how experiences and actions no longer occur in purely offline and online contexts but are situated at the intersection of these spaces. Digital spaces have thus become as important in shaping people's understanding of social and political occurrences as offline connections. This makes people more reliant on digital spaces, particularly social media, for information and attitudinal guidance.

As such, political action in offline spaces is not replaced by online activism, but political actions in both spaces mutually constitute each other and interact.¹⁶ Social media have thus shifted political action away from traditional organisational structures that focused on *collective action* to allowing people to engage individually in *connective action* in which they co-produce content through personal

¹⁴ Harmer and Lumsden; Majid KhosraviNik and Eleonora Esposito, 'Online Hate, Digital Discourse and Critique: Exploring Digitally-Mediated Discursive Practices of Gender-Based Hostility', *Lodz Papers in Pragmatics*, 14.1 (2018), pp. 45–68; Karsten Müller and Carlo Schwarz, 'Fanning the Flames of Hate: Social Media and Hate Crime', *Journal of the European Economic Association*, 19.4 (2021), pp. 2131–67; Jillian Peterson and James Densley, 'Cyber Violence: What Do We Know and Where Do We Go from Here?', *Aggression and Violent Behaviour*, 34 (2017), pp. 193–200; Piazza; Shawn Walker, Dan Mercea, and Marco Bastos, 'The Disinformation Landscape and the Lockdown of Social Platforms', *Information, Communication & Society*, 22.11 (2019), pp. 1531–43, art. 11.

¹⁵ Bastos; Gohdes; Harmer and Lumsden; Peterson and Densley.

¹⁶ John D. Gallacher, Marc W. Heerdink, and Miles Hewstone, 'Online Engagement Between Opposing Political Protest Groups via Social Media Is Linked to Physical Violence of Offline Encounters', *Social Media + Society*, 7.1 (2021), pp. 1–16; Hedy Greijdanus and others, 'The Psychology of Online Activism and Social Movements: Relations between Online and Offline Collective Action', *Current Opinion in Psychology*, 35 (2020), pp. 49–54; Zhuravskaya, Petrova, and Enikolopov.

expression.¹⁷ This individualising tendency of individual engagement can nonetheless create collective identities and foster offline action in which people are mobilised through their digital connections for physical participation in protests, rallies, or riots.¹⁸ The internet and social media can support offline collective action through mobilisation, information dissemination, and discursive framing of topics. Social media discourses can thus also mobilise and motivate individuals, allowing coordination, legitimising violence, and providing a space for representing acts of physical violence that have been committed.¹⁹ Furthermore, how groups interact online can also determine whether and to what degree violence ensues offline.²⁰ For example, individuals can be radicalised by incendiary content on social media irrespective of the factual accuracy and veracity of the information,²¹ resulting in a higher propensity to violence,²² and conflict actors have learned to use this potential strategically.²³

Defining Perpetrators of Physical Violence: Nature of Harm and Degree of Participation

Despite the burgeoning of perpetrator studies there are only a few scholars who consider conceptually what constitutes a perpetrator. I do not mean the rich and nuanced literature that demonstrates who and how

17 W. Lance Bennett and Alexandra Segerberg, 'The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics', *Information, Communication & Society*, 15.5 (2012), pp. 739–68 (p. 752).

18 Edward W. N. Bernroider, G. Harindranath, and Sherif Kamel, 'From Connective Actions in Social Movements to Offline Collective Actions: An Individual Level Perspective', *Information Technology & People*, 35.8 (2022), pp. 205–30; Greijdanus and others; Øystein Sæbø, Tommaso Federici, and Alessio M. Braccini, 'Combining Social Media Affordances for Organising Collective Action', *Information Systems Journal*, 30.4 (2020), pp. 699–732.

19 Pablo Barberá and others, 'Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?', *Psychological Science*, 26.10 (2015), pp. 1531–42, art. 10; Marlon Moojiman and others, 'Moralization in Social Networks and the Emergence of Violence during Protests', *Nature Human Behaviour*, 2.6 (2018), pp. 389–96, art. 6; Müller and Schwarz, 'Fanning the Flames of Hate'; Mattias Wahlström, Anton Törnberg, and Hans Ekbrand, 'Dynamics of Violent and Dehumanizing Rhetoric in Far-Right Social Media', *New Media & Society*, 23.11 (2021), pp. 3290–3311.

20 Gallacher, Heerdink, and Hewstone, p. 9.

21 Piazza, p. 4.

22 Lieven Pauwels and Nele Schils, 'Differential Online Exposure to Extremist Content and Political Violence: Testing the Relative Strength of Social Learning and Competing Perspectives', *Terrorism and Political Violence*, 28.1 (2016), pp. 1–29.

23 Gohdes, p. 90.

people or regimes are represented as or attributed the label of perpetrators in public discourse, media depictions, or cultural works, but academic definitions of who we are talking about when we analyse perpetrators. I will outline various approaches to conceptualising perpetrators through typologies, definitions, and action-based understandings.

First, most authors develop typological approaches. For example, legal scholar Kai Ambos distinguishes perpetrators according to their position as state or non-state actors, as well as whether they are low-, middle- or high-level actors.²⁴ Scholar of international law Mark Drumbl differentiates varying degrees of blameworthiness or culpability ranging from conflict entrepreneurs, further leaders, killers, those who assist, those who benefit, to bystanders.²⁵ Others such as sociologist Michael Mann or criminologist Alette Smeulers lay a stronger focus on the motivations that perpetrators have for their participation, highlighting motivations such as careerism, fear, authority, opportunism, camaraderie, or fanaticism and many others.²⁶ Though motivations are unquestionably important to understand,²⁷ it may be less helpful as the foundation for a typology of perpetrators, as it distracts from the possibility of individuals having multiple motivations or that their motivations can change over time. A different approach is to typologise the social role an individual fulfils in the violence; political scientist Lee Ann Fujii, for example, differentiated between leaders, collaborators, joiners, survivors, rescuers, evaders, witnesses, and resisters to distinguish these social roles.²⁸

Second, there are social scientific attempts to define what constitutes a perpetrator. For example, political scientist Omar McDoom defines a perpetrator as 'any individual who joined an attack group that killed at least one individual. The perpetrator then need not have struck a physical blow himself.'²⁹ Political scientist Scott Straus defines perpetrators 'as those who had a hand, directly or indirectly, in the physical destruction

24 Kai Ambos, 'Criminologically Explained Reality of Genocide, Structure of the Offence and the "Intent to Destroy" Requirement', in *Collective Violence and International Criminal Justice*, ed. by Alette Smeulers (Intersentia, 2010), pp. 153–73.

25 Mark Drumbl, *Atrocity, Punishment, and International Law* (Cambridge University Press, 2007).

26 Michael Mann, *The Dark Side of Democracy* (Cambridge University Press, 2005), pp. 27–29; Alette Smeulers, *Perpetrators of Mass Atrocities: Terribly and Terrifyingly Normal?* (Routledge, 2024).

27 Timothy Williams, *The Complexity of Evil: Perpetration and Genocide* (Rutgers University Press, 2021).

28 Lee Ann Fujii, *Killing Neighbours: Webs of Violence in Rwanda* (Cornell University Press, 2009); see also Eric Markusen, 'Mechanisms of Genocide', in *Will Genocide Ever End?*, ed. by Carol Rittner, John K. Roth, and James M. Smith (Paragon House, 2002), pp. 83–89.

29 Omar McDoom, *The Path to Genocide in Rwanda. Security, Opportunity, and Authority in an Ethnocratic State* (Cambridge University Press, 2020), pp. 37–38.

of individuals. Perpetrators take part in violence against non-combatants.³⁰ These definitions view killing as the key action of perpetrators, although they both emphasise the importance of not only those who directly killed but also those more broadly involved.

And, third, there has been a move in recent years to shift our perspective from individuals as perpetrators to a more nuanced understanding of perpetration as a form of action.³¹ This approach does not reify perpetrators as the sole identity of the individual in question but highlights the contingent nature of their perpetrations, focusing on each individual action as the unit of analysis that needs explaining, accounting for multiple motivations and changes over time. Further, it allows for individuals to be understood not exclusively as perpetrators but allows an understanding of them as complex political actors who may perpetrate but can at other times also engage in rescuing, be bystanders or even experience victimisation themselves.³² In this vein, in previous work I have suggested a typology of action that differentiates the contribution that an action makes to the violence in question, as well as the proximity of the action to the violence, establishing fourteen categories that span perpetration, bystanding, and rescuing.³³

From these myriad approaches, I would suggest that we can distill two main elements in how these authors have conceptualised perpetrators that may be helpful for this article's undertaking: the nature of harm and degree of participation. First, the harm that is in focus for most of the discussed definitions and typologies suggests that the perpetrators are involved in the killing of an individual. This is possibly a product of perpetrator studies' predominant focus on events

30 Scott Straus, 'Studying Perpetrators: A Reflection', *Journal of Perpetrator Research*, 1.1 (2017), pp. 28–38 (p. 36).

31 Christian Gudehus, 'Violence as Action', in *Perpetrators and Perpetration of Mass Violence: Dynamics, Motivations and Concepts*, ed. by Timothy Williams and Susanne Buckley-Zistel (Routledge, 2018), pp. 36–57; Christian Gudehus, 'Action Action Action', *Journal of Perpetrator Research*, 3.1 (2020), pp. 188–95; Timothy Williams, 'Thinking beyond Perpetrators, Bystanders, Heroes: A Typology of Action in Genocide', in *Perpetrators and Perpetration of Mass Violence: Dynamics, Motivations and Concepts*, ed. by Timothy Williams and Susanne Buckley-Zistel (Routledge, 2018), pp. 17–35.

32 Bradley Campbell, 'Contradictory Behavior During Genocides', *Sociological Forum*, 25.2 (2010), pp. 296–314; Lee Ann Fujii, 'Rescuers and Killer-Rescuers during the Rwanda Genocide', in *Resisting Genocide: The Multiple Forms of Rescue*, ed. by Jacques Sémelin, Claire Andrieu, and Sarah Gensburger (Columbia University Press, 2011), pp. 145–57; Erin Jessee, *Negotiating Genocide in Rwanda* (Palgrave Macmillan, 2017); Aliza Luft, 'Toward a Dynamic Theory of Action at the Micro Level of Genocide: Killing, Desistance, and Saving in 1994 Rwanda', *Sociological Theory*, 33.2 (2015), pp. 148–72.

33 Williams, 'Thinking beyond Perpetrators, Bystanders, Heroes'.

of genocide and mass killing, but, of course, for other contexts this harm can also be more broadly understood to encompass other crimes. However, this ignores the many forms of non-physical violence, such as discursive, cultural, or epistemic violence that also create significant harms but are not covered by these approaches. Second, the degree of participation is important across many of the discussed approaches, attending to the proximity to violence or distinguishing between direct and indirect participation. Straus elaborates on this distinction: 'If they are direct, they kill; they maim; they torture; they incite violence; they order violence; they distribute weaponry. If they are indirect, they contribute to an institution or organisation that itself participates in violence; they make meals for people that go out and kill; they reveal the location of would-be targets; they steal or take advantage of victims.'³⁴ This suggests the importance in any conceptualisation of including those who do not cause the harm finally, but also those who prepare the way, legitimise, or abet and aid in the implementation of the harm. Here, of course, genocidal leaders such as Adolf Hitler or Radovan Karadžić, and desk perpetrators such as Adolf Eichmann, are important to consider.³⁵

To clarify the concept of perpetrator, it is useful to distinguish it from the other roles in violent action mentioned above: bystanders and rescuers. Bystanders can be understood as 'individuals in a given territory, [...] who do not actively participate in violence, but who share the same politically salient identity as direct perpetrators (ethnicity, 'race,' class, etc.), or whom the perpetrators consider political allies. Bystanders are neither members of the victim group, nor associated with victims.'³⁶ Insufficiently complex approaches suggest that bystanders are uninvolved witnesses with 'no blood on their hands' and are thus morally exculpated,³⁷ but considering bystanding as an active decision focuses attention on the manifest consequences it also entails.³⁸ As Mary Fulbrook delineates for German society under Nazi

³⁴ Straus, p. 36.

³⁵ Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* (Penguin Books, 1994); Leora Bilsky and Rachel Klagsbrun, 'The Desk Perpetrator, the Expert Witness, and the Role of Law: The Trial of Arthur Greiser', *Journal of Genocide Research*, 27.1 (2025), pp. 1-22.

³⁶ Ernesto Verdeja, 'Moral Bystanders and Mass Violence', in *New Directions in Genocide Research*, ed. by Adam Jones (Routledge, 2012), pp. 153-68 (p. 154).

³⁷ Giorgia Donà, "'Situated Bystanderism" During and After the Rwandan Genocide', *Journal of Genocide Research*, 20.1 (2017), pp. 1-19.

³⁸ Mary Fulbrook, *Bystander Society: Conformity and Complicity in Nazi Germany and the Holocaust* (Oxford University Press, 2023), p. 3; Williams, 'Thinking beyond Perpetrators, Bystanders, Heroes'.

rule, to understand bystanders we need to go beyond thinking solely along lines of passivity, as the conformity that being a bystander entails can fade into complicity in laying the foundations for genocide when acting in a context of state-supported violence.³⁹ While bystanders may not have been involved in any of the actions that implemented the violence that defines perpetration, bystanders' presence alone shapes the situations they are in and can have psychological effects on the enactment of violence,⁴⁰ for example through suggesting their support for the violence if they raise no objections, thus legitimising the violence.⁴¹ Furthermore, the so-called 'bystander effect' suggests that the more bystanders present, the less likely an individual is to intervene or rescue.⁴² Adding nuance to these discussions on bystanders, other concepts have also emerged that suggest varying degrees of action vs. passivity, complicity vs. innocence for perpetrated violence, such as beneficiaries,⁴³ collaborators,⁴⁴ or observers.⁴⁵ Comparative literature scholar Michael Rothberg's proposal of the category of implicated subjects has garnered particular attention across different fields. He defines implicated subjects as those who 'occupy positions aligned with power and privilege without being themselves direct agents of harm; they contribute to, inhabit, inherit, or benefit from regimes of domination but do not originate or control such regimes.'⁴⁶ This renders actors not culpable in the sense of being a perpetrator, but implicated enough to go beyond ideas of passive bystanding.

Rescuers, by contrast, are individuals who save potential victims from the harm that perpetrators would have committed. For example, rescuers may hide those being targeted by genocide, provide them with

39 Fulbrook.

40 Fulbrook; Verdeja; Williams, 'Thinking beyond Perpetrators, Bystanders, Heroes'.

41 Ervin Staub, 'The Psychology of Bystanders, Perpetrators, and Heroic Helpers', in *Understanding Genocide: The Social Psychology of the Holocaust*, ed. by Leonard S. Newman and Ralph Erber (Oxford University Press, 2002), pp. 11–42 (p. 24).

42 Staub, p. 27; for a critical discussion see Mark Levine, Richard Philpot, and Anastasiia G. Kovalenko, 'Rethinking the Bystander Effect in Violence Reduction Training Programs', *Social Issues and Policy Review*, 14.1 (2020), pp. 273–96; Peter Fischer and others, 'The Bystander Effect: A Meta-Analytic Review on Bystander Intervention in Dangerous and Non-Dangerous Emergencies', *Psychological Bulletin (US)*, 137.4 (2011), pp. 517–37.

43 Bruce Robbins, *The Beneficiary* (Duke University Press, 2017).

44 P. E. Digeser, 'Collaboration and Its Political Functions', *American Political Science Review*, 116.1 (2022), pp. 200–12.

45 Haley S. Anderson, 'Culpability in Atrocity and the Role of Complicit Observer', *Temple International & Comparative Law*, 37.2 (2023), pp. 11–22.

46 Rothberg, p. 1.

resources in hiding, or aid them in safe passage in trying to flee the violence.⁴⁷ Sociologists Hollie Nyseth Nzitira and Nicole Fox suggest understanding rescue as a form of collective action in which ‘rescue efforts vary both in terms of the actions employed and their level of sustainment over time, as well as the level of coordination between actors.’⁴⁸ Furthermore, we can differentiate between active rescuing, where people save others through concrete action, and passive rescuing, in which people save others through acts of omission, ultimately locating this closer to the category of bystanders.⁴⁹ In other contexts, rescuers have also been labeled ‘upstanders,’ emphasising the active contrast between those who ‘stand up’ and intervene and those who ‘stand by’ passively allowing the violence.⁵⁰ As the term upstanders is more open also to smaller acts than that of rescue, such as speaking out on behalf of someone, this shall be the term that underlies our later thinking on actors in digital violence.

Finally, it should be emphasised that while these categories are defined separately and in relation to each other, empirically, individuals can be attributed to multiple categories at once. This is highlighted by oral historian Erin Jessee’s notion of complex political actors, or sociologist Aliza Luft’s behavioural boundary crossing, perpetrating violence in one moment but rescuing others at other times.⁵¹ Moreover, from an action-oriented understanding of violence as action, all perpetrators are in some situations also bystanders.⁵²

Conceptualising Perpetrators of Digital Violence

Let us now turn to what this can mean for considering who is a perpetrator of violence in digital spaces. The next two sections will interro-

47 Hollie Nyseth Brehm and others, ‘In the Aftermath: The Post-Conflict Social and Economic Consequences of Rescue During Genocide’, *Journal of Genocide Research*, 25.1 (2023), pp. 24–45 (p. 30).

48 Nicole Fox and Hollie Nyseth Brehm, ‘“I Decided to Save Them”: Factors That Shaped Participation in Rescue Efforts during Genocide in Rwanda’, *Social Forces*, 96.4 (2018), pp. 1625–48 (p. 1628).

49 Timothy Williams, ‘Agency, Responsibility, and Culpability: The Complexity of Roles and Self-Representations of Perpetrators’, *Journal of Perpetrator Research*, 2.1 (2018), pp. 39–64 (p. 53).

50 Zachary D. Kaufman, ‘Protectors of Predators or Prey: Bystanders and Upstanders amid Sexual Crimes’, *Southern California Law Review*, 92.6 (2019), pp. 1317–406.

51 Jessee; Luft.

52 Williams, ‘Thinking beyond Perpetrators, Bystanders, Heroes’.

gate the categories of harm and participation further for online spaces, before turning to how this can (or cannot) productively differentiate perpetrators from other categories of online actors.

HARM

The first category through which perpetrators have previously been defined is harm. For physical violence there was a decided focus on killing people. However, building on the above discussions of digital violence, it would be important to discuss harm not only referring to the killing of an individual, but by engaging with both psychological and physical harms. Importantly, this clearly allows us to delineate its meaning for both digital violence that remains in digital spaces, as well as hybrid violence that sees digital dynamics used in the preparation and implementation of physical violence.

Differentiating between hybrid and purely digital violence, we can define the harm experienced in hybrid violence primarily through the harm of the implemented physical violence, to bring this in line with the above perpetrator definitions. Depending on the type of violence examined, we can apply definitions of perpetration from physical violence, e.g. with killing and injuring individuals of the target group as central to the definition. Here, digital spaces are primarily used to recruit, motivate, and incite individuals to participate in the offline violence, to legitimise and justify the violence, to organise logistics and funding, as well as digitally document the physical violence for broader aims of propaganda. Of course, in hybrid violence, this would – beyond the social media in focus in this article – also include the use of technical infrastructures for cyber-attacks or control of drones in the pursuit of realising said harms.⁵³ These forms of hybrid violence can thus also be seen as indirect forms of perpetration for the physical violence, as they support its commission without actively and directly participating. For example, the Islamic State radicalised and recruited individuals from over 90 countries through social media,⁵⁴ and rioters at the

⁵³ Peterson and Densley.

⁵⁴ Imran Awan, 'Cyber-Extremism: Isis and the Power of Social Media', *Society*, 54.2 (2017), pp. 138–49; Donatella Della Ratta, *Shooting a Revolution: Visual Media and Warfare in Syria* (Pluto Press, 2018); Scott Gates and Sukanya Podder, 'Social Media, Recruitment, Allegiance and the Islamic State', *Perspectives on Terrorism*, 9.4 (2015), pp. 107–16; Anita Perešin, 'Fatal Attraction: Western Muslimas and ISIS', *Perspectives on Terrorism*, 9.3 (2015), pp. 21–38.

Capitol Hill on 6 January 2021 were motivated by online discussions and networks were forged in social media.⁵⁵

In the context of purely digital violence, the harms are necessarily different, as they do not cause death or physical injury to victims, although there can be very real consequences in the physical world through psychological harms. One attempt to distinguish forms of online harm comes from the World Economic Forum (WEF) which lists six types of online harm: threats to personal and community safety, harm to health and well-being, hate and discrimination, violation of dignity, invasion of privacy, and deception and manipulation.⁵⁶ Clearly, these harms go well beyond the threshold for harm of offline perpetration discussed previously, but are useful in thinking through what harms are relevant in digital spaces. As such, the focus here is necessarily more on psychological harms created through targeting of identities, emotional abuse, and hostile discourse. For example, hate speech attacks groups due to identity categories such as race, religion, gender, or sexual orientation, and is virulent in many social media platforms,⁵⁷ causing potential harm to any member of these groups that see the content, even if it has not been directed at them specifically. Other more targeted forms of online violence include cyber-bullying against specific individuals, trolling people with posts or comments designed to 'bait' people by deliberately provoking an argument or emotional reaction, and doxing: publishing private information about someone without their consent.⁵⁸

55 Luke Munn, 'More than a Mob: Parler as Preparatory Media for the U.S. Capitol Storming', *First Monday*, published online 7 February 2021; Lynnette Hui Xian Ng, Iain J. Cruickshank, and Kathleen M. Carley, 'Coordinating Narratives Framework for Cross-Platform Analysis in the 2021 US Capitol Riots', *Computational and Mathematical Organization Theory*, 29.3 (2023), pp. 470–86, art. 3; Wei Zhong and others, 'Proud Boys on Telegram', *Journal of Quantitative Description: Digital Media*, 4 (2024), pp. 1–47.

56 World Economic Forum, *Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms: Insight Report* (2023).

57 Michał Bilewicz and Wiktor Soral, 'Hate Speech Epidemic: The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization', *Advances in Political Psychology*, 41.S1 (2020), pp. 3–33; Ariadna Matamoros-Fernández and Johan Farkas, 'Racism, Hate Speech, and Social Media: A Systematic Review and Critique', *Television & New Media*, 22.2 (2021), pp. 205–24; Diana Rieger and others, 'Assessing the Extent and Types of Hate Speech in Fringe Communities: A Case Study of Alt-Right Communities on 8chan, 4chan, and Reddit', *Social Media + Society*, 7.4 (2021), pp. 1–14.

58 See, for example, Stephanie Bühner, Kevin Koban, and Jörg Matthes, 'The WWW of Digital Hate Perpetration: What, Who, and Why? A Scoping Review', *Computers in Human Behaviour*, 159 (2024), p. 108321.

In conceptualising perpetration, it is important to reflect on whether the harm itself needs to be intended by the person creating or sharing the content or even whether they necessarily perceive or fathom how it unfolds its harmful effect. For example, a social media post may include the slur 'gypsy', that is considered racist in most places, but has developed as a normalised term in the UK context.⁵⁹ Regardless of the context it is posted in, the intentionality of the racist connotation may remain unclear. And yet, whether intentionally racist and denigrating or not, its use could cause psychological harm to people seeing the post. While it may dilute the understanding of a perpetrator if we analytically approach it from the perspective of the (potential) victims and whether it is perceived as harmful, I would argue that any understanding of perpetration that necessitates intentionality is empirically untenable as it is impossible to actually study the intentionality of individual posts in combination with their effects. As such, I argue that we should not stipulate any specification of whether an individual is aware of or intends the harm in our conceptual approach to understanding perpetrators, but rather let the act of perpetration be foregrounded in the effect that it unfolds, tying into action-oriented approaches to violence discussed above.⁶⁰

The flipside to this discussion, however, is also to ask whether the definition of a perpetrator necessitates that a harm is actually created for a victimised person or group or whether it suffices for content to have been created or shared in ways that could *potentially* have harmed someone. Here, I would again argue for an effect-centric understanding of perpetrators and perpetration that puts the onus on the actual harm created. For example, if someone creates a video including racist slurs but no-one is actually harmed by the content, (if only few viewers even see the video and they are not members of the targeted group or Black viewers brush it off as unimportant). Here, as we cannot realistically develop an empirically tenable metric for *potential* harm beyond actual experienced harm, if no harm actually occurred, I would argue that those responsible for creating such content should not be classified as perpetrators.

59 Hazel Marsh and Julia Morgan, 'Decolonising the University Curriculum: An Investigation into Current Practice Regarding Gypsy, Roma and Traveller Communities', *Race Ethnicity and Education*, 28.6 (2025), pp. 855-75 (pp. 858-59).

60 Gudehus, 'Violence as Action'; Gudehus, 'Action Action Action'; Williams, 'Thinking beyond Perpetrators, Bystanders, Heroes'.

Finally, it should be emphasised here that I, of course, do not mean that *any* action perceived as harmful to an individual should be conceptualised as an act of perpetration. For example, if someone posts a picture of a spider that elicits psychological harm in someone with a phobia, this would not make them a perpetrator. Instead we can borrow from research on discursive violence and suggest that it is only harmful if and when a misrecognition or denigration of somebody's identity is involved.⁶¹

Overall, then, I would argue that for digital perpetration it is necessary for a harm to have been created and subjectively perceived by a victim through some form of digital content, irrespective of whether this harm was intentional on the part of the perpetrator or not. Though this subjective approach that prioritises the perspective of the victim broadens perpetration considerably, it strikes me as the only workable approach to understanding perpetrator harm in digital spaces as we cannot realistically adjudicate potential harms.

PARTICIPATION

Next, we turn to what participation in these harms can mean in defining perpetrators in online spaces. Kilger's definition refers mostly to the 'use of' various technologies for harm, leaving the actual acts this entails mostly open.⁶² Here, the previously mentioned WEF report suggests that harm can be created in content production, distribution, and consumption; furthermore, it differentiates harms that result from contact in online interactions and those from conduct that was facilitated by technology.⁶³ This broad approach highlights how difficult it is to pinpoint a precise definition of the *degree of participation* for online perpetration. The nature of cause and effect are considerably more convoluted in digital spaces than they are for perpetration in physical violence and there are several questions that emerge.

It seems sensible to centre the primary locus of responsibility in content production: to view those who create original content that harms others as perpetrators both of hybrid and digital violence. For example, this could include recording a racist diatribe that disparag-

61 Hannes Kuch and Steffen Herrmann, 'Philosophien Sprachlicher Gewalt. Eine Einführung', in *Philosophien Sprachlicher Gewalt: 21 Grundpositionen von Platon Bis Butler*, ed. by Hannes Kuch and Steffen Herrmann (Velbrück Wissenschaft, 2010).

62 Kilger (p. 606).

63 World Economic Forum.

es Black people or writing antisemitic content in the form of a conspiracy myth. Both forms of content creation relate to an entire group and could generate harm towards any member of the group that sees the content. Content can furthermore be created to more specifically target individuals, for example, phrasing a homophobic message that denigrates a man known to be gay or creating an Islamophobic meme to post on the feed of a specific Muslim person. In both forms of content, for an entire group or for a specific individual, the essence of the harm lies in the content creation through the discursive construction of discriminatory ideas, stereotypes, and misinformation: sometimes through subtle references, sometimes through blatant statements. Here, attributing the perpetrator in the creation process would mean understanding who generated the content, an undertaking that is not always clear-cut, when memes are developed; pictures are re-captioned; and ideas of other people are used as text and copied and pasted, for example. Multiple people can be involved in the process of creation, during which the content eventually becomes harmful. I would argue that the threshold here must qualitatively be placed in a way that seeks to identify who is responsible for bringing together ideas, images, and text in a way that ultimately unfolds harm. As such, there can, of course, be multiple actors deemed responsible.

Beyond content creation, the distribution of potentially harmful posts is also an important process in which perpetration can occur, not least as I argued previously that perpetration lies not just in the potential of harm in a post but in its realisation. Hence, only if a post is actually seen by individuals of the targeted group and actually harms them does the creation of the post even make the creator a perpetrator. As such, those who may not create their own content but contribute to the dynamics of its distribution can also be plausibly viewed as perpetrators. For example, if someone with a large following on a social media platform like X or Facebook shared an inciteful post this may then feature in the feeds of, be seen by, and harm many more people than if they had not shared the content. As such, even if the same thing is posted by two people, one of whom is an influencer and the other is not, the former is more likely to be deemed a perpetrator even though they performed the exact same action because the level of harm created by their action differs. It is also dependent on the algorithms of each platform as to who actually sees the posts, whether they remain within echo chambers, or whether they also reach those people who may

be harmed by them.⁶⁴ In terms of distribution, we must then also ask what threshold is necessary for deeming an individual a perpetrator. Does it necessitate an active promotion of content through sharing it? Or would one also consider attention paid to specific kinds of content, searching for it, and reading it, each of which also plays into algorithms and then makes its display for others more likely?⁶⁵ As I will discuss in the next section the latter could also easily be viewed from a perspective of what constitutes a bystander, highlighting the grey zones that abound with regard to responsibility for violence in digital spaces.

Overall, we see that the participation aspect in conceptualising who can be seen as a perpetrator of digital violence raises several problems in defining actual thresholds. However, this section has argued that the essence of perpetration needs to be viewed as how someone has *actually contributed* to the harm, that is it could be through creating original content, sharing it, liking it, retweeting it, or many of the other ways that different social media platforms allow their users to engage with the content and thus make it more likely that others will see it – and in the case of digital violence be harmed by it. While it is impossible to define a clear threshold for all platforms and all types of violence, the main thrust of this section has been to critically disaggregate responsibility for creating a connection between the content (in its creation or distribution) and the harm it creates in being consumed by a member of the target group.

Differentiating Perpetrators from other Roles

This article cannot interrogate the specific idea of perpetrators in digital spaces without also considering what other roles, like bystanders and upstanders, mean online. In conceiving of a bystander as someone who witnesses the violence but does not intervene,⁶⁶ this suggests in its application to online spaces that someone sees a post on social media that contains vitriolic content that psychologically harms a targeted group or individual, but the bystander does not actively engage with

64 Ludovic Terren Ludovic Terren and Rosa Borge-Bravo Rosa Borge-Bravo, 'Echo Chambers on Social Media: A Systematic Review of the Literature', *Review of Communication Research*, 9 (2021).

65 Jaeho Cho and others, 'Do Search Algorithms Endanger Democracy? An Experimental Investigation of Algorithm Effects on Political Polarization', *Journal of Broadcasting & Electronic Media*, 64.2 (2020), pp. 150–72.

66 Williams, 'Thinking beyond Perpetrators, Bystanders, Heroes'.

the post by sharing it, liking it, commenting on it, or similar (depending on the platform's architecture). This individual thus neither supports the post's messaging nor undermines or challenges its content. However, in the complex nature of how social media platforms work, even perception of a post impacts its algorithmic success. For example, if someone spends a longer time looking at a post but does not share it or engage in any other way, the algorithm may still use this information on the duration of attention given to it to highlight it to other users (who in turn may engage more actively). Thus, almost all decisions and engagements on social media impact how a post is displayed to others.

Here, Rothberg's idea of implicated subjects is useful as it highlights how people can be entangled with acts of violence and positions of privilege and power without themselves being agents in creating the harm itself. While Rothberg's focus is on power and privilege derived from historical legacies, most frequently in the context of racialised hierarchies, I would nonetheless argue that implication as a concept is useful in the digital realm to highlight a certain form of implication as users' actions and inactions help produce and reproduce the positions of victims and perpetrators.⁶⁷ Of course, these forms of implication are not structural as in Rothberg's original concept in which people's positionalities and group memberships are the foundation for their implication, but instead rest on their presence in an online space. However, while one could argue that people have the possibility to remove themselves from their implication by not being online, practically this is a relatively unrealistic suggestion in an era in which much social, political and economic life occurs in digital spaces.

Further, it helps to understand who can be deemed a perpetrator in online spaces by reflecting on what upstanding means in the context of digital violence. Most directly, I would argue that upstanding refers to direct intervention into situations of digital violence, for example when individuals push back against discriminatory language or seek to debunk conspiracy myths and disinformation. Furthermore, upstanders can emotionally support victims of digital violence or express their solidarity with individuals or communities. While these actions can all provide support in mitigating the harms of hybrid and digital violence, they can also be critically viewed in that any form of engagement with content foregrounds it for the algorithms and thus can also inadvert-

67 Rothberg, p. 1.

ently lead to popularisation.⁶⁸ Activists at times circumvent this issue by posting content in ways that avoid drawing algorithmic attention to digital violence, for example not re-posting problematic content but posting a message with the original post included as a screenshot.⁶⁹ Another method of upstanding that seeks to undermine digital violence's algorithmic power is to flag problematic content, a function that some social media platforms include to allow users to notify them about content that violates terms and conditions for example against hate speech.⁷⁰

More broadly, digital technologies can also be used to prevent physical violence in what international relations scholar Rhiannon Neilsen terms 'cyber humanitarian interventions'.⁷¹ Neilsen distinguishes between, first, hacking into perpetrators' digital technological infrastructures to disrupt communication, interfere with logistics, or remove finances, for example hampering ISIS' ability to plan and implement military operations. Second, 'targeted educational campaigns' can be devised for social media to delegitimise violence and impact perpetrators' motivations, for example, highlighting to Burmese youth a human rights agenda and moral qualms around participating in killings of Rohingya. And, third, 'atrocities suppressing disinformation campaigns' influence perpetrators' decision-making by intentionally deceiving them and changing their calculus for their violent action. Here upstanding does not just seek to counter digital violence, but uses digital tools for upstanding in and of itself.

68 Cho and others.

69 Anna Ricarda Luther and others, 'Social Media for Activists: Reimagining Safety, Content Presentation, and Workflows', *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, no. 956 (2025), pp. 1-18 (p. 9).

70 This functionality is being rolled back in several of the largest social media platforms, for example, on X since Elon Musk bought the platform or on Meta's platforms like Facebook or Instagram since the beginning of Donald Trump's second presidency.

71 Rhiannon Neilsen, 'Coding Protection: "Cyber Humanitarian Interventions" for Preventing Mass Atrocities', *International Affairs*, 99.1(2023), pp. 299-319; Rhiannon Neilsen, *Algorithms & Atrocities: Cyber Humanitarian Interventions to Prevent Genocide, War Crimes, and Crimes Against Humanity* (Oxford University Press, forthcoming).

Conclusion

Overall, this article ends with more questions than it began with. Research on perpetrators of physical violence already provides myriad understandings of what constitutes this role, but I showed here that underlying many of these understandings are two fundamental dimensions: the nature of harm and degree of participation. While mapping this understanding of perpetrators onto digital spaces is anything but clear-cut, this article has attempted to sketch some first ideas on what this could look like. First, I have argued that harm needs to be defined from the victims' perspective irrespective of perpetrators' intent in the creation of digital content, with psychological harms being rendered through purely digital violence as well as physical harm when violence is hybrid. Second, it is difficult to define thresholds of what constitutes a sufficient degree of participation in or contribution to digital violence to deem something an act of perpetration. However, I have argued that it is important to consider not only the creation of content but also how it is distributed to evaluate what impact each of these steps may have had in actually causing the relevant harm. Which actions this specifically entails depends somewhat on the platforms' architecture and each individual case of violence. Finally, I have shown how hard it is to attribute responsibility for individual actions in the generation of harm in digital violence, making the grey zones between perpetrators and implicated subjects – as bystanders can be more usefully conceived of in digital spaces – significant and large.

At this point, I am convinced that it is useful to think of many acts in digital spaces as constituting perpetration. And yet I am also skeptical regarding how the concept of perpetrator can be made operationalizable for digital violence and those involved, given struggles with pinpointing the specific nature of harm and which thresholds could be useful in defining participation. A minimal conclusion may be that it is useful to speak of perpetrators in hybrid violence, including those involved online in the genesis of physical violence through planning, mobilising, and legitimating, but that more considerably more research will be needed to think through how we can classify perpetrators in the context of purely digital violence.

But for both digital and hybrid violence the question of how to conceptually approach those who contribute to the violence is not just relevant and important for social scientific research. It also has practical repercussions: with regard to how to think of their actions (as crimes?

Or as inconsequential online postings?), how to deal with these actions legally, and what to do with the content. Here, on the one hand, digital violence has never been easier to prosecute as all necessary evidence is openly on display. On the other hand, however, it necessitates a political will to combat digital violence that does not rest on prosecution only occurring when the victims themselves pursue it. Also, there needs to be clearer legislation on how social media companies deal with digital violence, prioritising 'free speech' or violence prevention and deciding whether the onus needs to be on social media companies finding and removing digital violence from their platforms or passing evidence on to prosecutors in ways that facilitate trials against those we could call perpetrators.⁷²

While I end this article questioning how precisely we can fruitfully use the concept of perpetrators 2.0 for labelling individuals who have committed acts of perpetration in digital spaces, I hope this article can serve as the beginning of a research debate on the issue. The issue itself is important, and social scientists and other scholars need a way to conceptualise how digital violence occurs and what this means for responsibility, culpability, and implication, even if we shy away from using the word perpetrator.

72 Anna Veronica Banchik, 'Disappearing Acts: Content Moderation and Emergent Practices to Preserve at-Risk Human Rights-Related Content', *New Media & Society*, 23.6 (2021), pp. 1527–44; Susan Benesch, 'But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies', *Yale Journal on Regulation Bulletin*, 38 (2020), p. 86.

Works Cited

- Ambos, Kai, 'Criminologically Explained Reality of Genocide, Structure of the Offence and the "Intent to Destroy" Requirement', in *Collective Violence and International Criminal Justice: An Interdisciplinary Approach*, ed. by Alette Smeulers (Intersentia, 2010), pp. 153–73
- Anderson, Haley S., 'Culpability in Atrocity and the Role of Complicit Observer', *Temple International & Comparative Law*, 37.2 (2023), pp. 11–22
- Arendt, Hannah, *Eichmann in Jerusalem: A Report on the Banality of Evil* (Penguin Books, 1994)
- Awan, Imran, 'Cyber-Extremism: Isis and the Power of Social Media', *Society*, 54.2 (2017), pp. 138–49, doi:10.1007/s12115-017-0114-0
- Bail, Christopher A., and others, 'Exposure to Opposing Views on Social Media Can Increase Political Polarization', *Proceedings of the National Academy of Sciences*, 115.37 (2018), pp. 9216–21, doi:10.1073/pnas.1804840115

- Banchik, Anna Veronica, 'Disappearing Acts: Content Moderation and Emergent Practices to Preserve at-Risk Human Rights-Related Content', *New Media & Society*, 23.6 (2021), pp. 1527-44, doi:10.1177/1461444820912724
- Barberá, Pablo, and others, 'Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?', *Psychological Science*, 26.10 (2015), pp. 1531-42
- Bastos, Marco, *Spatializing Social Media: Social Networks Online and Offline* (Routledge, 2022)
- Benesch, Susan, 'But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies', *Yale Journal on Regulation Bulletin*, 38 (2020), pp. 86-111
- Bennett, W. Lance, and Alexandra Segerberg, 'The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics', *Information, Communication & Society*, 15.5 (2012), pp. 739-68, doi:10.1080/1369118X.2012.670661
- Bernroider, Edward W. N., G. Harindranath, and Sherif Kamel, 'From Connective Actions in Social Movements to Offline Collective Actions: An Individual Level Perspective', *Information Technology & People*, 35.8 (2022), pp. 205-30, doi:10.1108/ITP-08-2020-0556
- Bilewicz, Michał, and Wiktor Soral, 'Hate Speech Epidemic: The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization', *Advances in Political Psychology*, 41.S1 (2020), pp. 3-33
- Bilsky, Leora, and Rachel Klagsbrun, 'The Desk Perpetrator, the Expert Witness, and the Role of Law: The Trial of Arthur Greiser', *Journal of Genocide Research*, 27.1 (2025), pp. 1-22, doi:10.1080/14623528.2023.2255407
- Bührer, Stephanie, Kevin Koban, and Jörg Matthes, 'The WWW of Digital Hate Perpetration: What, Who, and Why? A Scoping Review', *Computers in Human Behavior*, 159 (2024), p. 108321, doi:10.1016/j.chb.2024.108321
- Campbell, Bradley, 'Contradictory Behavior During Genocides', *Sociological Forum*, 25.2 (2010), pp. 296-314
- Cho, Jaeho, and others, 'Do Search Algorithms Endanger Democracy? An Experimental Investigation of Algorithm Effects on Political Polarization', *Journal of Broadcasting & Electronic Media*, 64.2 (2020), pp. 150-72, doi:10.1080/08838151.2020.1757365
- Conway, Maura, 'Determining the Role of the Internet in Violent Extremism and Terrorism', *Studies in Conflict & Terrorism*, 40.1 (2017), pp. 77-98
- Della Ratta, Donatella, *Shooting a Revolution: Visual Media and Warfare in Syria* (Pluto Press, 2018)
- Digeser, P. E., 'Collaboration and Its Political Functions', *American Political Science Review*, 116.1 (2022), pp. 200-12, doi:10.1017/S0003055421000745
- Donà, Giorgia, "'Situated Bystanderism" During and After the Rwandan Genocide', *Journal of Genocide Research*, 20.1 (2017), pp. 1-19
- Drumbl, Mark, *Atrocity, Punishment, and International Law* (Cambridge University Press, 2007)

- Fischer, Peter, and others, 'The Bystander-Effect: A Meta-Analytic Review on Bystander Intervention in Dangerous and Non-Dangerous Emergencies', *Psychological Bulletin (US)*, 137.4 (2011), pp. 517–37
- Fox, Nicole, and Hollie Nyseth Brehm, "'I Decided to Save Them': Factors That Shaped Participation in Rescue Efforts during Genocide in Rwanda", *Social Forces*, 96.4 (2018), pp. 1625–48
- Fuchs, Christian, *Social Media: A Critical Introduction*, 3rd edn (SAGE, 2021)
- Fujii, Lee Ann, *Killing Neighbors: Webs of Violence in Rwanda* (Cornell University Press, 2009)
- , 'Rescuers and Killer-Rescuers during the Rwanda Genocide', in *Resisting Genocide: The Multiple Forms of Rescue*, ed. by Jacques Sémelin, Claire Andrieu, and Sarah Gensburger (Columbia University Press, 2011), pp. 145–57
- Fulbrook, Mary, *Bystander Society: Conformity and Complicity in Nazi Germany and the Holocaust* (Oxford University Press, 2023)
- Gallacher, John D., Marc W. Heerdink, and Miles Hewstone, 'Online Engagement Between Opposing Political Protest Groups via Social Media Is Linked to Physical Violence of Offline Encounters', *Social Media + Society*, 7.1 (2021), pp. 1–16 doi:10.1177/2056305120984445
- Gates, Scott, and Sukanya Podder, 'Social Media, Recruitment, Allegiance and the Islamic State', *Perspectives on Terrorism*, 9.4 (2015), pp. 107–16
- Gohdes, Anita R., 'Studying the Internet and Violent Conflict', *Conflict Management and Peace Science*, 35.1 (2018), pp. 89–106, doi:10.1177/0738894217733878
- Greijdanus, Hedy, and others, 'The Psychology of Online Activism and Social Movements: Relations between Online and Offline Collective Action', *Current Opinion in Psychology*, 35 (2020), pp. 49–54, doi:10.1016/j.copsyc.2020.03.003
- Gudehus, Christian, 'Action Action Action', *Journal of Perpetrator Research*, 3.1 (2020), pp. 188–95
- , 'Violence as Action', in *Perpetrators and Perpetration of Mass Violence: Dynamics, Motivations and Concepts*, ed. by Timothy Williams and Susanne Buckley-Zistel (Routledge, 2018), pp. 36–57
- Harmer, Emily, and Karen Lumsden, 'Online Othering: An Introduction', in *Online Othering*, ed. by Karen Lumsden and Emily Harmer (Springer International Publishing, 2019), pp. 1–33 <http://link.springer.com/10.1007/978-3-030-12633-9_1> [accessed 9 August 2021]
- Jessee, Erin, *Negotiating Genocide in Rwanda* (Palgrave Macmillan, 2017)
- Kaufman, Zachary D., 'Protectors of Predators or Prey: Bystanders and Upstanders amid Sexual Crimes', *Southern California Law Review*, 92.6 (2019), pp. 1317–406
- KhosraviNik, Majid, and Eleonora Esposito, 'Online Hate, Digital Discourse and Critique: Exploring Digitally-Mediated Discursive Practices of Gender-Based Hostility', *Lodz Papers in Pragmatics*, 14.1 (2018), pp. 45–68, doi:10.1515/lpp-2018-0003

- Kilger, Max, 'Interventions, Policies, and Future Research Directions in Cybercrime', in C. A. Cuevas and C. M. Rennison, *The Wiley Handbook on the Psychology of Violence* (Wiley, 2016), pp. 604–22
- Kuch, Hannes, and Steffen Herrmann, 'Philosophien Sprachlicher Gewalt. Eine Einführung', in *Philosophien Sprachlicher Gewalt: 21 Grundpositionen von Platon Bis Butler*, ed. by Hannes Kuch and Steffen Herrmann (Velbrück Wissenschaft, 2010)
- Levine, Mark, Richard Philpot, and Anastasiia G. Kovalenko, 'Rethinking the Bystander Effect in Violence Reduction Training Programs', *Social Issues and Policy Review*, 14.1 (2020), pp. 273–96, doi:10.1111/sipr.12063
- Luft, Aliza, 'Toward a Dynamic Theory of Action at the Micro Level of Genocide: Killing, Desistance, and Saving in 1994 Rwanda', *Sociological Theory*, 33.2 (2015), pp. 148–72
- Luther, Anna Ricarda, and others, 'Social Media for Activists: Reimagining Safety, Content Presentation, and Workflows', *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, no. 956 (2025), pp. 1–18
- Mann, Michael, *The Dark Side of Democracy* (Cambridge University Press, 2005)
- Markusen, Eric, 'Mechanisms of Genocide', in *Will Genocide Ever End?*, ed. by Carol Rittner, John K. Roth, and James M. Smith (Paragon House, 2002), pp. 83–89
- Marsh, Hazel, and Julia Morgan, 'Decolonising the University Curriculum: An Investigation into Current Practice Regarding Gypsy, Roma and Traveller Communities', *Race Ethnicity and Education*, 28.6 (2025), pp. 855–75, doi:10.1080/13613324.2025.2474936
- Matamoros-Fernández, Ariadna, and Johan Farkas, 'Racism, Hate Speech, and Social Media: A Systematic Review and Critique', *Television & New Media*, 22.2 (2021), pp. 205–24, doi:10.1177/1527476420982230
- McDoom, Omar, *The Path to Genocide in Rwanda: Security, Opportunity, and Authority in an Ethnocratic State* (Cambridge University Press, 2020)
- Meltwater, *Digital 2024: Global Overview Report* (Meltwater, 2024)
- Moojiman, Marlon, and others, 'Moralization in Social Networks and the Emergence of Violence during Protests', *Nature Human Behaviour*, 2.6 (2018), pp. 389–96
- Müller, Karsten, and Carlo Schwarz, 'Fanning the Flames of Hate: Social Media and Hate Crime', *Journal of the European Economic Association*, 19.4 (2021), pp. 2131–67, doi:10.1093/jeea/jvaa045
- Munn, Luke, 'More than a Mob: Parler as Preparatory Media for the U.S. Capitol Storming', *First Monday*, published online 7 February 2021, doi:10.5210/fm.v26i3.11574
- Neilsen, Rhiannon, *Algorithms & Atrocities: Cyber Humanitarian Interventions to Prevent Genocide, War Crimes, and Crimes Against Humanity* (Oxford University Press, forthcoming)
- , 'Coding Protection: "Cyber Humanitarian Interventions" for Preventing Mass Atrocities', *International Affairs*, 99.1 (2023), pp. 299–319, doi:10.1093/ia/iiaac261
- Ng, Lynnette Hui Xian, Iain J. Cruickshank, and Kathleen M. Carley, 'Coordinating Narratives Framework for Cross-Platform Analysis in the 2021 US Capitol Riots', *Computational*

- and *Mathematical Organization Theory*, 29.3 (2023), pp. 470–86, doi:10.1007/s10588-022-09371-2
- Nyseth Brehm, Hollie, and others, 'In the Aftermath: The Post-Conflict Social and Economic Consequences of Rescue During Genocide', *Journal of Genocide Research*, 25.1 (2023), pp. 24–45
- Osmundsen, Mathias, and others, 'Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter', *American Political Science Review*, 115.3 (2021), pp. 999–1015, doi:10.1017/S0003055421000290
- Pauwels, Lieven, and Nele Schils, 'Differential Online Exposure to Extremist Content and Political Violence: Testing the Relative Strength of Social Learning and Competing Perspectives', *Terrorism and Political Violence*, 28.1 (2016), pp. 1–29, doi:10.1080/09546553.2013.876414
- Perešin, Anita, 'Fatal Attraction: Western Muslimas and ISIS', *Perspectives on Terrorism*, 9.3 (2015), pp. 21–38
- Peterson, Jillian, and James Densley, 'Cyber Violence: What Do We Know and Where Do We Go from Here?', *Aggression and Violent Behaviour*, 34 (2017), pp. 193–200, doi:10.1016/j.avb.2017.01.012
- Piazza, James A., 'Fake News: The Effects of Social Media Disinformation on Domestic Terrorism', *Dynamics of Asymmetric Conflict*, published online 2021, doi:https://doi.org/10.1080/17467586.2021.1895263
- Rieger, Diana, and others, 'Assessing the Extent and Types of Hate Speech in Fringe Communities: A Case Study of Alt-Right Communities on 8chan, 4chan, and Reddit', *Social Media + Society*, 7.4 (2021), pp. 1–14, doi:10.1177/20563051211052906
- Robbins, Bruce, *The Beneficiary* (Duke University Press, 2017)
- Rothberg, Michael, *The Implicated Subject: Beyond Victims and Perpetrators* (Stanford University Press, 2019)
- Sæbø, Øystein, Tommaso Federici, and Alessio M. Braccini, 'Combining Social Media Affordances for Organising Collective Action', *Information Systems Journal*, 30.4 (2020), pp. 699–732, doi:10.1111/isj.12280
- Schradie, Jen, 'The Digital Activism Gap: How Class and Costs Shape Online Collective Action', *Social Problems*, 65.1 (2018), pp. 51–74
- Smeulders, Alette, *Perpetrators of Mass Atrocities: Terribly and Terrifyingly Normal?* (Routledge, 2024)
- Staub, Ervin, 'The Psychology of Bystanders, Perpetrators, and Heroic Helpers', in *Understanding Genocide: The Social Psychology of the Holocaust*, ed. by Leonard S. Newman and Ralph Erber (Oxford University Press, 2002), pp. 11–42
- Straus, Scott, 'Studying Perpetrators: A Reflection', *Journal of Perpetrator Research*, 1.1 (2017), pp. 28–38
- Sunstein, Cass R., *#Republic: Divided Democracy in the Age of Social Media* (Princeton University Press, 2017)

- Terren, Ludovic, and Rosa Borge-Bravo, 'Echo Chambers on Social Media: A Systematic Review of the Literature', *Review of Communication Research*, 9 (2021), pp.100-18 doi:10.12840/ISSN.2255-4165.028
- Verdeja, Ernesto, 'Moral Bystanders and Mass Violence', in *New Directions in Genocide Research*, ed. by Adam Jones (Routledge, 2012), pp. 153-68
- Wahlström, Mattias, Anton Törnberg, and Hans Ekbrand, 'Dynamics of Violent and Dehumanizing Rhetoric in Far-Right Social Media', *New Media & Society*, (2021), pp.3290-3311 doi:10.1177/1461444820952795
- Walker, Shawn, Dan Mercea, and Marco Bastos, 'The Disinformation Landscape and the Lockdown of Social Platforms', *Information, Communication & Society*, 22.11 (2019), pp. 1531-43
- Williams, Timothy, 'Agency, Responsibility, and Culpability: The Complexity of Roles and Self-Representations of Perpetrators', *Journal of Perpetrator Research*, 2.1 (2018), pp. 39-64
- , *The Complexity of Evil: Perpetration and Genocide* (Rutgers University Press, 2021)
- , 'Thinking beyond Perpetrators, Bystanders, Heroes: A Typology of Action in Genocide', in *Perpetrators and Perpetration of Mass Violence: Dynamics, Motivations and Concepts*, ed. by Timothy Williams and Susanne Buckley-Zistel (Routledge, 2018), pp. 17-35
- World Economic Forum, *Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms: Insight Report* (2023)
- Zhong, Wei, and others, 'Proud Boys on Telegram', *Journal of Quantitative Description: Digital Media*, 4 (2024), pp. 1-47, doi:10.51685/jqd.2024.003
- Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov, 'Political Effects of the Internet and Social Media', *Annual Review of Economics*, 12.1 (2020), pp. 415-38, doi:10.1146/annurev-economics-081919-050239

Timothy Williams is Junior Professor of Insecurity and Social Order as well as co-chairman of the research centre RISK, both at the University of the Bundeswehr Munich.