4 Cultural violence and fragmentation on social media

Interventions and countermeasures by humans and social bots

Jasmin Haunschild, Marc-André Kaufhold, and Christian Reuter

Mobile technologies and social media services are among the socio-technological innovations that have an enormous impact transforming modern culture and political processes. Social media are often defined as a "group of internet-based applications […] that allow the creation and exchange of user-generated content" (Kaplan and Haenlein 2010). Shaping opinions, politics, participation, and protest (Wulf et al. 2013), they are used by citizens for news consumption and social exchange (Robinson et al. 2017); by journalists for reporting, analyzing, and collecting information (Stieglitz et al. 2018a); and by organizations to monitor crises, emergencies, customer feedback, and sentiment, among others (Haunschild et al. 2020). Large-scale international events, such as the 2010 Arab Spring, showcased the potential of socio-technological transformations: Citizens were not passive victims but active and autonomous participants utilizing social media to coordinate protest and for crisis response (Reuter and Kaufhold 2018). However, in other cases, citizens' activities coordinated via social media also increased the complexity of tasks and pressure for formal authorities, since the lack of state control has not had only empowering or benign effects. Instead, on social media, false information spreads fast and it is easy for groups to find an audience there, either to enhance their profit or to target vulnerable groups with dangerous ideology.

To understand the role of social media in contributing to peace and conflict, the conceptions of war, peace, and security from the domains of peace and conflict research and security studies are helpful. They have identified the need to deepen and broaden understandings of the relevant actors, referent objects, and threats (Booth 2007). While traditionally, the state had been the only actor and threatened object, the conflict in former Yugoslavia showed that social groups can also be threatened by their own state and by other groups within the same state (Waever 1993). This is even more the case with regard to cyberspace, where it "is also often unclear whether the actors pursue military-strategic or commercial objectives and whether they have no political, but maybe commercial interests maybe on behalf of the private sector or on behalf of a state or group with political intents" (Reuter 2020: 13). Similarly, the conception of human security shines a light on the potential threats to individuals, which do not only concern security

DOI: 10.4324/9781003110224-5

aspects such as direct attacks, but also safety issues, such as health, development, and environmental threats (Booth 2007). This conception of the potential sources of harm and insecurity helps understand the role of social media as a socio-technological innovation that, along with its emancipatory power, also amplifies existing threats. In this way, social media cannot only contribute to direct, physical violence, e.g. through facilitating the recruitment of terrorists (Weimann 2016), but also to structural and cultural violence by creating, reinforcing, and escalating grievances and political fragmentation, e.g. through the dissemination of fake news and of extremist ideologies (Reuter et al. 2017), partly aided by social bots (Stieglitz et al. 2017). Cultural violence is understood as "all aspects of a culture that are used to justify direct or structural violence" (Galtung 2007: 341), while structural violence describes "unjust economic, social and political conditions and institutions that harm people by preventing them from meeting their basic needs" (Campbell et al. 2010: 390). Accordingly, socio-technological transformations with potential for structural violence can be witnessed (a) in the use and misuse of social media platforms to foster intercultural understanding, but also to disseminate harmful content; and (b) in the use of social bots that can feign widespread support and amplify the spread of harmful content. On the other hand, innovations and regulations are also developed to mitigate socio-technological uncertainties in a way that curbs the misuse while maintaining the positive potential of social media.

In this context, social media are relevant as an important platform for shaping culture, both to foster cultural peace, as well as to be abused for structural and cultural violence. Notions of cyber peace have already recognized the structural dimension, when cyber peace is described as "the peaceful application of cyberspace to the benefit of humanity and the environment [including] the renouncement of all cyberwar activities, but [also the use of] the whole of the communication infrastructure for international understanding" (FifF o.D.). The study of cyber peace should take into account insights from peace and conflict research on conditions that foster peace and conflict in other realms of society, as well as contributions from fields of human–computer iInteractions and IT security, to create designs and modes of interacting with technology that foster peace (Reuter 2019).

In a socio-technological setting, cultural violence might become tangible by the actual *content*, but is also driven by the motives of *actors* and mediated by the capabilities of *technology*. To address these three perspectives with emergent phenomena in cyberspace, the following chapters will examine (1) *fake news* and their exploiting of existing grievances and distrust; (2) *cyber terrorism* showing how actors exploit disadvantaged groups and further alienate them from the society they live in; and (3) the technology of *social bots*, networks of which can be bought by actors to further their political or economic agenda through manipulation and fake news. By conducting a narrative literature review, the chapter identifies challenges and explores socio-technological countermeasures to cultural violence perpetrated on social media, shedding light on the social grievances exploited by technology. It thus shows that both technological and

social interventions are fruitful. But it also shows that ultimately the question of how to differentiate the voicing of legitimate grievances and the organization of political opposition from malicious efforts at politically and financially motivated fragmentation remains open and cannot be solved by technology or social media firms who are currently the dominant actors for setting the rules on social media (DeNardis and Hackl 2015).

This chapter illustrates different phenomena that increase societal fragmentation and erode trust in communities and political institutions. First, the case of fake news shows that existing grievances can be nourished by fake news. Secondly, targeted propaganda on social media uses existing grievances to turn individuals against other societies in the process of terrorism recruitment. While social media primarily increases the reach of existing voices, the third case of social bots shows the potential to artificially amplify certain voices, skewing the discourse according to the financial and political agendas of those buying the service of bots. Each case closes by showing socio-technological countermeasures to the exploitation of social media. The chapter concludes by discussing the implications of the socio-technological transformation through social media for legitimacy and regulatory authority.

Fabricated, manipulated, and misinterpreted content: The issue of fake news in social media

By increasing communication among online users, social media can contribute to cultural violence, for instance, by emphasizing religious, ideological, and language divides, including by spreading misinformation and disinformation, commonly known as "fake news". While the term was originally used to mostly refer to comedy news shows, in 2016 the perception changed when many fake stories went viral and started to affect political parties globally and impacted opinions on a larger scale than before (Becker 2016). Although "fake news" is a popular and frequent term, it is often mingled with other phenomena, facilitating misuse of the term to discredit undesired news (Cooke 2017), political opponents, and conspiracy theories.

Dissemination of fake news in social media

Fake news are news articles that are "intentionally and verifiably false and could mislead readers" (Allcott and Gentzkow 2017: 213). The topics of fake news often lead to high emotions and are associated with controversial discussions like migration, child abuse, or war (Ziegele et al. 2014), but prevalent types of fake news differ across states and cultures (Humprecht 2019). Fake news can have serious consequences, e.g. influencing elections, stock markets, or leading to direct violence (Kaufhold and Reuter 2019). In an illustrative case in South Africa, foreign shops were attacked, leading to the deaths of 12 people, mostly nationals, while tensions between South Africans and Nigerians increased with footage on social media from different times and places claiming to portray attacks against

Nigerians (News Afrika 2016). This case shows how already existing xenophobia is exacerbated by social media, leading to retribution for violence that did not actually take place.

Often, political and financial motivations exist for generating fake news. Links from social media posts can result in vast advertising revenues if they are successfully published and shared and fake news have been used to manipulate the public opinion and debate. Well-known incidents are the recent US presidential election (McCarthy 2017) and the UK "brexit" referendum where false information have often been employed in combination with social bots (Mostrous et al. 2017).

Countermeasures against fake news

Three enablers and corresponding response vectors have been identified for countering fake news: To address the susceptibility of the "host" (news readers and social media users), education and clarification is the most promising avenue. Another enabler is a "conducive environment", consisting of toxic and complicit platforms, which can be addressed through regulation. Finally, the various types of fakes acting as "virulent pathogens" can be addressed through auto-detection (Rubin 2019). This leads to four possible approaches to countering fake news (see Table 4.1).

Most social networks have taken measures such as curating, deleting, and censoring. In doing so, even initially independent platforms now take the traditional journalistic role of information gatekeeper (Wohn et al. 2017). Many platforms provide mechanisms for users to flag content that they believe to be false. These annotations are then checked by experts, belonging either to the platform or to national independent fact-checking organizations. This expert-oriented checking of facts is based on human work and deals with the exposure of false statements. The experts check their researched and already created lists with the articles

Table 4.1 Measures against fake news in social media

Gatekeeping	Gatekeeping is the process through which information, including fake news, is filtered for dissemination, e.g. for publication, broadcasting, social media, or some other mode of communication (Barzilai-Nahon 2009).
Media literacy	The purpose of media literacy, which is a multidimensional process allowing people to access, evaluate, and create media, is to help people protect themselves against the potentially negative effects of (mass) media (Potter 2010).
Regulation/Law	Laws assist in fighting fake news and hate speech by forcing platforms to quickly delete illegal content, but potentially threaten freedom of speech (Müller and Denner 2017).
Algorithms/Tools	Algorithmic detection of fake news comprises classification-based, propagation-based, and survey-based approaches (Viviani and Pasi 2017) as well as user assistance tools (Hartwig and Reuter 2019).

flagged by Facebook users. In addition, technological means are used to limit the visibility of fake news on social media by reducing their relevance in news feeds and to limit their spread, e.g. reducing the amount of possible forwarding on messenger apps to five (Hern 2020).

Furthermore, efforts are made to increase the populations' media literacy. People with good media literacy can better navigate today's media and are able to identify and critique false news, but also create fake news themselves (Mihailidis and Viotty 2017). Hancock et al. (2008) show that the style of disinformation often differs from real news: Fraudsters rely more on sense-based, less on self-oriented, and more on other-oriented words. In addition, they use more negatively associated words, which provides guidance for people to detect fake news emotions (Newman et al. 2003). Furthermore, diverse non-state actors and associations are developing tools, such as the app Fake News Check (Neue Wege des Lernens e.V. 2017). Instead of the automatic flagging of fake news, the app aims to sensitize for the critical handling of news by helping users to ask the right questions and identify fake news through guided reflection of a set of 19 questions.

Regarding regulation, in many countries, laws have entered into force that require platforms to quickly delete illegal content, including hate speech. While celebrated for giving support to victims, it has also been widely criticized for threatening freedom of speech. Deleting fake news from social networks may create reactance and thus an even more fertile ground for conspiracy theories (Müller and Denner 2017). Additionally, such laws may incentivize social networks to delete content preemptively if there is any suspicion of fake news.

There are several approaches to use algorithms and tools for fake news detection. Such algorithms use classification-based (including machine learning), propagation-based (including social network analysis), and survey-based (including representative samples) approaches (Viviani and Pasi 2017). This also includes user assistance tools, for instance, *Fake Tweet Buster* helps Twitter users to identify a tweeted image as fake and tools such as *Trusty Tweet* and *Alethiometer* provide indicators and a browser plugin on the trustworthiness of tweets (Hartwig and Reuter 2019; Kaufhold and Reuter 2019).

These approaches place the responsibility for dealing with disinformation on different groups. While media literacy targets the recipients of fake news, regulation demands that either governments or social media platforms make and enforce rules about limiting the availability or spread of fabricated content. Gatekeeping can be performed either by experts employed by social media platforms or by journalists organized in independent fact-checking institutions (Graves 2018). Their results can either prevent fake news from being shown or can be used to inform consumers. Similarly, algorithmic solutions support any of the actors, pointing out identified fake news either to media consumers, to platforms, gatekeepers, or regulators, depending on who is deemed responsible. While citizens are undecided about who should take that responsibility, the majority of Germans supports relevant authorities' swift reaction to fake news, but also transparent journalism (Reuter et al. 2019).

Terrorist actors: Propaganda and recruitment in social media

As indicated, the spread of disinformation is strongly driven by the motivations of different actors. The recent past saw an increase in terrorist attacks across Europe, such as the November 2015 Paris attacks, the 2016 Brussels bombings, or the 2017 London Bridge attack (Stieglitz et al. 2018b). Besides direct violence and extensive media coverage of such events, the internet and especially social media are also used to promote cultural violence, e.g. by disseminating ideologies of terrorism and recruiting new members. Again, radicalization and recruitment into terrorist and extremist organizations is only possible where terrorist propaganda meets experiences or perceptions of injustice and grievances (Al-Saggaf 2016).

Terrorist propaganda and recruitment in social media

Research about terrorist organizations and social media mainly deals with the so-called Islamic State (IS, a.k.a. ISIS, ISIL, DEASH). Neer and O'Toole (2014) emphasize that especially Twitter is used by IS as a strategic tool to gain support from young jihadists, Ba'ath officials, and women. Klausen et al. (2012) stress that the British terrorist group al-Muhajiroun uses its international network of YouTube-channels elaborately for propaganda and the presentation of violent contents. Social media are used to incite phantasies and to normalize extreme views by creating an echo chamber of like-minded individuals (Awan 2017; Torok 2015). This leads to IS developing and disseminating "its central narratives, often by reframing familiar concepts such as jihad and martyrdom" (Torok 2015). In addition to propaganda targeted at vulnerable and like-minded people, terrorists also use tools such as Kik or Skype for "direct, real-time communication between recruiters and their audiences" (Weimann 2016: 82).

The IS propaganda helps in the recruitment not only of potential new fighters, but also of "technically proficient and talented users of social media to sustain the machinery of recruitment" (Gates and Podder 2015: 109). Since May 2014, IS videos or other media have been produced by the al-Hayat Media Center, a special production unit for Western recruitment. Their material exists in many languages and is spread via social media. For example, "IS released a video inciting Muslims to come and participate in jihad, featuring a German chant with an English translation" (Weimann 2016: 80).

Counter-terrorism in social media

A variety of different measures to counter-terrorism have been identified in research (see Table 4.2). Reuter et al. (2017) identify three categories of countermeasures: Clarification, parody/satire, and hacking. They show that private users are more adapt at reaching a wider audience as opposed to institutional accounts aiming to clarify. Satirical content is shown to receive most attention, while the success of hacking scenes is judged as limited due to the ease of reopening accounts and moving content to other platforms.

Table 4.2 Measures against terrorism

Clarification	Countering terrorist propaganda with logic to invalidate false information and simplistic portrayals.
Parody/Satire	Humorous imitation working through distortion and exaggeration (parody), critique and mockery (satire) of serious issues.
Hacking	Illegal "hacktivist" activities like attacking and blocking of pro-IS accounts and websites, supported by crowdsourced reporting of accounts of suspected terrorists. Includes legal activities of multiplying anti-IS parodist content.
Counter-narratives	A narrative that competes with another narrative. Narratives are compelling storylines which can explain events convincingly and from which inferences can be drawn.

Terrorists' activity and dependence on social media propaganda can also be seen as a weak spot that can be attacked with small and quick units that refute IS propaganda, expose untrue aspects, and damage the IS's credibility (Gartenstein-Ross 2015). Jeberson and Sharma (2015) focus on methods to identify terror suspects in social networks. Cheong and Lee (2011) describe that these data could be collected in a knowledge base in connection with intelligent data mining, visualization, and filter methods. They could be used by authorities for quick reaction and control. Furthermore, Sutton et al. (2008) deal with the application of backchannels as a special form of data mining for acquiring information. Instead of a strict censorship of radical contents, "terrorist communication strategies [should therefore be disturbed] by a mixture of technical (hacking) and especially psychological (anti-propaganda) means" (Weimann and Jost 2015). Gartenstein-Ross (2015) concludes that it would be a significant victory to weaken the strategic communication campaign of the IS. Weimann (2016) sees the security community and governments as well as researchers in the role of a counter-terrorism force. For the security community, according to Weimann (2016), it is necessary to include cyberspace in counter-terrorism strategies.

Hussain and Saltman (2014) emphasize that general censorship, similar to that of fake news, can be counterproductive, suggesting positive measures such as counter-narratives (Freedman 2006). Yet, (believable) anti-propaganda does not only come from abroad: Hundreds of Arabic YouTubers transformed an IS-video with religious singing into a funny dance clip after its release (Al-Rawi 2016). Moreover, it is possible to focus on preventive measures in combination with (offline) information at schools, universities, or prisons (Saltman and Russell 2014), focusing on social work and vulnerable populations. An effort that combines social and technological intervention uses machine learning to identify grievances which can then be politically and socially addressed, before radicalization turns into violence (Al-Saggaf and Davies 2019).

Automated technology-driven manipulation: the impact of social bots

When fake news and terrorism propaganda lead to the dissemination of cultural violence across social media, technologies such as social bots and large-scale botnets may be misused as multipliers of cultural violence. "A social bot is a computer algorithm that automatically produces content and interacts with humans on social media, trying to emulate and possibly alter their behavior" (Ferrara et al. 2016: 96). Bots' behavior can establish realistic social networks and produce credible content with human-like patterns. They can be classified along their intent and capacity to imitate human behavior (Stieglitz et al. 2017). The use of bots facilitates the targeted spread of particular ideological content and views on social media, disguised as organic, natural human support, creating new socio-technological phenomena.

Account hijacking and astroturfing by social bots

Bots, in addition to human hackers, can be involved in compromising accounts temporarily or entirely through account hijacking. Login details are received via phishing, malware, or cross-site scripting. Often attackers use compromised accounts for further phishing activities to gain access to additional accounts, misusing trust of befriended users (Stein et al. 2011). Hijacked accounts disseminate malware- or phishing-infected websites with the goal of identity theft (Almaatouq et al. 2016). Account hijacking can be used for political purposes, with compromised accounts abusing the trust of legitimate users within the network, who are then more likely to believe misinformation and propaganda (Trang et al. 2015). The added value of accounts taken over increases when profiles are associated with a popular person or organization. Bots are also used to intervene in online discourse through confusion or misinformation, e.g. by associating a hashtag with non-related content for distraction ("misdirection"), or to hide relevant content amidst unrelated content ("smoke screening").

As a further phenomenon, astroturfing describes the imitation of grassroot movements with the aim of feigning a local, social initiative or organization to influence economic or political conditions (Cho et al. 2011). Using bots to suggest wide-spread support, astroturfing is often conducted by political or economic groups. Similar to lobbying, it aims at manipulating public opinion and political decisions by strengthening its own views and discrediting contrary arguments. However, this type of lobbying is inherently extremely intransparent and involves the payment of individuals to set up the structures and campaigns that suggest a legitimate grassroots organization. In this context, bots can be a cost-effective way of simulating wide-spread support. In addition, illegal or gray area content is frequently distributed, e.g. ad fraud, questionable political statements, or defamatory rumors (Wang et al. 2012). Instead of targeting the outcome of a particular policy, the Russian bot firm "Internet Research Agency" (IRA) was used to manipulate voters in the 2016 US election (Diresta et al. 2019). It had set up

accounts across all main social media platforms and used astroturfing to, among other things, encourage and discourage certain voter groups. Research shows that the bot firm co-opted current debates such as the #BlackLivesMatter movement and spread posts both on the extreme spectrum of both the right and left positions, and used existing grievances and distrust to increase fragmentation, societal insecurity, and distrust in the democratic institutions (Stewart et al. 2018).

Algorithmic and crowd-based social bot detection

To counteract social bots, it is first necessary to identify the respective bot accounts. For this purpose, the field of social bot detection has developed various approaches (Ferrara et al. 2016). Social bots may be identified through human engagement or through algorithmic analysis of features and social networks, both complemented by hybrid approaches (see Table 4.3).

To begin with, the approach of crowdsourcing assumes that humans are uniquely able to identify social bot accounts due to their human cognitive skills required to detect human verbal shades of sarcasm, humor, or commitment which cannot be easily imitated by social bots nor recognized by automated bot detection mechanisms. An online platform based on crowdsourcing was thus developed (Wang et al. 2012), with thousands helping to identify bot accounts on Facebook and Renren, a popular Chinese social network. Appling and Briscoe (2017) examine the effectiveness of human identification of social bots and compare it to automated determination of bots. One class of algorithmic detection systems include graph-based approaches which model a respective social network as a finite graph, the participating users constituting vertices and edges illustrating relationships between them. These approaches identify social bots based on analysis of the network topology of the social graph (Yan 2013). Social bots rely on social connections to other accounts for presenting a trustworthy image. It is assumed that bots can only establish a disproportionally small number of social links with legitimate users and are therefore more connected with other bot

Table 4.3 Approaches for social bot detection

Crowdsourcing	Relies on identification of social bots by human actors, assuming humans to be the most able to recognize linguistic nuances like sarcasm, humor, or commitment (Wang et al.2012).
Social graph analysis	Model social networks visually as finite graphs. Nodes illustrate participants of the respective network; edges represent relationships (Yan 2013).
Feature analysis	Identify social bots by determining unique characteristics and behaviors, using machine learning or entropy approaches (Ramalingam and Chinnaiah 2018).
Hybrid approach	Combine different methods, such as adding features to a graph-based approach, to increase the accuracy of social bot detection (Gao et al. 2015).

accounts. This characteristic of close-knit communities of bots within a network is used to identify them through community detection algorithms.

Furthermore, feature-based approaches detect defining characteristics and behaviors of social bot accounts to distinguish them from human users (Ramalingam and Chinnaiah 2018). The examined features are diverse and include the number of followers or tweets, chronological activities of users, content of posts, profile pictures, account names, and friend lists. This group of detection systems may be subclassified into machine learning systems and entropy-based detection systems. Approaches based on machine learning first learn conspicuous training data and subsequently apply a classification algorithm to real data. Entropic-based detection systems do not rely on a prior learning process but identify bots through algorithms searching for anomalies in data sets. Finally, hybrid approaches combine different types of algorithms, for instance, a graph-based approach may be supplemented with features to increase the accuracy of detection (Gao et al. 2018). The simultaneous improvements of both the human-like behavior of bots and of detection systems are leading to an arms race similar to that observed for spam. The experience with spam shows that technical interventions can be powerful, but they must be complemented with social aspects such as knowledge about the mechanisms of abuse to empower users to protect themselves where technical solutions fail.

Conclusion

In this chapter we examined three phenomena that take place in social media where human and (semi-)automatic interventions potentially inflict cultural violence and incite inter-societal conflict through fragmentation. To prevent negative impacts of these phenomena, a variety of different countermeasures are applied which potentially improve cultural peace in social media (see Table 4.4).

In terms of (manual) human interventions, we see that fabricated, misinterpreted, and manipulated content, as well as propaganda and terrorist recruitment,

Table 4.4 Actors and intentions for cultural violence and peace

		Actor	
		Human	*Machine*
Intention	**Malicious interventions**	Fabricated, misinterpreted, manipulated content; propaganda, recruitment	Account hijacking, astroturfing, fake accounts, fake posts, spam
	Countermeasures	Gatekeeping, media literacy, laws, clarification, parody/satire, hacking, counter-narratives	Crowdsourcing platforms, detection algorithms, user assistance tools

may inflict structural or direct violence. Here, countermeasures are similar and include gatekeeping, media literacy and laws, as well as clarification, parody/satire, and hacking. Further research could examine how so far largely neglected actors, such as the crowd and IT-related civil society groups, can contribute to solutions, bringing together IT knowledge and society-level interventions. These can be inspired by established peace interventions from other domains, such as reconciliation. Considering (semi-)automatic machine interventions, we identified account hijacking, astroturfing, fake accounts, fake posts, and spam as potentials for cultural violence exacerbating existing divides and eroding trust in legitimate protest and institutions. Respective countermeasures contain detection algorithms and crowdsourcing for malicious content. Experiences in countering spam show the power of technical arms races, but also spammers' adaptability in using sophisticated social engineering to deceive detection mechanisms and humans by exploiting trust detection mechanisms. Similarly, the Russion bot firm IRA has adapted its strategy to feigning affiliation with established, trusted institutions (Wired 2020). Technical arms races can thus be powerful, but never all-encompassing, leaving the necessity for social interventions. Hybrid forms of intervention include solutions that, without outright censoring posts, limit the visibility or spreading speed of harmful content, or provide technical assistance for users to better judge the truthworthiness of online information, or can identify social media users at risk of radicalization. However, as long as legitimate grievances exist, actors such as terrorists will be able to co-opt these grievances and resistance. Therefore, organizations such as ICT4Peace use communication technology to address community grievances at the root level, helping overcome fragmentation and societal insecurity.

This limit of technical interventions also applies to disinformation and terrorist propaganda: While deletion and flagging of false content are possible, this raises questions about the authority over defining the truth and dangers of censorship. The dominant technical interventions are not addressing the root causes that make people gullible to disinformation and even lead them to potentially sign away their future to join extremist groups. This also raises new questions about the definition of victims and perpetrators of online structural violence: Are people who spread misinformation and propaganda perpetrators of societal fragmentation and structural violence, or victims of a society that has left them with low media literacy and the feeling of being alienated by the society they live in? Similar to fake news, it is difficult to differentiate legitimate protest movements from those instigated by politically and economically motivated bot firms that specialize in feigning public support for radical or partisan opinions. As is in many countries required to start a new political party, for sensitive topics with the potential to fragment society, new organizations could be required to proof their legitimacy through referral by an organization that is trusted by that community. Though a difficult task, such measures may be necessary to save the legitimacy of grassroot protest in the long run. The frame of structural and cultural violence can help to identify issues and populations that are particularly vulnerable to social media incitement of resentment, or topics and corporations that may profitably use disinformation and social bots, suggesting a need for societal interventions.

A promising first step is the social media analytics, which can be used to better understand the social side of social media abuse, e.g. by making situational assessments of specific discourses and events (Kaufhold et al. 2020a), including the identification of fake news or hate speech as potential instances of cultural violence using (supervised) machine learning approaches (Kaufhold et al. 2020b). As an intermediary, technical tools can be developed to flag false content and provide transparency over actors and organizations that fuel the extremes and follow partisan interests. This will require identifying the actors and incentive structures that motivate disinformation and the buying of social bot systems as well as addressing the societal structures, mainly mistrust and grievances, which allow malicious interventions to take devastating effects.

Further research should overcome the limitation of this explorative contribution by first including more socio-technological technological transformations seen in social media that can contribute to structural violence. As this chapter focused on the cultural areas of ideology, a more comprehensive examination should further address issues such as cultural diversity, religion, and economy as factors for cultural violence in social media, e.g. through an apposite mapping to Galtung's (2007) cultural areas of religion, ideology, language, art, and empirical and formal science.

Acknowledgments

Parts of this chapter are based on our previous work (Kaufhold and Reuter 2019). This work has been co-funded by the German Federal Ministry of Education and Research (BMBF) and the Hessen State Ministry for Higher Education, Research and Arts (HMKW) within the SecUrban mission of the National Research Center for Applied Cybersecurity ATHENE and by the LOEWE initiative (Hesse, Germany) within the emergenCITY center.

References

All links checked on August 20, 2021.

Al-Rawi, A. (2016). Anti-ISIS Humor: Cultural Resistance of Radical Ideology. *Politics, Religion and Ideology*, 7689(May): 1–17.

Al-Saggaf, Y. (2016). Understanding Online Radicalisation Using Data Science. *International Journal of Cyber Warfare and Terrorism (IJCWT)*, 6(4): 13–27.

Al-Saggaf, Y. and Davies, A. (2019). Understanding the Expression of Grievances in the Arabic Twitter-Sphere Using Machine Learning. *Journal of Criminological Research, Policy and Practice*, 5(2): 108–119.

Allcott, H. and Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2): 211–236.

Almaatouq, A., Shmueli, E., Nouh, M., Alabdulkareem, A., Singh, V. K., Alsaleh, M., … Pentland, A. (2016). If it Looks Like a Spammer and Behaves Like a Spammer, it Must Be a Spammer: Analysis and Detection of Microblogging Spam Accounts. *International Journal of Information Security*, 15(5): 475–491.

Appling, D. S. and Briscoe, E. J. (2017). The Perception of Social Bots by Human and Machine. In: *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference*, pp. 20–25.

Awan, I. (2017). Cyber-Extremism: Isis and the Power of Social Media. *Society*, 54(2): 138–149.

Becker, B. W. (2016). The Librarian's Information War. *Behavioral and Social Sciences Librarian*, 35(4): 188–191.

Booth, K. (2007). *Theory of World Security* (Vol. 105). Cambridge: Cambridge University Press.

Campbell, P. J., MacKinnon, A. S. and Stevens, C. (2010). *An Introduction to Global Studies*. Hoboken, NJ: Wiley-Blackwell.

Cheong, M. and Lee, V. C. S. (2011). A Microblogging-Based Approach to Terrorism Informatics: Exploration and Chronicling Civilian Sentiment and Response to Terrorism Events via Twitter. *Information Systems Frontiers*, 13(1): 45–59.

Cho, C. H., Martens, M. L., Kim, H. and Rodrigue, M. (2011). Astroturfing Global Warming: It Isn't Always Greener on the Other Side of the Fence. *Journal of Business Ethics*, 104(4): 571–587.

Cooke, N. A. (2017). Posttruth, Truthiness, and Alternative Facts: Information Behavior and Critical Information Consumption for a New Age. *Library Quarterly: Information, Community, Policy*, 87(3): 211–221.

DeNardis, L. and Hackl, A. M. (2015). Internet Governance by Social Media Platforms. *Telecommunications Policy*, 39(9): 761–770.

Diresta, R., Shaffer, K., Ruppel, B., Matney, R., Fox, R., Albright, J. and Johnson, B. (2019). *The Tactics and Tropes of the Internet Research Agency*. Report for the United States Senate Select Committee on Intelligence. Retrieved from: https:/digitalcommons.unl.edu/senatedocs/2/.

Ferrara, E., Varol, O., Davis, C., Menczer, F. and Flammini, A. (2016). The Rise of Social Bots. *Communications of the ACM*, 59(7): 96–104. Retrieved February 25 2021, from: http://dl.acm.org/citation.cfm?id=2818717.

Freedman, L. (2006). *The Transformation of Strategic Affairs*. Abingdon: Routledge.

Galtung, J. (2007). *Frieden mit friedlichen Mitteln. Friede und Konflikt, Entwicklung und Kultur*. Münster: Agenda Verlag.

Gao, P., Gong, N. Z., Kulkarni, S., Thomas, K. and Mittal, P. (2018). SybilFrame: A Defense-in-Depth Framework for Structure-Based Sybil Detection. In: *Computing Research Repository*. Retrieved February 25 2021, from: http://arxiv.org/abs/1503.02985.

Gartenstein-Ross, D. (2015). Social Media in the Next Evolution of Terrorist Recruitment. *Hearing before the Senate Committee on Homeland Security and Governmental Affairs*, Foundation for Defense of Democracies, 1–11.

Gates, S. and Podder, S. (2015). Social Media, Recruitment, Allegiance and the Islamic State. *Perspectives on Terrorism*, 9(4): 107–116.

Graves, L. (2018). Boundaries Not Drawn: Mapping the Institutional Roots of the Global Fact-Checking Movement. *Journalism Studies*, 19(5): 613–631.

Hancock, J. T., Curry, L. E., Goorha, S. and Woodworth, M. (2008). On Lying and Being Lied To: A Linguistic Analysis of Deception in Computer-Mediated Communication. *Discourse Processes*, 45(1): 1–23.

Hartwig, K. and Reuter, C. (2019). TrustyTweet: An Indicator-Based Browser-Plugin to Assist Users in Dealing with Fake News on Twitter. In: *Proceedings of the International Conference on Wirtschaftsinformatik (WI)*. Siegen.

Haunschild, J., Kaufhold, M.-A. and Reuter, C. (2020). Sticking with Landlines? Citizens' and Police Social Media Use and Expectation during Emergencies. In: *Proceedings of the International Conference on Wirtschaftsinformatik (WI)*. Potsdam, Germany: AIS Electronic Library (AISel).

Hern, A. (2020, April 7). WhatsApp to Impose New Limit on Forwarding to Fight Fake News. *The Guardian*. Retrieved from: https://www.theguardian.com/technology/2020/apr/07/whatsapp-to-impose-new-limit-on-forwarding-to-fight-fake-news.

Humprecht, E. (2019). Where 'Fake News' Flourishes: A Comparison across Four Western Democracies. *Information, Communication and Society*, 22(13): 1973–1988.

Hussain, G. and Saltman, E. M. (2014). Jihad Trending: A Comprehensive Analysis of Online Extremism and How to Counter It. Report from the Quilliam Foundation. Retrieved from: https://preventviolentextremism.info/jihad-trending-comprehensive-analysis-online-extremism-and-how-counter-it.

Jeberson, W. and Sharma, L. (2015). Survey on Counter Web Terrorism. *COMPUSOFT: An International Journal of Advanced Computer Technology*, 4(5): 1744–1747.

Kaplan, A. M. and Haenlein, M. (2010). Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons*, 53(1): 59–68.

Kaufhold, M.-A., Bayer, M. and Reuter, C. (2020b). Rapid Relevance Classification of Social Media Posts in Disasters and Emergencies: A System and Evaluation Featuring Active, Incremental and Online Learning. *Information Processing and Management*, 57(1): 1–32.

Kaufhold, M.-A. and Reuter, C. (2019). Cultural Violence and Peace in Social Media. In: C. Reuter (ed.). *Information Technology for Peace and Security - IT-Applications and Infrastructures in Conflicts, Crises, War, and Peace*. Wiesbaden: Springer Vieweg, pp. 361–381.

Kaufhold, M.-A., Rupp, N., Reuter, C. and Habdank, M. (2020a). Mitigating Information Overload in Social Media during Conflicts and Crises: Design and Evaluation of a Cross-Platform Alerting System. *Behaviour and Information Technology (BIT)*, 39(3): 319–342.

Klausen, J., Barbieri, E. T., Reichlin-Melnick, A. and Zelin, A. Y. (2012). The YouTube Jihadists: A Social Network Analysis of Al-Muhajiroun's Propaganda Campaign. *Perspectives on Terrorism*, 6(1): 36–53.

McCarthy, T. (2017, October 14). How Russia Used Social Media to Divide Americans. *The Guardian*. Retrieved from: https://www.theguardian.com/us-news/2017/oct/14/russia-us-politics-social-media-facebook.

Mihailidis, P. and Viotty, S. (2017). Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in "Post-Fact" Society. *American Behavioral Scientist*, 61(4): 441–454.

Mostrous, A., Bridge, M. and Katie, G. (2017, November 15). Russia Used Twitter Bots and Trolls 'to Disrupt' Brexit Vote. *The Times*. Retrieved from: https://www.thetimes.co.uk/article/russia-used-web-posts-to-disrupt-brexit-vote-h9nv5zg6c.

Müller, P. and Denner, N. (2017). Was tun gegen "Fake News"? Eine Analyse anhand der Entstehungsbedingungen und Wirkweisen gezielter Falschmeldungen im Internet. Retrieved from: https://madoc.bib.uni-mannheim.de/50564/.

Neer, T. and O'Toole, M. E. (2014). The Violence of the Islamic State of Syria (ISIS): A Behavioral Perspective. *Violence and Gender*, 1(4): 145–156.

Neue Wege des Lernens e.V. (2017). Fake News Check. Retrieved Februrary 25 2021, from: https://www.neue-wege-des-lernens.de/apps/.

Newman, M. L., Pennebaker, J. W., Berry, D. S. and Richards, J. M. (2003). Lying Words: Predicting Deception from Linguistic Styles. *Society for Personality and Social Psychology*, 29(5): 665–675.

News Afrika. (2016, September). Fake News Fuels Xeniphobic Tensions in South Africa.

Ramalingam, D. and Chinnaiah, V. (2018). Fake Profile Detection Techniques in Large-Scale Online Social Networks: A Comprehensive Review. *Computers and Electrical Engineering*, 65: 165–177.

Reuter, C. (2019). Information Technology for Peace and Security – Introduction and Overview. In: C. Reuter (ed.). *Information Technology for Peace and Security*. Wiesbaden: Springer, pp. 3–9.

Reuter, C. (2020). Towards IT Peace Research: Challenges on the Interception of Peace and Conflict Research and Computer Science. *S+F Sicherheit Und Frieden / Peace and Security*, 38(1): 10–16.

Reuter, C., Hartwig, K., Kirchner, J. and Schlegel, N. (2019). Fake News Perception in Germany: A Representative Study of People's Attitudes and Approaches to Counteract Disinformation. In: *Proceedings of the International Conference on Wirtschaftsinformatik (WI)*. Siegen, Germany: AIS, pp. 1069–1083.

Reuter, C. and Kaufhold, M.-A. (2018). Fifteen Years of Social Media in Emergencies: A Retrospective Review and Future Directions for Crisis Informatics. *Journal of Contingencies and Crisis Management (JCCM)*, 26(1): 41–57.

Reuter, C., Pätsch, K. and Runft, E. (2017). IT for Peace? Fighting against Terrorism in Social Media – An Explorative Twitter Study. *I-Com: Journal of Interactive Media*, 16(2): 181–193.

Robinson, T., Callahan, C., Boyle, K., Rivera, E. and Cho, J. K. (2017). I ♥ FB: A Q-Methodology Analysis of Why People 'Like'' Facebook.' *International Journal of Virtual Communities and Social Networking (IJVCSN)*, 9(2): 46–61.

Rubin, V. L. (2019). Disinformation and Misinformation Triangle: A Conceptual Model for "Fake News" Epidemic, Causal Factors and Interventions. *Journal of Documentation*, 75(5): 1013–1034.

Saltman, E. M. and Russell, J. (2014). *White Paper – The Role of Prevent in Countering Online Extremism*. London: Quilliam Foundation.

Stein, T., Chen, E. and Mangla, K. (2011). Facebook Immune System. *Proceedings of the 4th Workshop on Social Network Systems*, 5: 1–8.

Stewart, L. G., Arif, A. and Starbird, K. (2018). Examining Trolls and Polarization with a Retweet Network. *Proceedings of WSDM Workshop on Misinformation and Misbehavior Mining on the Web (MIS2)*, 6.

Stieglitz, S., Brachten, F., Ross, B. and Jung, A.-K. (2017). Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts. *Proceedings of the Australasian Conference on Information Systems*, 1–11.

Stieglitz, S., Mirbabaie, M., Ross, B. and Neuberger, C. (2018a). Social Media Analytics – Challenges in Topic Discovery, Data Collection, and Data Preparation. *International Journal of Information Management*, 39: 156–168.

Stieglitz, S., Mirbabaie, M. and Milde, M. (2018b). Social Positions and Collective Sense-Making in Crisis Communication. *International Journal of Human-Computer Interaction*, 34(4): 328–355.

Sutton, J., Palen, L. and Shklovski, I. (2008). Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires. In: F. Friedrich and B. Van de Walle (eds). *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)*. Washington, DC, pp. 624–632.

Torok, R. (2015). ISIS and the Institution of Online Terrorist Recruitment. Retrieved February 25 2021, from: https://www.mei.edu/publications/isis-and-institution-online-terrorist-recruitment.

Trang, D., Johansson, F. and Rosell, M. (2015). Evaluating Algorithms for Detection of Compromised Social Media User Accounts. *Proceedings - 2nd European Network Intelligence Conference, ENIC 2015*, 75–82.

Viviani, M. and Pasi, G. (2017). Credibility in Social Media: Opinions, News, and Health Information—A Survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(5): e1209.

Waever, O. (1993). Societal Security: The Concept. In: O. Waever, et al. (eds). *Identity, Migration and the New Security Agenda in Europe*. London: Pinter, pp. 17–40.

Wang, G., Wilson, C., Zhao, X., Zhu, Y., Mohanlal, M., Zheng, H. and Zhao, B. Y. (2012). Serf and Turf: Crowdturfing for Fun and Profit. *Arxiv Preprint ArXiv:1111.5654*, 10.

Weimann, G. (2016). The Emerging Role of Social Media in the Recruitment of Foreign Fighters. In: A. de Guttry, F. Capone, and C. Paulussen (eds). *Foreign Fighters under International Law and Beyond*. The Hague: T.M.C. Asser Press, pp. 77–95.

Weimann, G. and Jost, J. (2015). Neuer Terrorismus und Neue Medien. *Zeitschrift Für Außen- Und Sicherheitspolitik*, 8(3): 369–388.

Wired (2020, March 5). Russia Is Learning How to Bypass Facebook's Disinfo Defenses. *Wired*. Retrieved from: https://www.wired.com/story/russia-ira-bypass-facebook-disinfo-defenses/.

Wohn, D. Y., Fiesler, C., Hemphill, L., De Choudhury, M. and Matias, J. N. (2017). How to Handle Online Risks? Discussing Content Curation and Moderation in Social Media. In: *CHI 2017 Extended Abstracts*, pp. 1271–1276.

Wulf, V., Aal, K., Ktesh, I. A., Atam, M., Schubert, K., Yerousis, G. P., … Bank, W. (2013). Fighting against the Wall: Social Media Use by Political Activists in a Palestinian Village. In: *Proceedings of the Conference on Human Factors in Computing Systems* (*CHI*). Paris: ACM, pp. 1979–1988.

Yan, G. (2013). Peri-Watchdog: Hunting for Hidden Botnets in the Periphery of Online Social Networks. *Computer Networks*, 57(2): 540–555. https://doi.org/10.1016/j.comnet.2012.07.016.

Ziegele, M., Breiner, T. and Quiring, O. (2014). What Creates Interactivity in Online News Discussions? An Exploratory Analysis of Discussion Factors in User Comments on News Items. *Journal of Communication*, 64(6): 1111–1138.