# Hate and harm

Frischlich, Lena

Erstveröffentlichung / Primary Publication
Sammelwerksbeitrag / collection article

Digital
Communication
Research.de

**Recommended citation:** Frischlich, L. (2023). Hate and harm. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), *Challenges and perspectives of hate speech research* (pp. 165–183). Digital Communication Research. https://doi.org/10.48541/dcr.v12.10

**Abstract:** From a psychological point of view, hate speech can be conceptualized as harmful intergroup communication. In contrast to other forms of incivility, hate speech is directed toward individuals because of their (perceived) social identity. This explains why the harm of hate speech can extend to entire social groups and societies. Hate speech therefore cannot be separated from pre-existing power structures and resource inequalities, as its harm is particularly severe when coping resources are already deprived. Psychological research on the perpetrators of hate speech links hate speech to a lack of empathy and the acceptance of, or even desire for social inequalities. In summary, hate speech jars the norms of democratic discourses by denying fellow humans basic respect and violating the democratic minimal consent of human equality. Overall, the chapter demonstrates the usefulness of a (social) psychological perspective on the harms of hate speech for both researchers and practitioners.

DOI 10.48541/dcr.v12.10

*Lena Frischlich*

# Hate and Harm

## 1 Hate speech as "going against" a social identity

In this chapter, I understand hate speech as a specific form of *incivility*, a communication that violates norms (e.g., Kenski et al., 2020; Mutz, 2015; Papacharissi, 2004). Incivility is thereby a "notoriously difficult concept to define" (Coe et al., 2014, p. 660), not least because different perspectives have to be taken into account: the perspective of the perpetrator, the perspective of the attacked, and the perspective of the observer (O'Sullivan & Flanagin, 2003). Further, it is often unclear which norms exactly have been violated, and while explicit vulgar insults are relatively consistently rated as uncivil, more subtle norm transgressions, such as dehumanizing metaphors, or formal forms of incivility (e.g., the use of multiple exclamation marks or grammatically wrong expressions) are less agreed upon (for a discussion, see Chen et al., 2019; Bormann & Ziegele in this volume).

In the following, I focus on hate speech as a subtype of communication "going against" a target (Gagliardone et al., 2016, p. 6). Following Gagliardone et al. (2016), two types of targets can be distinguished in this context, although they may overlap (see also Rossini, 2020). The first type, which Gagliardone et al. (2016) label *offensive speech*, is directed toward individuals and is often studied under labels such as (*cyber-)bullying* (e.g., Festl, 2016) or *trolling* (for a comprehensive overview, see Phillips, 2012). Offensive speech violates interpersonal norms of

politeness (Mutz, 2015, p. 6), for example, by using swear words, and insults, or by mocking the target. The current chapter focuses on the second type of incivility—*hate speech*, which is directed against individuals because of their collective or social identity and reflects a biased attitude toward the targeted group rather than a personal dislike (Silva et al., 2016, p. 3). Although hate speech is not necessarily formally uncivil (i.e., detectable by exclamation marks) or offensive (e.g., using well-known swear words), it can have specific severe effects on those attacked and their social context, wherefore it is sometimes described as "harmful speech" (Bilewicz & Soral, 2020, p. 2).

In the following, I describe these harms for both the individual target and its social context from a predominantly psychological point of view. First, I will show how a social psychological perspective allows for describing the harmful "fallout" of hate speech compared to other types of incivility. Notably, this does not imply that offensive speech, such as cyberbullying, is harmless; however, as I will argue in the following, the fundamentally social nature of hate speech is unique and thus should be treated as such. Second, I will show how individual characteristics, such as personality, attitudes, and emotions, shape the spread of hate speech on the individual micro-level. I will close the chapter by arguing that the suggested perspective allows us to consider both the individual and the social levels when examining the harms of hate speech.

## 2 Hate speech is directed toward people's social identities

The fallout of hate speech can be explained by social identity and self-categorization mechanics. *Social identity theory* (Tajfel & Turner, 1979) and *self-categorization theory* (Turner et al., 1987) postulate that people not only own a *personal identity* that distinguishes them from others and makes them unique but also multiple *social identities* resulting from their social roles and memberships in social groups or categories. The *ingroups* to which people belong are perceptually and functionally distinct from *outgroups*, that is, groups or categories people do not belong to. The more people identify with their social group, the more they think, feel, and act on behalf of that group. For instance, people can feel nostalgic or guilty on behalf of their nation (Martinovic et al., 2017) and sad or joyful because of their sports teams' performances (for an overview of intergroup

emotions, see Smith & Mackie, 2015). Even random categorizations in artificial groups motivate a distinct treatment of ingroup versus outgroup members and change the neural processing of ingroup and outgroup members (Brewer, 1979; Crocker & Schwartz, 1985; Ratner & Amodio, 2013).

Due to the central role of social groups in ones' identity, people are motivated to see their ingroup(s) in a positive light and to perceive them as positively distinct from outgroups (Tajfel & Turner, 1979). Such positive ingroups are an important factor in psychosocial well-being (Haslam et al., 2016; Jetten et al., 2012) and a pillar of individuals' resilience in light of hardships (Muldoon et al., 2019). To preserve a positive group image, people are biased to perceive ingroup compared to outgroup members as being more trustworthy (Yamagishi & Kiyonari, 2000) and less flawed (Koval et al., 2012). In times of uncertainty (Hogg et al., 2007), when people feel socially ostracized (Pfundmair & Wetherell, 2019), or are reminded of their inevitable decay (Frischlich et al., 2015), ingroup biases can even extend to tolerate extremist and violent ingroup members more than outgroup members, as social identities can help individuals cope with these kinds of existential threats (Jonas et al., 2014).

People's social identities differ in their stability (or variability). While some groups are relatively easy to change through re-categorization processes (e.g., when changing one's employer), other social categories are more difficult to change and are repeatedly ascribed to individuals even without their intervention. This is especially true for membership in disadvantaged or visually marked groups (such as gender or ethnicity). Hate speech primarily attacks such stable identities (Bilewicz & Soral, 2020), often relying on century-old stereotypes and longstanding prejudice. To understand the damage caused by hate speech, it is therefore crucial to consider the perspective of socially marginalized groups (Dieckmann et al., 2018) and to understand hate speech as "harmful language" (Leets & Giles, 1999).

From this perspective, hate speech is more closely related to *hate crimes* (Walters et al., 2016) than to impoliteness. *Hate crime*, as a legal category in the UK, is defined as "any crime or incident where the perpetrator's hostility or prejudice against an identifiable group of people is a factor in determining who is victimised" (College of Policing, 2020). Typical hate crimes are incidents of discrimination or even violence against people who are interpreted as members of a certain social group, such as a religious or sexual minority.

Hate crimes are often a "stranger-danger" where perpetrator and victim are unknown to each other though physically (or virtually) existing in the same space (Mason, 2005). Hate crimes are often driven by the motivation to preserve the perceived "natural superiority"[1] of the perpetrator (Perry & Alvi, 2012). Through stereotypes and prejudices, hate speech is embedded in a specific socio-cultural system, with its specific power relationships and specific histories of intergroup conflicts. What makes hate speech specifically harmful speech are the resources for coping with the threat of hate speech, which are unequally distributed among the benefiters of human history and those struggling for their place at the table. A cross-country survey showed that hate speech varies in both reported frequency and attacked targets along socio-cultural lines and long-term narratives in a given context (Reichelmann et al., 2020). For instance, female (compared to male) journalists and politicians are disproportionally often overflooded with hate directed toward them (for media reports, see Carter, 2021; Gardiner et al., 2016).

## 3 Putting the harm in hate speech: Effects on victims and observers

Hate speech (and other types of incivility not in focal attention here) can have severe negative effects. For instance, a large German study (Geschke et al., 2019) found that among those who had experienced hate speech, only one-third reported no personal consequences, another third reported emotional distress, and 17% reported depression as a consequence of the attacks. The same study also showed that 46% of those who had experienced hate speech refrained from online discussions at least sometimes to avoid attacks, and 51% did not speak about their political orientation online. Similar silencing effects were reported by indigenous Australians in a study by Gelber and McNamara (2015).

Computational simulation studies warn that hate speech can over time erode norms for civil interactions via desensitization (Soral et al., 2018), leading to a "hate speech epidemic," as Bilewicz and Soral (2020, p. 3) termed it. Another computational simulation indicates that even subtle discrimination by a societal

1 The author distances herself from the idea of a natural order in which some human beings or social groups are supreme to others.

majority can cement prejudiced intergroup relationships over time by eroding trust in outgroups (Uhlmann et al., 2018). Hate speech thus diminishes social trust (Näsi et al., 2015), potentially contributing to "spirals of distrust" (Frischlich & Humprecht, 2021, p. 4) and endangering societal cohesion. Further, an experimental study by Hsue et al. (2015) showed that reading uncivil comments targeting minorities motivated more negative attitudes toward the targeted group. Once someone is classified as an outgroup member, people become less able to detect that person's pain (Ma et al., 2011), which in turn makes it less likely that they will respond with empathy in future interactions (Timmers et al., 2018). Not surprising, hate speech can also impact helping behavior. For instance, Ziegele et al. (2018) showed that reading hate speech reduced readers' pro-social intentions towards the attacked group. In summary, hate speech does jar the foundations of the democratic contract (Papacharissi, 2004) by denying human equality.

## 4 Interindividual differences and motivations for hate speech

Not all people are equally likely to spread hate. Interindividual differences in personality traits, attitudes, and emotions are all associated with a different likelihood of becoming a hate speech perpetrator. With regards to personality traits, different studies have indicated an association between incivility and the so-called *dark tetrad* (Međedović & Petrović, 2015). The dark tetrad describes four sub-clinical forms of offensive personalities, the so-called dark triad of *narcissism*, *Machiavellianism*, and *psychopathy* (Paulhus & Williams, 2002) plus everyday *sadism* (Buckels et al., 2013). Narcissists are characterized by grandiosity perceptions (Paulhus & Williams, 2002)—although their self-esteem can be brittle at the same time (Miller et al., 2017)—and social manipulativeness (Raskin & Hall, 1981). Machiavellianism involves manipulative and cold behavior, psychopathy describes impulsive and thrill-seeking behavior by individuals showing reduced levels of anxiety (Paulhus & Williams, 2002), and sadism describes the enjoyment of cruelty and others' harm (Buckels et al., 2013). People scoring higher on the dark triad are more likely to admit to engaging in uncivil (Frischlich et al., 2021) and aggressive online behavior (Buckels et al., 2014; Kurek et al., 2019), although the direct link to hate speech is unclear (Koban et al., 2018). One component that could link the dark triad and uncivil and hateful speech is empathy. People scoring higher on the

dark tetrad tend to have deficits in their empathy abilities (e.g., see Heym et al., 2021), and empathic people engage less in trolling (March, 2019) and hate speech dissemination (Bilewicz & Soral, 2020).

Hate speech is also associated with people's generalized attitudes—that is, their ideological evaluation frameworks across situations, time, and/or persons. Two of these generalized attitudes are particularly relevant with regards to discrimination and prejudice (Duckitt et al., 2002; Duckitt & Sibley, 2010): Individual's level of *right-wing authoritarianism* (RWA*)* and their *social dominance orientation* (SDO).

Right-wing authoritarianism reflects a general psychological tendency to submit to authorities, support conventional values, and punish those who transgress the rules (Altemeyer, 1988; Duckitt, 2015). Social dominance orientation (Sidanius & Pratto, 1999) reflects a preference for group-based inequalities in society (Ho et al., 2015), either such that powerful groups should forcefully oppress lower status groups or in a more subtle hierarchy-enhancing way, for instance, by endorsing policies that stabilize group-based inequalities.

Bilewicz et al. (2017) showed that individuals with a larger social dominance orientation were particularly likely to consider hate speech to be acceptable—mirroring the idea that hate crimes are often an attempt to restore the presumably "natural order" (Perry & Alvi, 2012). Authoritarians, by contrast, were eager to prohibit hate speech expressions in a study by Bilewicz et al. (2017)—likely because the norm deviant character of hate speech conflicts with authoritarians' preference for adherence to established norms. Of note, our own research found that high authoritarians are more open to hateful right-wing extremist propaganda (Frischlich et al., 2015; Rieger et al., 2013, 2017), suggesting that further research into the interplay between authoritarianism, norm perceptions, and hate speech is needed.

Research also points toward ideological *asymmetries* regarding the association between political attitudes and hate speech. Survey data from the US showed that conservatives, compared to liberals, evaluated hate speech as being less disturbing (Costello et al., 2019), and research from Germany showed that supporters of the right-wing populist *Alternative for Germany* (AfD) were particularly active in supporting hate speech in online media (Frischlich et al., 2021; Kreißel et al., 2018). Hate speech is also a prominent communication style in alt-right online circuits (Marwick & Lewis, 2017). Although ideological asymmetries between those leaning toward the right versus toward the left have been demonstrated

for a wide array of human characteristics (Jost, 2017), one aspect could be particularly relevant regarding hate speech: differences in moral evaluations across the political spectrum. Based on *moral foundation theory* (Graham et al., 2011; Haidt & Joseph, 2008), humans have an intuitive ethic that has evolved to fulfil specific adaptive needs and whose violation is disregarded. Five of these moral foundations are particularly well established: The intuition of (a) *care*, evolved through the adaptive challenge to care for humans' vulnerable offspring but also the larger tribe, as research on *parochial altruism* suggests (Bernhard et al., 2006; Choi & Bowles, 2007). Following Graham et al. (2013), care is assumed to be related to empathic responses to others' suffering, and its violation is described as *harm*. The other four dimensions are (b) *fairness/cheating* (evolved as humans' response to interaction partners' lack of reciprocity in interactions); (c) *loyalty/betrayal* (related to humans' devotedness to their ingroup or tribe); (d) *authority/subversion* (reflecting the social order within the tribe); and (e) sanctity or *purity* versus *degradation*, reflecting disgust toward devaluated behaviors.

Individuals with different political orientations differ with regard to the relevance they ascribe to the violation and upholding of these five moral foundations. While liberals value individualizing moral intuitions of care and fairness particularly highly, conservatives also uphold biding moral intuitions of loyalty, authority, and purity (Graham et al., 2009). This difference is even reflected in the extreme case of terrorists' self-explanations: Hahn et al. (2019) showed that right-wing terrorists and religious fundamentalists justified their deeds more with binding moral values, whereas left-wing terrorists and those acting for animal rights relied more often on individualizing moral foundations. People who highly value the individual moral foundations of care and fairness are also more likely to report hate speech, whereas those valuing loyalty, authority, and purity are less likely to do so (Wilhelm et al., 2020).

Hate speech and other forms of incivility are also associated with different negative emotions. Following the *appraisal theories of emotion* (for a comprehensive overview, see Scherer, 2005), emotions can be understood as a process that ranges from (1) the cognitive appraisal of a specific internal or external stimulus over, (2) the psychophysiological response to that stimulus, (3) a verbal or non-verbal response, and (4) a motivational activation specific to the given emotion, up to (5) a distinct feeling such as joy, fear, or awe (e.g., Scherer, 1987). For instance, evaluating a situation as unjust and someone guilty of this injustice triggers anger (Nabi, 2002). Anger is associated with increased blood pressure (Lindquist et al., 2016),

the motivation to change the anger-inducing condition, and subjective feelings of being annoyed or in rage (Harmon-Jones & Harmon-Jones, 2016).

Research on political incivility has shown that people who respond with anger to incivilities write more uncivil comments in response (Gervais, 2017, 2019). Although emotions of anger partially overlap with hate, Bilewicz et al. (2017) argued that most of the phenomena labeled as hate speech are actually driven by emotions of disgust and contempt rather than anger and hate. Both hate and contempt increase the willingness to harm the target; however, contempt is associated with perceiving the target as inferior, whereas hate often targets seemingly powerful targets. Consequentially, contempt is often a better predictor of hate speech than hate or anger (for an overview, see Bilewicz et al., 2017).

It is likely that the different personality, attitudinal, and emotional variables lead to different types of haters, as a study by Erjavec and Kovačič (2012) shows. Based on a series of interviews, the authors identified four distinct types. The first two types tap into the social identity and social dominance components of hate speech: (1) "the soldier" (p. 909), who is described as an active member of a political party or (nationalist) organization who engages in organized hate speech as part of a "contemporary war" (p. 909), and (2) "the believer[s]" (p. 911), who has a similar worldview but lacks the organizational affiliation. The third type, (3) the "player," is someone who derives pleasure from disturbing the discourse, implying that dark personality traits might play a role here. Lastly, (4), the "watch-dog[s]" uses hate speech to draw attention to what is perceived to be unjust, seemingly underlining the role of morality.

## 5 Equality and empathy against hate and harm

In summary, hate speech can be conceptualized as harmful intergroup communication. This harm is particularly severe when coping resources are low, for instance, when stable social identities such as gender or ethnicity are under attack and/or for those belonging to socially underrepresented and marginalized groups. The suggested social psychological perspective provides a solid social scientific base for legal-rooted terms, such as hate crime or hate speech, and allows for describing the fallout of this specific type of attacks. Hate speech not only harms those directly attacked but also the entire social group; it jars social trust

and contributes to lasting social frictions by fueling prejudice, reducing prosocial behavior, and endangering empathy for fellow humans.

Although civility very often lies "in the eye of the beholder" (Herbst, 2010, p. 3), hate speech in the narrower sense described in this chapter is bound to specific socio-cultural spaces and norms, often reflecting traditional stereotypes and power imbalances in a society. The tendency of hate speech to attack those already deprived of coping resources, and the fact that these attacks fall out toward larger social groups, underlines the specific harms of hate speech. Although offensive speech can also be harmful, hate speech denies fellow humans their right to equity, thus crossing the borders of "reasonable disagreement" in a normative sense (Nussbaum, 2011, p. 4). Consequently, counter-measures such as the moderation of online content—which always need to strike the fragile balance between the freedom of expression and the preservation of a reasonable democratic discourse—might not only refer to individual harms when it comes to deciding about hate speech but can, psychologically speaking, also take the broader intergroup and societal context into account (for an excellent fusion of legal and social science perspectives, see Leets & Giles, 1999).

The psychology-rooted perspective of this chapter also demonstrates that not all people are equally likely to engage in hate speech. Dark personality traits characterized by empathy deficits, binding moral foundations that weigh loyalty, authority, and purity at least as highly as caring for others and fairness, convictions that society is, and should be, composed of inequal groups with different rights, and emotions of contempt all are associated with a larger propensity to spread hate speech.

This observation has meaningful implications for prevention: fostering empathy (Miklikowska, 2017) and creating unified super-ordinated social identities within a society (Dovidio et al., 2007) or with all humankind (McFarland, 2017) can help reduce stereotypes and prejudice. Social-dominance orientation can be a barrier to such endeavors (Sidanius et al., 2013); thus, it is also necessary to address the larger context in which social dominance orientation thrives. For instance, meta-analyses have shown that social dominance orientation is larger among individuals perceiving the world as a competitive struggle (Perry et al., 2013) and living in more hierarchical societies (Fischer et al., 2012). Taking the psychological factors on the micro-level of the individual hater as well as on the

meso-level of social groups into account can help to understand the roots and harms of hate speech, and find new ways to heal them.

*Lena Frischlich* is a junior research group leader at the University of Münster, Germany. https://orcid.org/0000-0001-5039-5301

## References

Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism.* Jossey-Bass.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, *442*(7105), 912–915. https://doi.org/10.1038/nature04981

Bilewicz, M., Kamińska, O. K., Winiewski, M., & Soral, W. (2017). From disgust to contempt-speech: The nature of contempt on the map of prejudicial emotions. *Behavioral and Brain Sciences*, *40*, 1–63. https://doi.org/10/gg5dxr

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, *41*(1), 3–33. https://doi.org/10.1111/pops.12670

Bilewicz, M., Soral, W., Marchlewska, M., & Winiewski, M. (2017). When authoritarians confront prejudice. Differential effects of SDO and RWA on support for hate-speech prohibition. *Political Psychology*, *38*(1), 87–99. https://doi.org/10/bfk8

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, *86*(2), 307–324. https://doi.org/10/dg5fn8

Buckels, E. E., Jones, D. N., & Paulhus, D. L. (2013). Behavioral confirmation of everyday sadism. *Psychological Science*, *24*(11), 2201–2209. https://doi.org/10.1177/0956797613490749

Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, *67*, 97–102. https://doi.org/10/f58bzw

Carter, L. (2021, March 13). Finland's women-led government targeted by online harassment. *Politico*. https://www.politico.eu/article/sanna-marin-finland-online-harassment-women-government-targeted/

Chen, G., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We should not get rid of incivility online. *Social Media + Society*, *5*(3), 205630511986264. https://doi.org/10/gghcnh

Choi, J.-K., & Bowles, S. (2007). The coevolution of parochial altruism and war. *Nature*, *318*(October), 636–640. https://doi.org/10.1126/science.1144237

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679. https://doi.org/10/f6dxrx

College of Policing. (2020, October 20). Major investigation and public protection. *Responding to Hate*. https://www.app.college.police.uk/app-content/major-investigation-and-public-protection/hate-crime/responding-to-hate/#agreed-definitions

Costello, M., Hawdon, J., Bernatzky, C., & Mendes, K. (2019). Social group identity and perceptions of online hate. *Sociological Inquiry*, *89*(3), 427–452. https://doi.org/10/gghcnc

Crocker, J., & Schwartz, I. (1985). Prejudice and ingroup favoritism in a minimal intergroup situation: Effects of self-esteem. *Personality and Social Psychology Bulletin*, *11*(4), 379–386. https://doi.org/10.1177/0146167285114004

Dieckmann, J., Geschke, D., & Braune, I. (2018). *Für die Auseinandersetzung mit Diskriminierung ist die Betroffenen Perspektive von großer Bedeutung* [For dealing with discrimination the perspective of the victims is of high relevance]. Institut für Demokratie und Zivilgesellschaft. https://doi.org/10.19222/201702/4

Dovidio, J. F., Gaertner, S. L., & Saguy, T. (2007). Another view of "we": Majority and minority group perspectives on a common ingroup identity. *European Review of Social Psychology*, *18*(1), 296–330. https://doi.org/10/bwt4jb

Duckitt, J. (2015). Authoritarian personality. In J.D. Wright (Ed.). *International Encyclopedia of the Social& Behavioral Sciences* (2nd. Ed). Elsevier: https://doi.org/10.1016/B978-0-08-097086-8.24042-7

Duckitt, J., & Sibley, C. G. (2010). Personality, ideology, prejudice, and politics: A dual-process motivational model. *Journal of Personality*, *78*(6), 1861–1893. https://doi.org/10/dq9ptj

Duckitt, J., Wagner, C., du Plessis, I., & Birum, I. (2002). The psychological bases of ideology and prejudice: Testing a dual process model. *Journal of Personality and Social Psychology*, *83*(1), 75–93. https://doi.org/10/cgq3sx

Erjavec, K., & Kovačič, M. P. (2012). "You don't understand, this is a new war!" Analysis of hate speech in news web sites' comments. *Mass Communication and Society*, *15*(6), 899–920. https://doi.org/10/gfgnmm

Festl, R. (2016). Perpetrators on the internet: Analyzing individual and structural explanation factors of cyberbullying in school context. *Computers in Human Behavior*, *59*, 237–248. https://doi.org/10/f8hw28

Fischer, R., Hanke, K., & Sibley, C. G. (2012). Cultural and institutional determinants of social dominance orientation: A cross-cultural meta-analysis of 27 societies. *Political Psychology*, *33*(4), 437–467. https://doi.org/10.1111/j.1467-9221.2012.00884.x

Frischlich, L., & Humprecht, E. (2021). *Trust, democratic resilience, and the infodemic*. Public Policy Institute.

Frischlich, L., Rieger, D., Hein, M., & Bente, G. (2015). Dying the right-way? Interest in and perceived persuasiveness of parochial extremist propaganda increases after mortality salience. *Frontiers in Psychology: Evolutionary Psychology and Neuroscience*, *6*(1222). https://doi.org/10/f7n3q8

Frischlich, L., Schatto-Eckrodt, T., Boberg, S., & Wintterlin, F. (2021). Roots of incivility: How personality, media use, and online experiences shape uncivil participation. *Media and Communication*, *9*(1), 195–208. https://doi.org/10.17645/mac.v9i1.3360

Gagliardone, I., Pohjonen, M., Zerai, A., Beyene, Z., Aynekulu, G., Bright, J., Bekalu, M. A., Seifu, M., Moges, M. A., Stremlau, N., Taflan, P., Gebrewolde, T. M., & Teferra, Z. M. (2016). *MECHACHAL: Online debates and elections in Ethiopia–From hate speech to engagement in social media*. Oxford University.

Gardiner, B., Mansfield, M., Anderson, I., Hoolder, J., Louter, D., & Ulumanu, M. (2016). The dark side of Guardian comments. *The Guardian*. https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments

Gelber, K., & Mcnamara, L. (2015). Evidencing the harms of hate speech. *Social Identities, 22*(3), 234–341. https://doi.org/10.1080/13504630.2015.1128810

Gervais, B. T. (2017). More than mimicry? The role of anger in uncivil reactions to elite political incivility. *International Journal of Public Opinion Research*, *29*(3), 384–405. https://doi.org/10/gftpws

Gervais, B. T. (2019). Rousing the partisan combatant: Elite incivility, anger, and antideliberative attitudes. *Political Psychology*, *40*(3), 637–655. https://doi.org/10/gghcns

Geschke, D., Klaßen, A., Quent, M., & Richter, C. (2019). *Hass im Netz–Der schleichende Angriff auf unsere Demokratie [Hate on the net – the creeping attack on our democracy]*. Institut für Demokratie und Zivilgesellschaft.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10/fhfs36

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/10/cq64hc

Hahn, L., Tamborini, R., Novotny, E., Grall, C., & Klebig, B. (2019). Applying moral foundations theory to identify terrorist group motivations. *Political Psychology*, *40*(3), 507–522. https://doi.org/10.1111/pops.12525

Haidt, J., & Joseph, C. (2008). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind Volume 3: Foundations and the future* (pp. 367–391). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195332834.003.0019

Harmon-Jones, E., & Harmon-Jones, C. (2016). Anger. In L. Feldmann Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (4th ed., pp. 774–791). The Guilford Press.

Haslam, C., Cruwys, T., Haslam, S. A., Dingle, G., & Chang, M. X. L. (2016). Groups 4 Health: Evidence that a social-identity intervention that builds and strengthens social group membership improves mental health. *Journal of Affective Disorders*, *194*, 188–195. https://doi.org/10/gf3hcv

Herbst, S. (2010). *Rude democracy: Civility and incivility in American politics*. Temple University Press.

Heym, N., Kibowski, F., Bloxsom, C. A. J., Blanchard, A., Harper, A., Wallace, L., Firth, J., & Sumich, A. (2021). The dark empath: Characterising dark traits in the presence of empathy. *Personality and Individual Differences*, *169*, 110172. https://doi.org/10.1016/j.paid.2020.110172

Ho, A. K., Sidanius, J., Kteily, K., Jennifer, J., Pratto, F., Henkel, K. E., Foels, R., & Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO7 scale. *Journal of Personality and Social Psychology*, *109*(6), 1003–1028. https://doi.org/10.1037/pspi0000033

Hogg, M. A., Sherman, D. K., Dierselhuis, J., Maitner, A. T., & Moffitt, G. (2007). Uncertainty, entitativity, and group identification. *Journal of Experimental Social Psychology*, *43*(1), 135–142. https://doi.org/10.1016/j.jesp.2005.12.008

Hsueh, M., Yogeeswaran, K., & Malinen, S. (2015). "Leave your comment below": Can biased online comments influence our own prejudicial attitudes and behaviors? *Human Communication Research*, *41*(4), 557–576. https://doi.org/10.1111/hcre.12059

Jetten, J., Haslam, A. S., & Haslam, C. (2012). The case for a social identity analysis of health and well-being. In J. Jetten, C. Haslam, & S. A. Haslam (Eds.), *The Social Cure: Identity, Health and Well-being* (pp. 3–21). Psychology Press.

Jonas, E., McGregor, I., Klackl, J., Agroskin, D., Fritsche, I., Holbrook, C., Nash, K., Proulx, T., & Quirin, M. (2014). Threat and defense: From anxiety to approach. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 49, pp. 219–286). Elsevier. https://doi.org/10.1016/B978-0-12-800052-6.00004-4

Jost, J. T. (2017). Ideological asymmetries and the essence of political psychology. *Political Psychology*, *38*(2), 167–208. https://doi.org/10/ggmpfr

Kenski, K., Coe, K., & Rains, S. A. (2020). Perceptions of uncivil discourse online: An examination of types and predictors. *Communication Research*, *47*(6), 795–814. https://doi.org/10/gghcnf

Koban, K., Stein, J. P., Eckhardt, V., & Ohler, P. (2018). Quid pro quo in Web 2.0. Connecting personality traits and Facebook usage intensity to uncivil commenting intentions in public online discussions. *Computers in Human Behavior*, *79*, 9–18. https://doi.org/10/gf3gv4

Koval, P., Laham, S. M., Haslam, N., Brock, B., & Whelan, J. (2012). Our flaws are more human than yours: Ingroup bias in humanizing negative characteristics. *Personality and Social Psychology Bulletin*, *38*(3), 283–295. https://doi.org/10/bwt484

Kreißel, P., Ebner, J., Urban, A., & Guhl, J. (2018). Hass auf Knopfdruck– Rechtsextreme Trollfabriken und das Ökosystem koordinierter Hasskampagnen im Netz [Hate on the press of the button: Right-wing extremist troll campaigns and the ecosystem of coordinated hate campaigns]. *Institute for Strategic Dialogue*, 28–28.

Kurek, A., Jose, P. E., & Stuart, J. (2019). 'I did it for the LULZ': How the dark personality predicts online disinhibition and aggressive online behavior in adolescence. *Computers in Human Behavior, 98*, 31–40. https://doi.org/10/ggft9d

Leets, L., & Giles, H. (1999). Harmful speech in intergroup encounters: An organizational framework for communication research. *Annals of the International Communication Association*, *22*(1), 91–137. https://doi.org/10.1080/23808985.1999.11678960

Lindquist, K. A., Gendron, M., & Satpute, A. B. (2016). Language and emotion putting words into feelings and feelings into words. In L. Feldmann Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (4th ed., pp. 579–594). The Guilford Press.

Ma, Y., Wang, C., & Han, S. (2011). Neural responses to perceived pain in others predict real-life monetary donations in different socioeconomic contexts. *NeuroImage*, *57*(3), 1273–1280. https://doi.org/10.1016/j.neuroimage.2011.05.003

March, E. (2019). Psychopathy, sadism, empathy, and the motivation to cause harm: New evidence confirms malevolent nature of the Internet Troll. *Personality and Individual Differences*, *141*, 133–137. https://doi.org/10.1016/j.paid.2019.01.001

Martinovic, B., Jetten, J., Smeekes, A., & Verkuyten, M. (2017). Collective memory of a dissolved country: Collective nostalgia and guilt as predictors of interethnic relations between diaspora groups from former Yugoslavia. *Journal of Social and Political Psychology*, 588–607. https://doi.org/10/gf3gzb

Marwick, A., & Lewis, R. (2017). *Media manipulation and disinformation online*. Data & Society Research Institute. https://datasociety.net/library/media-manipulation-and-disinfo-online/

Mason, G. (2005). Hate crime and the image of the stranger. *The British Journal of Criminology*, *45*(6), 837–859. https://doi.org/10.1093/bjc/azi016

McFarland, S. (2017). Identification with all humanity: The antithesis of prejudice, and more. In C. G. Sibley & F. K. Barlow (Eds.). *The Cambridge handbook of the psychology of prejudice* (pp. 632–654). Cambridge University Press. https://doi.org/10.1017/9781316161579.028

Međedović, J., & Petrović, B. (2015). The dark tetrad: Structural properties and location in the personality space. *Journal of Individual Differences*, *36*(4), 228–236. https://doi.org/10.1027/1614-0001/a000179

Miklikowska, M. (2017). Empathy trumps prejudice: The longitudinal relation between empathy and anti-immigrant attitudes in adolescence. *Developmental Psychology*, *54*(4), 703. https://doi.org/10.1037/dev0000474

Miller, J. D., Lynam, D. R., Hyatt, C. S., & Campbell, W. K. (2017). Controversies in narcissism. *Annual Review of Clinical Psychology*, *13*(1), 291–315. https://doi.org/10/gghcm9

Muldoon, O. T., Haslam, S. A., Haslam, C., Cruwys, T., Kearns, M., & Jetten, J. (2019). The social psychology of responses to trauma: Social identity pathways associated with divergent traumatic responses. *European Review of Social Psychology*, *30*(1), 311–348. https://doi.org/10/gg4whk

Mutz, D. C. (2015). *In-your-face politics: The consequences of uncivil media*. Princeton University Press.

Nabi, R. L. (2002). Anger, fear, uncertainty, and attitudes: A test of the cognitive-functional model. *Communication Monographs*, *69*(3), 204–216. https://doi.org/10/dchzh4

Näsi, M., Räsänen, P., Hawdon, J., Holkeri, E., & Oksanen, A. (2015). Exposure to online hate material and social trust among Finnish youth. *Information Technology & People*, *28*(3), 607–622. https://doi.org/10.1108/ITP-09-2014-0198

Nussbaum, M. C. (2011). Perfectionist liberalism and political liberalism. *Philosophy & Public Affairs*, *39*(1), 3–45. https://doi.org/10.1111/j.1088-4963.2011.01200.x

O'Sullivan, P. B., & Flanagin, A. J. (2003). Reconceptualizing "flaming" and other problematic messages. *New Media & Society*, *5*(2), 69–94. https://doi.org/10/b3txz4

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, *6*(2), 259–283. https://doi.org/10/dz4rp6

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10/d2jxm9

Perry, B., & Alvi, S. (2012). "We are all vulnerable": The in terrorem effects of hate crimes. *International Review of Victimology*, *18*(1), 57–71. https://doi.org/10/fxsq7s

Perry, R., Sibley, C. G., & Duckitt, J. (2013). Dangerous and competitive worldviews: A meta-analysis of their associations with social dominance orientation and right-wing authoritarianism. *Journal of Research in Personality*, *47*(1), 116–127. https://doi.org/10/tvc

Pfundmair, M., & Wetherell, G. (2019). Ostracism drives group moralization and extreme group behavior. *The Journal of Social Psychology*, *159*(5), 518–530. https://doi.org/10/ggbhgv

Phillips, W. M. (2012). *This is why we can't have nice things: The origins, evolution and cultural embeddedness of online trolling*. ProQuest Dissertations Publishing.

Raskin, R., & Hall, C. S. (1981). The narcissistic personality inventory: Alternate form reliability and further evidence of construct validity. *Journal of Personality Assessment*, *45*(2), 159–162. https://doi.org/10/ch4b6b

Ratner, K. G., & Amodio, D. M. (2013). Seeing "us vs. them": Minimal group effects on the neural encoding of faces. *Journal of Experimental Social Psychology*, *49*(2), 298–301. https://doi.org/10/f4rvcr

Reichelmann, A., Hawdon, J., Costello, M., Ryan, J., Blaya, C., Llorent, V., Oksanen, A., Räsänen, P., & Zych, I. (2020). Hate knows no boundaries: Online hate in six nations. *Deviant Behavior*, advanced online publication, https://doi.org/10.1080/01639625.2020.1722337

Rieger, D., Frischlich, L., & Bente, G. (2013). *Propaganda 2.0: Psychological effects of right-wing and Islamic extremist internet videos* (Vol. 44). Wolters Kluwer Deutschland.

Rieger, D., Frischlich, L., & Bente, G. (2017). Propaganda in an insecure, unstructured world: How psychological uncertainty and authoritarian attitudes shape the evaluation of right-wing extremist internet propaganda. *Journal for Deradicalization*, *10,* 203–229.

Rossini, P. (2020). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research*, advanced online publication. https://doi.org/10.1177/0093650220921314

Scherer, K. R. (1987). *Toward a dynamic theory of emotion: The component process model of affective states* (Geneva Studies in Emotion, pp. 1–96).

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10/fwmgvv

Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Ho, A. K., Sibley, C., & Duriez, B. (2013). You're inferior and not worth our concern: The interface between empathy and social dominance orientation. *Journal of Personality*, *81*(3), 313–323. https://doi.org/10.1111/jopy.12008

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression.* Cambridge University Press.

Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016). Analyzing the targets of hate in online social media. *Cornell University Library*, *June*. http://arxiv.org/abs/1603.07709

Smith, E. R., & Mackie, D. M. (2015). Dynamics of group-based emotions: Insights from intergroup emotions theory. *Emotion Review*, *7*(4), 349–354. https://doi.org/10.1177/1754073915590614

Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*, *44*(2), 136-146. https://doi.org/10/gf3gx2

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In S. Worchel & W. G. Austin (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33–47). Brooks-Cole. https://doi.org/10.1016/S0065-2601(05)37005-5

Timmers, I., Park, A. L., Fischer, M. D., Kronman, C. A., Heathcote, L. C., Hernandez, J. M., & Simons, L. E. (2018). Is empathy for pain unique in its neural correlates? A meta-analysis of neuroimaging studies of empathy. *Frontiers in Behavioral Neuroscience*, *12*. https://doi.org/10.3389/fnbeh.2018.00289

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory.* Basil Blackwell.

Uhlmann, E. L., Korniychuk, A., & Obloj, T. (2018). Initial prejudices create cross-generational intergroup mistrust. *Plos One*, *13*(4), e0194871. https://doi.org/10.1371/journal.pone.0194871

Walters, M. A., Brown, R., & Wiedlitzka, S. (2016). Causes and motivations of hate crime. *Equality and Human Rights Commission Research Report*, *102*, 61–61.

Wilhelm, C., Joeckel, S., & Ziegler, I. (2020). Reporting hate comments: Investigating the effects of deviance characteristics, neutralization strategies, and users' moral orientation. *Communication Research*, *47*(6), 921–944. https://doi.org/10.1177/0093650219855330

Yamagishi, T., & Kiyonari, T. (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly*, *63*(2), 116–132. https://doi.org/10.2307/2695887

Ziegele, M., Koehler, C., & Weber, M. (2018). Socially destructive? Effects of negative and hateful user comments on readers' donation behavior toward refugees and homeless persons. *Journal of Broadcasting & Electronic Media*, *62*(4), 636–653. https://doi.org/10/gf8pn4