



The virtual disengagement hypothesis: A neurophysiological framework for reduced empathy on social media

Maria Tavares¹ · Ben Rein²

Accepted: 5 August 2024
© The Psychonomic Society, Inc. 2024

Abstract

Social media is a hotbed of interpersonal conflict and aggression. Platforms such as Twitter and Instagram are used by more than 62% of the global population, facilitating billions of user interactions every day. However, many of these exchanges involve hostile, insensitive, and antisocial behaviors. This raises the question: is empathy blunted on social media? Substantial evidence demonstrates that humans tend to behave more rudely in virtual settings, but considering the scarcity of physiological data collected under these circumstances, it remains unclear how the neural systems guiding social cognition and empathy may function differently in online interactions. We propose the “Virtual Disengagement Hypothesis,” a conceptual framework to explain the prevalence of hostility online. It posits that interactions occurring on social media omit social cues that facilitate the assessment of a social partner’s affective state, such as facial expressions and vocal tone, and thus fail to sufficiently recruit brain circuitry involved in empathy, such as the anterior cingulate cortex, insula, and prefrontal cortex. Additionally, interactions on social media occur asynchronously and in a “replayed” context, which may further limit recruitment of empathy systems. As a result of this diminished sensitivity to others’ states, users may be predisposed to inconsiderate or outright antisocial behaviors. Given the massive and growing base of users on these platforms, we urge researchers to expand efforts that focus on neuroimaging in virtual settings with a particular emphasis on developing social media-relevant behavioral designs.

Keywords Empathy · Social media · Social behavior · Emotional empathy

In the past several years, social media platforms such as Instagram, Facebook, Twitter, and TikTok have boomed into worldwide hubs for a global community of users. As of January 2024, social networking sites boasted a whopping total of 5.04 billion users (Ani Petrosyan, 2024), gaining more than a billion new users since 2020 (Dixon, 2023). With 62.3% of the world’s population now using social media, including 84% of 18–29 year olds (Wong & Bottorff, 2023), we must recognize these apps as an unignorable element of modern life. Moreover, consumers invest a significant amount of time: in 2023, the average user spent 151 min per day on social media, up from 97 min in 2013 (Dixon, 2024a).

Modern platforms, such as Instagram, boast a range of features, including news, shopping, and entertainment, but social media remains a social enterprise at heart. In 2022, the most commonly cited reason for using social media was to keep in touch with friends and family (Dixon, 2024b). Despite this core mission of promoting social exchange, social networking sites have unfortunately grown into a breeding ground for hostility, inconsiderate behavior, and even hate speech (Jonathan Vanian, 2023). Several studies have examined the characteristics of polarization and animosity on social media (Brady et al., 2019; Rathje et al., 2021), but it remains unknown how the brain regions responsible for social cognition and empathy may be influenced by the virtual context of these exchanges.

We propose that interactions occurring on social media fail to fully engage empathy-related brain circuitry, because they exclude the necessary social cues required for the assessment of others’ affective states. This may explain why hate speech and conflict are so prominent online, because users may be less inclined to perspective-taking than in

✉ Ben Rein
brein@mindsience.org

¹ University of California San Diego School of Medicine, La Jolla, CA, USA

² Mind Science Foundation, San Antonio, TX, USA

face-to-face interaction. We title this the *Virtual Disengagement Hypothesis*, referring to the disengagement of empathy-related circuitry in virtual settings. In the sections below, we will outline the conceptual framework and existing data that support this hypothesis.

“Impressions” on social media as interactions

Social networking sites such as Facebook, Twitter, Instagram, YouTube, and TikTok provide users with numerous momentary social interactions. The user interface is populated with serial content in the format of text, photo, or video. The identity of the user who posted the content is represented as an image paired with a username. When users view a piece of content, it is called an *impression*. Impressions are thus analogous to a virtual interaction between a demonstrator (the poster) and an observer (the viewer). Observers may engage by *liking* or *commenting* on the post or *sharing* with another user. Multiple observers may interact by *replying* to one another’s comments. Scrolling social media thus constitutes a series of momentary social interactions that occur in distinct and replicable contexts.

Interactions on social media are inclined to hostility

These massive platforms connect and facilitate interactions between users across the globe, although many such interactions are hostile. Between 2014 and 2020, the proportion of users who received physical threats on social media doubled from 7 to 14%, and the incidence of sexual harassment rose from 5 to 11% (Vogels, 2021). In a 2020 survey of more than 10,000 American adults, 41% had experienced online harassment (Vogels, 2021). A more recent survey in 2023 found that this number had risen to 52% (ADL Center for Technology & Society, 2023). It is estimated that 10–40% of children and adolescents experience cybervictimization (Kowalski et al., 2014); some studies reported rates as high as 72% (Juvonen & Gross, 2008). In 2023, 51% of teens reported experiencing online harassment in the past 12 months alone (ADL Center for Technology & Society, 2023). Importantly, victims of virtual offenses report significant emotional responses. In studies that use controlled paradigms to induce virtual exclusion and rejection, recipients reported feelings of anger, hurt, or sadness (Donate et al., 2017; Wolf et al., 2015) and showed neural responses representative of social pain (Eisenberger et al., 2003; Radke et al., 2021). As such, cyberbullying is a major risk factor for suicidality in youth; a meta-analysis of 33

studies suggests a 2.57-fold increased risk in victims of online harassment (John et al., 2018).

Virtual conflict is not a rare occurrence. More than one in five Americans in a 2017 survey admitted to getting in online arguments “sometimes” or “often,” and 43% of arguers attributed these conflicts to disagreeing with a stranger over something one of them posted (Barna, 2017). According to an analysis of 11 billion organic posts on social media where one user tagged another user, more than one in every 15 posts was negative (Whatman, 2022). Interactions may be particularly unhealthy in certain settings, such as comments sections on news articles. One study, which evaluated more than 6,000 comments on stories published in the *Daily Star*, found that 22% contained “uncivil” behaviors, such as name-calling and vulgarity (Coe et al., 2014). Online discussions are less likely to achieve consensus than in-person debates and are perceived as more negative by participants (Baek et al., 2012).

Empathy in virtual interactions

Empathy can be defined as the adoption of a cognitive or affective state that is more appropriate to another’s situation than one’s own. It involves an observer and a demonstrator, whereby the former understands and/or shares the state of the latter. This phenomenon can be subdivided into two components: emotional empathy and cognitive empathy.

Emotional empathy entails an emotional response whereby the observer “steps into” and shares the emotions of a demonstrator. This occurs rapidly and often without conscious thought, permitting an instantaneous understanding of a conspecific’s condition. Emotional empathy can thus facilitate communication in groups and increase the likelihood of prosocial helping behaviors by imposing a demonstrator’s negative emotional states on observers (De Waal & Preston, 2017). Emotional empathy is thought to be driven primarily by activity in the anterior cingulate cortex (ACC) and insular cortex (IC) (De Waal & Preston, 2017).

Cognitive empathy is the process whereby an observer takes the perspective of a demonstrator and gains a conceptual understanding of what they are experiencing. Unlike emotional empathy, it is a deliberate and conscious process. Cognitive empathy involves more “mentalizing” areas, such as the prefrontal cortex (PFC) (De Waal & Preston, 2017). Therefore, we can come to understand another’s condition by modeling their emotional state, cognitively taking their perspective, or both.

The role of social cues

Various social cues, such as facial expressions and vocal tone, contribute to the acquisition of empathy in face-to-face interactions, and detection of these cues has been shown

to engage empathy-related brain regions in observers. For example, processing facial expressions is associated with increased activity in the ACC (Chan et al., 2016), and listening to vocalizations (i.e., laughing and crying) from infants or adults is associated with increased activity in ACC, insula, and medial prefrontal cortex (Sander & Scheich, 2005; Sander et al., 2007; Sun et al., 2022). These findings indicate that social cues that convey emotional information engage empathy-related brain areas. However, virtual interactions, such as those occurring on social media, may fail to represent these cues, thereby blunting one's ability to empathize. There also is evidence that olfactory signals, which are entirely omitted from virtual interactions, may play a role in mediating empathic responses. A recent paper demonstrated that aggressive behavior was significantly attenuated in men after sniffing emotional tears from women (Agron et al., 2023). Depending on the modality of the social media content (i.e., text or video), these various sensory cues may be entirely absent or incompletely represented, thereby limiting the engagement of these empathy-related brain areas.

Certain forms of content (i.e., videos) preserve certain cues, including facial expressions and vocal tone. However, the asynchronous nature in which this content is viewed also may contribute to the lack of empathy. Research suggests that live interactions with real-time feedback of social cues more robustly engage empathy circuitry. For example, research subjects display greater ACC activity and enhanced functional connectivity between ACC and right anterior insula when engaging in live eye contact with a partner versus focusing on the partner's eyes in a prerecorded video (Koike et al., 2019). Considering that many social media impressions involve one user viewing a prerecorded video of another, this "replayed" context of virtual interactions may diminish empathy.

Empathy also is facilitated by the subconscious modeling of a demonstrator's physiological cues during in-person interactions (Decety & Hodges, 2006). Through mimicry of facial expressions, vocal tone, and body language of those demonstrating emotional states, an observer may adopt a related state (Decety & Hodges, 2006). For example, observers show increased activity in facial muscles responsible for smiling when viewing happy facial expressions and increased activity in muscles involved in frowning when viewing negative expressions (Fujimura et al., 2010). However, evidence suggests that prerecorded videos do not facilitate mimicry as effectively. Women display weaker activity of frowning-related facial muscles when viewing prerecorded versus live negative facial expressions and show reduced activity in regions comprising the mirror neuron system, such as the inferior frontal gyrus (Hsu et al., 2022), which is implicated in emotion recognition (Perry et al., 2017). The reduced mimicry and weaker engagement of this circuitry in subjects viewing prerecorded expressions

provides direct support for the hypothesis that replayed content on social media may be less effective at recruiting empathic support from observers by occluding emotion recognition and mimicry of emotions.

Furthermore, user interactions on social media include an asynchronous "engagement" phase, during which an observer may engage with a post by taking an action, such as commenting. During this period, there is a total absence of real-time social cues from the creator that would typically convey information about their emotional response to the commenter's actions. This prevents hostile commenters from witnessing live emotional responses to any harmful messages they leave, potentially lowering the emotional cost of harassment.

In summary, we hypothesize that (1) the absence or incomplete representation of social cues on social media and (2) the asynchronous nature of social media impressions may diminish recruitment of empathy-related brain circuitry. Whereas content formats, such as photos or videos, may partially represent certain cues, these content modalities fail to comprehensively represent the multisensory nature of in-person interaction and thus may not evoke the same physiological responses. Additionally, the replayed context of social media impressions and the absence of social cues during the engagement phase may diminish empathic responses in observers.

Supportive behavioral data

The Virtual Disengagement Hypothesis would be further supported by real-world data that indicate an absence of empathy in virtual interactions. So, do humans truly display behavioral tendencies indicating a loss of empathy in virtual interactions? Evidence appears to suggest so. A large study found that young adults self-report higher levels of empathy in real-life scenarios than online interactions (Carrier et al., 2015). Studies comparing behavior in virtual versus in-person interactions have revealed trends of unusual and uninhibited behavior online, which has been previously referred to as the "online disinhibition effect" (Suler, 2004). This can include toxic behaviors, such as swearing, name-calling, rude language, criticism, hostility, and threats (Adam Joinson, 1998; Suler, 2004). Political discussions that take place online also are more hostile than those that occur in-person (Bor & Petersen, 2022), and study subjects are more likely to voice opinions on controversial topics when interacting with others online than in person (Ho & McLeod, 2008). We suspect that these tendencies may be driven by an impaired ability to mentalize with others online and a lower propensity to consider their perspectives when choosing behaviors. Of note, this discrepancy between virtual and in-person behavior was documented as early as 1984 (Kiesler

et al., 1984); participants in these studies displayed a propensity for “insults, name-calling, and hostile comments” when interacting via computer. This suggests that the tendency to behave crudely in virtual settings may not be simply a by-product of modern social media, but rather an innate and biological behavioral tendency. In support of this, a recent mouse study showed that mice will make altruistic decisions and share treats with conspecifics but will shift toward more selfish choices when the recipient animal is placed behind an opaque divider (Scheggia et al., 2022). Even in mice, prosocial behaviors appear to become less prevalent when the subject is out of sight.

Additionally, meta-analyses have revealed that the strongest and most consistent predictor of antisocial online behaviors (such as trolling, bullying, and harassment) was psychopathy (Moor & Anderson, 2019). This is noteworthy, because psychopathy is characterized by reduced empathy and aberrant activity in empathy-related brain regions when observing or imagining others in discomfort (Decety et al., 2013a, 2013b). The reported relationship between psychopathy and online antisocial behavior may suggest a shared mechanism, potentially supporting the proposed hypothesis.

Proposed involvement of contextual factors

While we propose that empathy is diminished in virtual interaction settings, we emphasize that this process is likely sensitive to various contextual factors. Inter-user interactions on social media are characterized by certain features, including a focus on a certain topic or piece of content, a particular modality of the interaction (i.e., video versus text), and the interpersonal relationship between those participating (i.e., friends or strangers). We thus propose that these factors are liable to influence the level of empathy a user experiences in virtual interactions.

1. **Topic:** Many social media posts address contentious or controversial topics, such as politics and social issues, whereas others do not. We propose that virtual interactions surrounding charged topics are more likely to evoke antisocial behavior, because they highlight ingroup/outgroup differences, which are known to modulate empathy (Vanman, 2016).
2. **Modality:** Various social media platforms represent content through different modalities (i.e., text-based on Twitter and video-based on TikTok). We predict that modalities that represent a more limited range of social cues will be associated with more hostile behavior. Considering the evidence that processing social cues engages empathy-related brain areas, modalities that represent fewer social cues should engage weaker neural correlates of empathy. Therefore, we propose that social

media platforms that represent interactions through text only will engender weaker empathic responses than video-based platforms.

3. **Participants:** Social media platforms offer the unique ability to interface with broad and indiscriminate audiences, which often includes strangers. We propose that users who are unfamiliar will be more prone to virtual disengagement, resulting in hostility. People tend to empathize less with those that they perceive as less close to them (Depow et al., 2021), and virtual interactions may exaggerate this perceived social distance.

Alternative theories and explanations

It is unlikely that virtual disengagement alone is entirely responsible for the antisocial trends observed in virtual settings, such as social media. Of course, we must acknowledge the involvement of several other factors at play.

1. **Anonymity:** Multiple studies have shown that subjects behave differently in online settings when granted anonymity; they are more likely to engage in trolling behavior (Nitschinsk et al., 2022) and report greater self-esteem (Joinson, 2001). The ability to create and use anonymous profiles on social media may be a major contributing factor. Anonymity grants users the ability to behave inappropriately while remaining protected from culpability. However, anonymity cannot entirely explain the cruel behaviors of social media users. One report indicated that only a small minority (5.9%) of Twitter users are anonymous, whereas most accounts (67.9%) are operated by identifiable users (Peddinti et al., 2014). Given the prevalence of hostility online, it is unlikely that this small share of anonymous accounts is responsible. Rather, we suspect that conditions of anonymity may exacerbate the lack of empathy in virtual settings.
2. **Perceived distance:** Users engaging in interactions online are, of course, separated physically. It is likely that the lack of physical threat or tangible repercussions presented in this setting plays a major role in bad behavior online. No doubt, users are more prone to make threatening comments when they are not at risk of immediate physical harm as a result.
3. **Online disinhibition:** Online disinhibition has been defined as “a psychological state in which individuals feel more relaxed and willing to engage in certain behaviors in the online environment” (Cheung et al., 2020). Several previous studies have reported that subjects behave in a less restrained manner in virtual settings. This too is compatible with the virtual disengagement hypothesis, because these loose behaviors may be facilitated by reduced empathy, which diminishes sensitivity

to an interaction partner's perspective. Notably, online disinhibition only refers to a change in behavior in virtual settings but does not present a neurophysiological explanation.

4. **Scale/mismatch:** A recent theory proposes that users on social media take in an abnormally large amount of social and moral content, which may not be conducive for empathy (Robertson et al., 2024). This creates a “mismatch” between evolved social behaviors and the environments that we are exposed to online. The Virtual Disengagement Hypothesis could offer a neurophysiological explanation for why this mismatch occurs, because this disparity in social behaviors could be partially driven by the absence of evolutionarily adapted social cues that facilitate empathy.

Given these alternative explanations, we propose that the virtual disengagement of empathy circuitry is but one piece that interfaces with a broader puzzle driving these behaviors. Of note, none of these proposed explanations incorporate neurophysiological data, but rather offer psychological explanations for abnormal behavior in virtual interactions. We propose that the virtual disengagement hypothesis is not incompatible with these explanations, but rather the disengagement of empathy-related circuitry likely interacts with each of these variables to produce the observed behavioral trends.

Conclusion and future directions

Social networking sites offer great potential for humanity through bridging global communities, enabling large-scale teamworking efforts (Rein, 2022), and even facilitating dissemination of scientific and medical information (Lowe-Calverley et al., 2022; Rein, 2023; Yammine et al., 2018). However, these platforms also engender widespread harassment that seems to particularly affect minors, increasing risk for suicidality. As the number of social media users is projected to reach 5.85 billion by 2027 (Dixon, 2023), there will almost certainly be a corresponding increase in abusive interactions that may cause preferential harm to younger users joining these apps. While it is apparent that online interactions differ considerably from face-to-face exchanges, the precise neurophysiological nature of these differences is unknown. Considering the large and ever-growing presence of social media, there is an unaddressed need for social neuroscience research to examine the interactions occurring there.

To our awareness, no studies have directly compared the activity of empathy-related brain areas between in-person interactions and those occurring in virtual settings reminiscent of social media environments. We thus encourage

researchers to develop and validate social media-relevant behavioral designs for use in conjunction with traditional neuroimaging methods. Several examples are provided below.

- a) **Characterize contributions of various sensory cues:** Future studies should examine how neural responses to social stimuli are altered in modalities that omit certain social cues. For example, subjects could be positioned for neuroimaging before a confederate, posing as a research assistant, enters the room and loudly, expressively suffers a painful experience, such as hitting their head or stubbing their toe. Activity in empathy-related areas could be measured in observers and contrasted with subjects that observe (1) a video of the confederate demonstrating this pain, (2) the same video with audio removed, or (3) the audio only. The systematic deconstruction of in-person interaction and progressive removal of various sensory cues is intended to permit the characterization of their individual contributions.
- b) **Distinguish brain responses to in-person versus virtual social stimuli:** Future studies should compare regional brain activity in human subjects observing live (in-person) versus virtual (on-screen video/audio) social cues (of positive or negative valence), such as facial expressions, spoken language, and body language. The presentation of virtual cues can be further manipulated to explore differences in neural responses to video presented in raw format versus a social media-relevant design (i.e., including a username, profile picture, and caption).
- c) **Validate the role of interpersonal relationship:** Are social media users more likely to criticize and attack other users they do not know personally? Can these attacks be predicted by neural activity? Future studies could compare behavioral tendencies and corresponding brain activity in subjects interacting with strangers or familiar individuals in virtual or in-person settings.
- d) **Examine the role of topic:** Future studies should examine the interaction between topic and modality, examining whether subjects are more prone to disengaging empathy-related brain areas when discussing controversial or uncontroversial topics in virtual settings or in-person. For example, subjects could be asked to discuss controversial topics (i.e., politics) or uncontroversial topics (i.e., favorite dog breed) with a stranger in various modalities while corresponding brain activity is measured. Multiple modalities could be explored, such as (1) in-person, (2) video chat, and (3) text chat.

A primary future goal should be to design safer online spaces and cultivate more mindful interactions. Understanding how social media environments disrupt empathy

may empower us to achieve this. Certain interventions have already been explored; for example, promoting empathy-based counterspeech has been shown to reduce hate speech on social media and encourage users to delete hateful posts (Hangartner et al., 2021). By characterizing the neural underpinnings of virtual aggression and validating corresponding biomarkers, we may be able to harness neuroscience to accelerate our path to a more friendly virtual world. As social media platforms become a more permanent part of modern life, social neuroscience must continue to invest in understanding virtual interactions to ensure that the scientific enterprise keeps pace with the ever-growing digital age.

Funding Not applicable.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflicts of interest The authors have no conflicts of interest to disclose.

References

- ADL Center for Technology & Society. (2023). *Online Hate and Harassment: The American Experience*. https://www.adl.org/sites/default/files/pdfs/2023-12/Online-Hate-and-Harassmen-2023_0_0.pdf. Accessed 22 Mar 2024.
- Agron, S., De March, C. A., Weissgross, R., Mishor, E., Gorodisky, L., Weiss, T., Furman-Haran, E., Matsunami, H., & Sobel, N. (2023). A chemical signal in human female tears lowers aggression in males. *PLOS Biology*, 21(12), e3002442. <https://doi.org/10.1371/journal.pbio.3002442>
- Baek, Y. M., Wojcieszak, M., & Delli Carpini, M. X. (2012). Online versus face-to-face deliberation: Who? Why? What? With what effects? *New Media & Society*, 14(3), 363–383. <https://doi.org/10.1177/1461444811413191>
- Barna. (2017). *Why People Fight Online*. <https://www.barna.com/research/people-fight-online/>. Accessed 22 Mar 2024.
- Bor, A., & Petersen, M. B. (2022). The Psychology of Online Political Hostility: A Comprehensive, Cross-National Test of the Mismatch Hypothesis. *American Political Science Review*, 116(1), 1–18. <https://doi.org/10.1017/S0003055421000885>
- Brady, W. J., Wills, J. A., Burkart, D., Jost, J. T., & Van Bavel, J. J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, 148(10), 1802–1813. <https://doi.org/10.1037/xge0000532>
- Carrier, L. M., Spradlin, A., Bunce, J. P., & Rosen, L. D. (2015). Virtual empathy: Positive and negative impacts of going online upon empathy in young adults. *Computers in Human Behavior*, 52, 39–48. <https://doi.org/10.1016/j.chb.2015.05.026>
- Chan, S. W. Y., Sussmann, J. E., Romaniuk, L., Stewart, T., Lawrie, S. M., Hall, J., McIntosh, A. M., & Whalley, H. C. (2016). Deactivation in anterior cingulate cortex during facial processing in young individuals with high familial risk and early development of depression: F MRI findings from the Scottish Bipolar Family Study. *Journal of Child Psychology and Psychiatry*, 57(11), 1277–1286. <https://doi.org/10.1111/jcpp.12591>
- Cheung, C. M. K., Wong, R. Y. M., & Chan, T. K. H. (2020). Online disinhibition: Conceptualization, measurement, and implications for online deviant behavior. *Industrial Management & Data Systems*, 121(1), 48–64. <https://doi.org/10.1108/IMDS-08-2020-0509>
- 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/10.1111/jcom.12104>
- Decety, J., Chen, C., Harenski, C., & Kiehl, K. A. (2013a). An fMRI study of affective perspective taking in individuals with psychopathy: Imagining another in pain does not evoke empathy. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00489>
- Decety, J., & Hodges, S. D. (2006). The Social Neuroscience of Empathy. In *Bridging social psychology: Benefits of transdisciplinary approaches* (pp. 103–109). Lawrence Erlbaum Associates Publishers.
- Decety, J., Skelly, L. R., & Kiehl, K. A. (2013b). Brain response to empathy-eliciting scenarios involving pain in incarcerated individuals with psychopathy. *JAMA Psychiatry*, 70(6), 638. <https://doi.org/10.1001/jamapsychiatry.2013.27>
- Depow, G. J., Francis, Z., & Inzlicht, M. (2021). The experience of empathy in everyday life. *Psychological Science*, 32(8), 1198–1213. <https://doi.org/10.1177/0956797621995202>
- Dixon, S. J. (2023). *Number of social media users worldwide from 2017 to 2027 (in billions)*. <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>. Accessed 22 Mar 2024
- Dixon, S. J. (2024a). *Daily time spent on social networking by internet users worldwide for 2012 to 2024 (in minutes)*. Statista. <https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/>. Accessed 22 Mar 2024.
- Dixon, S. J. (2024b). *Most popular reasons for internet users worldwide to use social media as of 3rd quarter 2023*. <https://www.statista.com/statistics/715449/social-media-usage-reasons-worldwide/>. Accessed 22 Mar 2024.
- Donate, A. P. G., Marques, L. M., Lapenta, O. M., Asthana, M. K., Amodio, D., & Boggio, P. S. (2017). Ostracism via virtual chat room—Effects on basic needs, anger and pain. *PLoS ONE*, 12(9), e0184215. <https://doi.org/10.1371/journal.pone.0184215>
- Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, 302(5643), 290–292. <https://doi.org/10.1126/science.1089134>
- Fujimura, T., Sato, W., & Suzuki, N. (2010). Facial expression arousal level modulates facial mimicry. *International Journal of Psychophysiology*, 76(2), 88–92. <https://doi.org/10.1016/j.ijpsycho.2010.02.008>
- Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., Demirci, B. B., Derksen, L., Hall, A., Jochum, M., Munoz, M. M., Richter, M., Vogel, F., Wittwer, S., Wüthrich, F., Gilardi, F., & Donnay, K. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences*, 118(50), e2116310118. <https://doi.org/10.1073/pnas.2116310118>
- Ho, S. S., & McLeod, D. M. (2008). Social-psychological influences on opinion expression in face-to-face and computer-mediated communication. *Communication Research*, 35(2), 190–207. <https://doi.org/10.1177/0093650207313159>
- Hsu, C.-T., Sato, W., Kochiyama, T., Nakai, R., Asano, K., Abe, N., & Yoshikawa, S. (2022). Enhanced mirror neuron network activity and effective connectivity during live interaction among

- female subjects. *NeuroImage*, 263, 119655. <https://doi.org/10.1016/j.neuroimage.2022.119655>
- John, A., Glendenning, A. C., Marchant, A., Montgomery, P., Stewart, A., Wood, S., Lloyd, K., & Hawton, K. (2018). Self-harm, suicidal behaviours, and cyberbullying in children and young people: Systematic review. *Journal of Medical Internet Research*, 20(4), e129. <https://doi.org/10.2196/jmir.9044>
- Jonson, A. (1998). Causes and implications of disinhibited behavior on the Internet. In *Psychology and the Internet: Intrapersonal, interpersonal, and transpersonal implications* (pp. 43–60). Academic Press.
- Jonson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology*, 31(2), 177–192. <https://doi.org/10.1002/ejsp.36>
- Juvonen, J., & Gross, E. F. (2008). Extending the school grounds?—Bullying experiences in cyberspace. *Journal of School Health*, 78(9), 496–505. <https://doi.org/10.1111/j.1746-1561.2008.00335.x>
- Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, 39(10), 1123–1134. <https://doi.org/10.1037/0003-066X.39.10.1123>
- Koike, T., Sumiya, M., Nakagawa, E., Okazaki, S., & Sadato, N. (2019). What makes eye contact special? Neural substrates of on-line mutual eye-gaze: a hyperscanning fMRI study. *Eneuro*, 6(1), ENEURO.0284–18.2019. <https://doi.org/10.1523/ENEURO.0284-18.2019>
- Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, 140(4), 1073–1137. <https://doi.org/10.1037/a0035618>
- Lowie-Calverley, E., Barton, M., & Todorovic, M. (2022). Can we provide quality #MedEd on social media? *Trends in Molecular Medicine*, 28(12), 1016–1018. <https://doi.org/10.1016/j.molmed.2022.08.002>
- Moor, L., & Anderson, J. R. (2019). A systematic literature review of the relationship between dark personality traits and antisocial online behaviours. *Personality and Individual Differences*, 144, 40–55. <https://doi.org/10.1016/j.paid.2019.02.027>
- Nitschinsk, L., Tobin, S. J., & Vanman, E. J. (2022). The Disinhibiting Effects of Anonymity Increase Online Trolling. *Cyberpsychology, Behavior, and Social Networking*, 25(6), 377–383. <https://doi.org/10.1089/cyber.2022.0005>
- Peddinti, S. T., Ross, K. W., & Cappos, J. (2014). “On the internet, nobody knows you’re a dog”: A twitter case study of anonymity in social networks. *Proceedings of the Second ACM Conference on Online Social Networks*, 83–94. <https://doi.org/10.1145/2660460.2660467>
- Perry, A., Saunders, S. N., Stiso, J., Dewar, C., Lubell, J., Meling, T. R., Solbakk, A.-K., Endestad, T., & Knight, R. T. (2017). Effects of prefrontal cortex damage on emotion understanding: EEG and behavioural evidence. *Brain*, 140(4), 1086–1099. <https://doi.org/10.1093/brain/awx031>
- Petrosyan, A. (2024). *Number of internet and social media users worldwide as of January 2024 (in billions)*. Statista. <https://www.statista.com/statistics/617136/digital-population-worldwide/>. Accessed 22 Mar 22 2024.
- Radke, S., Jankowiak, K., Tops, S., Abel, T., Habel, U., & Derntl, B. (2021). Neurobiobehavioral responses to virtual social rejection in females—Exploring the influence of oxytocin. *Social Cognitive and Affective Neuroscience*, 16(3), 326–333. <https://doi.org/10.1093/scan/nsaa168>
- Rathje, S., Van Bavel, J. J., & Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26), e2024292118. <https://doi.org/10.1073/pnas.2024292118>
- Rein, B. (2022). Harnessing social media to challenge scientific misinformation. *Cell*, 185(17), 3059–3065. <https://doi.org/10.1016/j.cell.2022.07.001>
- Rein, B. (2023). Making science education more accessible: A case study of TikTok’s utility as a science communication tool. *Neuroscience*, 530, 192–200. <https://doi.org/10.1016/j.neuroscience.2023.08.021>
- Robertson, C. E., Shariff, A., & Van Bavel, J. J. (2024). Morality in the anthropocene: The perversion of compassion and punishment in the online world. *PNAS Nexus*, 3(6), page 193. <https://doi.org/10.1093/pnasnexus/pgae193>
- Sander, K., Frome, Y., & Scheich, H. (2007). fMRI activations of amygdala, cingulate cortex, and auditory cortex by infant laughing and crying. *Human Brain Mapping*, 28(10), 1007–1022. <https://doi.org/10.1002/hbm.20333>
- Sander, K., & Scheich, H. (2005). Left auditory cortex and amygdala, but right insula dominance for human laughing and crying. *Journal of Cognitive Neuroscience*, 17(10), 1519–1531. <https://doi.org/10.1162/089892905774597227>
- Scheggia, D., La Greca, F., Maltese, F., Chiacchierini, G., Italia, M., Molent, C., Bernardi, F., Coccia, G., Carrano, N., Zianni, E., Gardoni, F., Di Luca, M., & Papaleo, F. (2022). Reciprocal cortico-amygdala connections regulate prosocial and selfish choices in mice. *Nature Neuroscience*, 25(11), 1505–1518. <https://doi.org/10.1038/s41593-022-01179-2>
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7(3), 321–326. <https://doi.org/10.1089/1094931041291295>
- Sun, L., Lukkarinen, L., Putkinen, V., Karlsson, H. K., Hirvonen, J., Tiihonen, J., Lauerma, H., Scott, S., & Nummenmaa, L. (2022). Mu-opioid receptor system modulates responses to vocal bonding and distress signals in humans. *Philosophical Transactions of the Royal Society b: Biological Sciences*, 377(1863), 20210181. <https://doi.org/10.1098/rstb.2021.0181>
- Vanian J. (2023). Elon Musk’s X sued this nonprofit after it exposed hate speech, and its new research shows little has changed. *CNBC*. <https://www.cnn.com/2023/09/13/x-formerly-twitter-hate-speech-running-rampant-cddh.html>. Accessed 22 Mar 2024.
- Vanman, E. J. (2016). The role of empathy in intergroup relations. *Current Opinion in Psychology*, 11, 59–63. <https://doi.org/10.1016/j.copsyc.2016.06.007>
- Vogels, E. A. (2021). *The state of online harassment*. Pew Research Center. <https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/>. Accessed 22 Mar 2024.
- Whatman P. (2022). *Analysis of 11 Billion Mentions: Social Media is More Negative Than Ever*. <https://mention.com/en/blog/social-media-mentions-analysis/>. Accessed 22 Mar 2024.
- Wolf, W., Levordashka, A., Ruff, J. R., Kraaijeveld, S., Lueckmann, J.-M., & Williams, K. D. (2015). Ostracism Online: A social media ostracism paradigm. *Behavior Research Methods*, 47(2), 361–373. <https://doi.org/10.3758/s13428-014-0475-x>
- Wong, B. & Botorff, C. (2023). Top Social Media Statistics and Trends of 2024. *Forbes Advisor*. <https://www.forbes.com/advisor/business/social-media-statistics/>. Accessed 22 Mar 2024.
- Yammine, S. Z., Liu, C., Jarreau, P. B., & Coe, I. R. (2018). Social media for social change in science. *Science*, 360(6385), 162–163. <https://doi.org/10.1126/science.aat7303>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.