Article

Challenges for Research on Groups in Cyberspace

Small Group Research
2025, Vol. 56(3) 469–486

DOI: 10.1177/10464964251318625
journals.sagepub.com/home/sgr
Sage

Sara Kiesler[1]

Abstract

Throughout my career, advances in technology and global communications have enabled larger and more dispersed groups worldwide, new services and products that make groups more efficient, effective, and cohesive. As well, we see new threats and harms to group members and others: online harassment and hate speech, invasions of privacy, financial crimes, cyberbullying, AI-created false information, and threats to vulnerable group members. More than half of young women and girls survey state they have been harassed online and a quarter reported feeling unsafe (https://planusa-org-staging.s3.amazonaws.com/public/uploads/2021/04/Freedom_Online_2020_State_Of_The_Worlds_Girls_PIUSA.pdf). Health misinformation spread by online groups, such as that HIV does not cause AIDS, may have killed thousands in Africa. Misinformation promulgated by anti-vaccine groups has encouraged millions in the U.S. to forgo life-saving vaccines. Cyberspace today spills over into the physical world and all forms of telecommunication, furthering fraudulent services and goods, polarization of the electorate and distrust in elections, threats to journalism, misleading science, and real-world violence. High quality research and investigative journalism has made some gains in understanding and mitigating the tsunami of accidental and intentional harms online, but opposing lies and threats with truth can feel like a losing arms race. Understanding what causes these harms enabled by technology and tech-company business models, will rest in fundamental principles of group member psychology, group dynamics, and intergroup conflict. Ongoing and future group research could help us understand the benefits of new technology and counter negative group behaviors in the online world.

Keywords

knowledge management, social media, online groups, benefits and harms in cyberspace

[1]Carnegie Mellon University, Pittsburgh, PA, USA

Corresponding Author:
Sara Kiesler, Human Computer Interaction Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
Email: kiesler@andrew.cmu.edu

When I was a child, our nation's capital was segregated. My family lived in a whites-only apartment complex in northeast Washington D.C. Many black families lived in our neighborhood, shut out of apartment rentals and living in dank basements of the complex, huddled near the furnaces. I used to roam the complex and visit with the children of these families, but they were not allowed to visit me. These experiences, and many others like them, were a big reason I ended up a graduate student in social psychology.

The 1960s was an exciting time to study group research, especially to read about group beliefs and conflict: Sherif et al.'s (1961) work on young campers he separated and then brought together (demonstrating in-group solidarity vs. out-group hostility); Asch's (1956) experiments (demonstrating conformity with false judgements in groups), and Tajfel et al.'s (1971) seminal research on the power of group identify. In the 1950s, Professor Leon Festinger and his colleagues had carried out an ethnographic study of a small self-named group, the "Seekers," that predicted a great flood and the end of the world on December 21st. When the date arrived and the world did not end, the strongest believers redoubled their proselytizing. I was Leon's graduate student at Stanford and never forgot this example and his fascination with false beliefs in groups.[1]

When I began teaching at Carnegie-Mellon University, 99% of the world's population had never used a computer, much less to communicate with another person, but my university was at the forefront of the emerging technology. Two researchers, Hiltz and Turoff (1978), had recently published a seminal book called *Network Nation: Human Communication via Computer*. Their description of what they called "computer conferencing" prompted my first experiments in online group communication. I quote from my description of what ensued:

> [Our undergraduate assistant, Mina's] face-to-face sessions were fine but her computer chat sessions were a mess. One of her first groups was not able to reach consensus until 2am in the morning. Subjects were getting upset with each other, refusing to reach consensus. . .I asked Mina to follow a strict protocol and to wear more formal clothes so subjects would "take the experiment seriously." We continued to see problems in reaching consensus when participants used [online] chat. My colleague, Lee llll, suggested I talk with [computer science professor] Allen Newell. He told me, smiling, "Oh, that's just flaming." Jane [Siegel, my graduate student] decided to run the experiments herself, and we began to document carefully the processes that ensued when people had to reach consensus using the computer to communicate. (Kiesler, 2007, p. 5).

> With those early experiments began a new program of research, and our concerns about the emerging unintended consequences of online group

communication (e.g., Kiesler et al., 1984; McGuire et al., 1987). Early on, we attributed disinhibited online communication to the absence of in-person social context cues (Sproull & Kiesler, 1986), a theory that perhaps may have some merit but certainly does not explain everything that has ensued since.

Skip to today, and I am working at the National Science Foundation, managing projects in the (recently renamed) Secure and Trustworthy Cyberspace research grants program. We receive proposals from researchers in disciplines as disparate as communication, political science, and computer science who bring concepts from groups research to bear on societal challenges.[2]

Research on Cyberspace

Today's research environment contrasts with the situation, years ago, when to study groups, we were mostly confined to observing small numbers of people (often students) in "the lab." Or we might visit a factory floor or a company in person, observe a crowd, or interview members of the public. After the internet arrived, we could conduct online email or web surveys or experiments (some of them embedded in surveys). Technologies like email and texting on small devices became everyday tools for group research. Later it became possible to collect observational data on the interactions of online groups and teams and gather thousands of posted communications from a myriad of platforms such as Twitter (now X), Facebook, and Instagram. These technological and methodological breakthroughs have helped us understand, if incompletely, the nature of groups encountered in today's world, among them teams who share their work no matter their location, families, friends, and neighborhoods connected online, and huge online crowds following online influencers and comments posted on current events. Cyberspace supports online groups and organizations with a myriad of different common interests ranging from sports and games, fashion, gambling, pornography, education and training, hobbies like sewing and dog rescue and snake collecting, political change and advocacy, and a mix of criminal enterprises.

Computational technologies and advanced statistical methods today can be used to describe and predict online group behavior through crowd-sourced reports and observations of behavior in online social media. Remote sensors can be used to track such disparate phenomena such as people's travel, health, and street protests along with their online interactions. Machine learning, deep learning, and generative AI can be used to collect multimodal data on both online and offline group behavior over time. Researchers can track platform-supported social media and even interactions on the dark web; they can

use new methods, such as data fusion, data interoperability, and differential privacy for sharing data while maintaining individual privacy. Network scientists and computational linguists use ever-improving machine learning approaches, such as deep neural network models, to track the evolution and spread of manipulated information across social media. These methods also allow for creating synthetic data and simulations for training and testing the predictive computer models. Modern computational modeling and analysis techniques have brought about a greater interest by researchers in detecting and predicting behavior alongside methods for testing the mechanisms that explain social behavior (Yarkoni & Westfall, 2017).

Unfortunately, most studies of online behavior are based on biased samples of publicly available data. A major barrier to research on manipulated information and online maliciousness is that researchers lack access to important data that are protected by corporations and platforms' proprietary algorithms. This limitation is a key problem for research because a major way information spreads today is thought to be significantly influenced by these platforms' unobservable proprietary algorithms. Researchers who work outside platform organizations must use workaround methods with publicly available data.[3] The situation is slowly improving with European regulation and international efforts to require, and guide, researchers' access to platform data.

Misinformation in Groups

The wide distribution of misinformation through social media, and advances in the computational and social sciences, have motivated considerable research on misinformation in groups worldwide.[4] Information threat models and theories of abuse aim to detect and describe the sources of misinformation in cyberspace, and its recipients, strategies, and harms, including misleading claims intended to influence the public that are hidden in ordinary news and pseudo-scientific publications. The spread of false information carries through time, platform, social networks, and groups. Although the sources of misinformation are often hard to detect, computationally enabled investigative methods have identified hundreds of fabulist, extremist, and state-sponsored groups that originate misinformation. Emerging detection methods can identify false claims and misrepresentations of people in text, audio, images, video, or even virtual reality, and that spread through social networks and across groups. Advances in natural language processing and computational linguistics have led to better detection and understanding of the use of in-group code words,[5] humor, and symbolic visualizations to manipulate information in all languages. New methods are emerging for

ground truth evidence and cross-validation in studies of false news and rumor, manipulated information, and harmful outcomes (e.g., Jiang et al., 2021). This work has the potential to explain, predict, and perhaps prevent serious harms from manipulated information, such as hate crimes, consumer fraud, and suicide (Ophir et al., 2022).

In cyberspace, the flow of misinformation to new individuals and groups is often far out of proportion to the size of the originating group. Falsehoods may emerge first within a small online group, then spread rapidly to others. False claims that millions see have, in some cases, originated with a single group propagandist who posted hateful and conspiratorial narratives.[6] A story in the Washington Post describes a California woman who originated a social media account she called "Libs of TikTok." She gathered a huge audience (2.8 million on Twitter/X in February 2024), first by denying COVID, inventing stories of child sex trafficking, doubting the 2020 presidential election results, and urging the firing of gay teachers. Encouraged by online responses, she urged Oklahomans (a state she had visited just once) to pull books featuring LGBTQ people from school libraries. She has been "blamed for sparking bomb threats, property damage, shooting threats, written and verbal harassment and other forms of violence against individuals, hospitals and schools across the country."[7]

Some writers assert that only a small percentage of all information on television, radio, and online is false, but digital information architectures and algorithms encourage the rapid spread of false claims within groups and then to huge numbers of individuals, who in turn spread them further to members of other groups. Often a false claim refers to "we" in trying to attract a collective response, for example, "@Wendys, your beef is frozen and we know it. Y'all know we laugh. . ." Humor, sarcasm, and making fun of others can make these messages more sensational and entertaining, which in turn aids platforms' business models. During the 2016 presidential campaign, a theory that Podesta's hacked email contained coded messages of human trafficking and a child sex ring spread almost immediately through 4chan, 8chan, Reddit, and Twitter. Even when people do not believe such fictions, misinformation online can lead to public concern about the truth of what is available on TV, radio, and the internet. Perhaps it also leads to malaise, or a general feeling of "I give up trying to figure it out" such that significant portions of the public take whatever information they get at face value (Figure 1).

Disinformation

Some groups have been formed for the express purpose of spreading online propaganda and reinforcing conspiracy mindedness in online group members

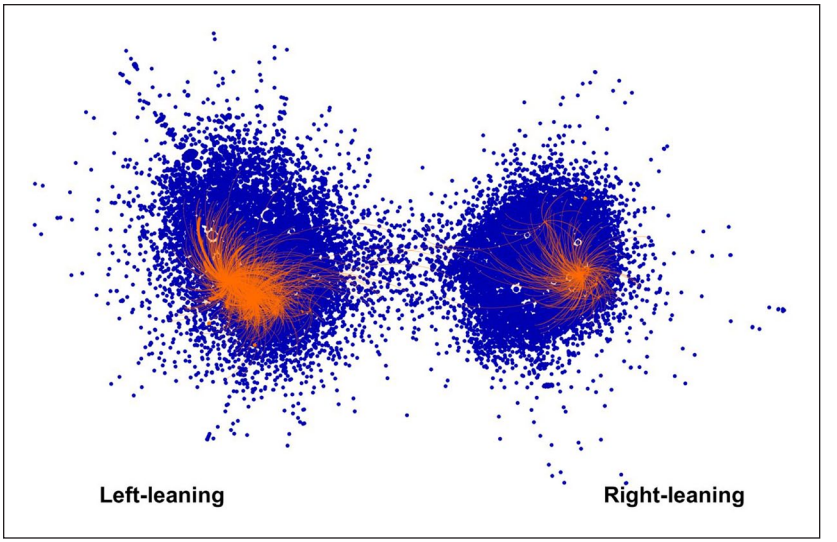

Figure 1. Tweets (blue) and retweets by RU-IRA (Russian) trolls that encourage polarization.
Source. From Stewart et al. (2018, Figure 2).

and vulnerable public audiences. These so-called disinformation campaigns use social engineering, including sophisticated advertising techniques and falsified content such as deepfakes and rumors hidden within correct news of real events, taking advantage of a distracted and uncritical public. In a longitudinal study of Twitter/X posts on reproduction, tweeted misinformation evolved to fit changes in the rest of the information environment. Pre-vaccines, tweets voiced conspiratorial and anti-vax rhetoric. When vaccines were imminent, tweets more frequently included fears about threats to pregnancy and fertility. After vaccines were distributed, first-person tweets reported menstrual cycle disruption (a real phenomenon that was harmless but not broadly addressed by public health authorities). These accounts were exploited in disinformation campaigns (Diamond et al., 2022; L. Palen, personal communication, 2024). It may be that sensational lies spread further and faster than truth, in part because surprising or emotionally charged news drives audience traffic, advertising, and sponsors (Silverman, 2015).

Detecting disinformation campaigns early enough to stop them before their messages spread across online platforms, and rapidly countering false narratives and providing trustworthy information, is challenging and, in some cases, has proved impossible. Researchers are studying the effectiveness of

fact checking and content moderation groups, artificial intelligence for anomaly detection, internet measurement to track diffusion, digital forensics to identify, expose and explain information manipulation campaigns, and labeling and nudging to flag inauthentic content (e.g., Kozyreva et al., 2020). There are ongoing efforts to make these techniques more scalable, interoperable, well-integrated, and applicable in multi-modal contexts. But often they are ineffective or arrive too late to halt false narratives and correct beliefs, as in the now long-standing certainty among many in the U.S. that the 2020 U.S. presidential election was stolen and that even the measles vaccine is harmful.

Group Identity and Members' Beliefs

The science of information manipulation is aimed in part at explaining and predicting group member beliefs in false information, especially when these beliefs lead to sustained social norms and behaviors that harm people's health and welfare or damage critical social institutions, as with fears of fraudulent elections and rejection of public health messages. There is a growing literature in the human structure of group beliefs, enabled by significant advances in methodologies for robust, reproducible science, computational tools, data science, and brain sciences. For example, neuroscientists have identified brain regions activated by misinformation that appeals to group moral values—the precuneus, medial prefrontal cortex (MFPC), and inferior parietal lobes, interconnected brain regions, often referred to as the Default Mode Network or DMN (Raichle, 2015). Analyses of huge datasets also point to cognitive dispositions that predict ideological, moralistic, and dogmatic preferences in groups (Zmigrod et al., 2021).

Controlled online experiments and large-scale data collection of information spread across online social networks suggest that partisanship and group loyalty are a major predictor of group members' misinformed beliefs. False beliefs can strengthen rather than weaken group ties; members copy the beliefs and behaviors of others in their group and those common beliefs, even if under attack, reinforce group allegiance (Mosleh et al., 2021). In one of our own studies, members of groups accepted the word of a collaborator, even if it was clearly wrong—a phenomenon we called "inaccuracy blindness" (Kane et al., 2018). Group identity can swamp other factors such as fact checking and source credibility. In one online game involving 540 people, members of an arbitrarily composed online group copied the preferences of their assigned group, overriding the effects of the reliability, warmth, or competence of others (Montrey & Shultz, 2022). The authors of this paper argue that such tendencies comprise fundamental processes that have, over centuries, reinforced cultural divergence and cumulative social evolution.

Cognitive and psychological scientists have been studying how long people remember false information, whether or not they have rejected it, and how sustained this rejection is (often, not very long). We still do not entirely understand why a considerable minority of the public (globally, not just in the U.S.) shares information they know is false, promulgates false information from questionable sources, and mistrusts the findings of science, medicine, and high-quality news sources. The reasons for these phenomena probably reside in many factors related to group identity: "facts" diverging from group ideology, vested group interests (including financial and reputational), and group-encouraged fear (e.g., Hornsey, 2020).

Polarization and Outgroup Hostility

Polarization of groups, including groups separated by political differences, socioeconomic status, or personal circumstance, has increased with the advent of online algorithms that point people to like-minded groups regardless of information quality. Platform algorithms and greater in-group similarity amplify the impact of poor information content and create more extreme and less politically diverse audiences (Bhadani et al., 2022; Kim et al., 2021). Researchers have begun to study the incentives and strategies of news organizations and platforms that increase polarization and poor-quality, divisive content (e.g., Acemoglu et al., 2021).

State-sponsored trolls have been used to polarize the electorate (Stewart et al., 2018; Zannettou et al., 2019). False claims sometimes reference strongly held moralized group beliefs that motivate allegiance and spread rapidly across sympathetic online social networks (e.g., Hoover et al., 2021; Moorijman et al., 2018). The strategy includes hiding polarizing content while seeming only to reinforce what people want to hear. False news creators, including teens, may be hired by organizations (foreign governments, unethical businesses, politicians). These individuals profit through monetary rewards and reputational benefits and do so with minimum effort or risk of punishment (Kshetri & Voas, 2017). The woman described above who spread falsehoods about LGBTQ people was given financial support by a like-minded sponsor, and was able to quit her job as a realtor and spend full time as an influencer. She told her interviewer that her online activism made her feel important.

A related phenomenon linked to polarization is the strengthening of so-called extremist groups that promulgate hatred, conspiracy theories, and dehumanization of other groups (see Harris & Fiske, 2006). Some researchers have studied how societal upheavals such as the 911 attacks on the World Trade Center, the recession in 2008, the 2019 viral pandemic, demographic

change, and increasing automation has encouraged out-group hostility and conspiracy theories that help explain people's anxiety and anger, and fix blame on other groups (Douglas et al., 2017; van Prooijen & Douglas, 2017). Moral rhetoric used in social media surrounding political demonstrations has been associated with violent protests (Gallacher et al., 2021; Hassan et al., 2018; Moorijman et al., 2018).

Changing Group Structures

With the advent of computer networking, group structures became more fluid than when groups mostly interacted in geographic proximity. At first, it was simply that groups of all kinds, including families and work teams and activists, were able to connect more easily with members who were geographically dispersed, thus including more peripheral members into their core functions (e.g., Hesse et al., 1993). As more people used internet communication, they could form dispersed groups, and larger teams and collaborations. Distributed teams could make teamwork more inclusive but also more challenging, such as for arranging meetings across time zones (e.g., Balakrishnan et al., 2011; Hinds & Mortensen, 2005). More recently, mass participation in social media and the internet globally seems to have encouraged the interconnection of small and large groups in overlapping spheres, embedding them in larger ideological, political, and cultural domains.

Many large organizations have one or more headquarters overseeing many online and geographically dispersed offices or teams that perform much of the core business of the organization. For instance, I belong to an organization that raises, trains, and gives service dogs to people with disabilities. Canine Companions (canine.org) is a nonprofit organization headquartered in California—with a mission and history, policies and leaders, but it is hierarchically connected, much through the internet, to regional and local groups all over the U.S. that carry out the mission. These groups are ideologically tied to the top but with local cultures and practices that influence how top-down guidelines and requests are met. Changes in the local environment can exert reverse influence on headquarters. Scholars have begun to study these kinds of organizational structures (e.g., Dolata & Schrape, 2016), raising many questions for group researchers, such as how different these structures are from one another, how they interact with, or threaten, traditional structures, how leadership maintains control, how groups change with challenges in the organizational environment, and effects on group members.

Changes in group structure online have had both beneficial and harmful effects. An example of the latter is described by historian, Belew (2018, 2023), who documents the spread of the white power movement with the

founding of Liberty Net in 1983 to 1984. Activists connected through Liberty Net distributed stolen funds and resources to white power groups around the U.S. About group structure, she says:

> We might consider these in concentric circles. Scholars who have attempted to estimate the size of the movement in the 1980s broadly agree that the innermost circle would include only 25,000 people but this inner ring would be quite dedicated. . . Outside this dedicated center, we would find another 150,000-175,000 activists: those who attend public actions, subscribe to movement literature, and similar actions. Then there is a larger circle of some 450,000 people who would not themselves subscribe to movement literature but who regularly read that literature all the same. (p. 257)

Belew argues that disagreement about which groups did what at the insurrection January 6, 2021 ignores that many of the groups that participated—the Proud Boys, Oath Keepers, Three Percenters, The Base, Boogaloo Boys, and Atomwaffen Division, among others—were part of the same white power movement, with years of experience organizing and connecting diverse and geographically dispersed groups, with Jan 6th aimed not just at exerting violent power in that moment in history, but also as a means to radicalize and draw in more groups.

Group Complacency (Groupthink, Social Loafing, and Free Riding)

We know that groups often reject, and defend against, outsiders' attacks on their practices and beliefs. But group members also sometimes resist dialogue and debate within their own group, particularly arguments that challenge group norms and comfortable social interactions, jeopardize members' social standing in the group, or threaten group leadership (Janis, 2008; McCauley, 1989). Perhaps it is merely that, in the online environment, group members are not paying sufficient attention to notice or counter false information (Kane et al., 2018; Kang et al., 2014).

With the advent of artificial intelligence (AI) that may take over some group tasks or supervision, whether, when, and why online group (or team) members fail to challenge misinformation inside the group becomes a new research challenge. Researchers are already speculating that social loafing (Latané et al., 1979) will increase when automated agents are considered teammates (Liu et al., 2023). AI can seriously interfere with group deliberations when the automated system fails, for example, when it mislabels a member as unproductive, or decides that the group is inefficient or is failing

to comply with requirements or recommends discriminatory decisions (e.g., Rosenthal-von der Pütten & Abrams, 2020). The data that AI used to generate these recommendations may consist of thousands (or millions) of data points that cannot be interpreted. For instance, imagine a court system where the judges rule unjustly that a defendant deserves incarceration or a large fine because of the predictions of a biased AI analyses of past records. These predictions may not be easily (or at all) interrogated. AI systems are meant to be autonomous, and if there are errors in design, or if the system experiences security breaches, or has software or hardware faults, these systems can have unforeseeable effects. Their recommendations cannot be made transparent in same way that human or traditional software decision making is transparent because AI models are always learning and adapting to new data. Thus, it may be impossible to reverse engineer the system to discover what mechanisms or factors led to the unforeseen or undesirable consequences. And because of this unpredictability and lack of transparency and explainability, the contestability of decisions may absent, as might blame and redress for harm (see Scherer, 2016).[8]

Interventions

Platforms and news organizations can remove misinformation before it is made public, or soon after, but few researchers have had access to this process. Once misinformation is made public, fact checkers and fact checking websites may be helpful to the public for evaluating the quality of information (Porter & Wood, 2021), but fact checking is an effortful task and consumes time during which manipulated information can spread widely. Recent research projects are aimed at aiding the early detection of false information and its provenance to increase the efficiency and value of fact checking and corrections of false information. A serious challenge to fact checking, however, is that even when it improves the accuracy of people's knowledge, it may have much less influence on people's beliefs, actions, and long-term reasoning ability. A standing research question is whether correcting false information aids misguided beliefs in the long run.

Two strategies that researchers are exploring are pre-emptive intervention (prebunking) and reactive intervention (debunking). Prebunking seeks to help individuals recognize and resist subsequently encountered misinformation, even if it is novel. For example, they might be warned to think about the accuracy of incoming information. Debunking responds to misinformation after exposure to demonstrate why it is false. The effectiveness of these corrections seems to depend on such factors as how quickly and well social media enable them, and which groups people belong to. In one study of

Facebook groups, Zollo et al. (2017) found that groups that discussed science (the so-called science echo chamber) commented positively on misinformation debunking posts, but groups inclined to entertain misinformation paid little attention to debunking: . . . "[O]nly few users usually exposed to unsubstantiated claims actively interact with the corrections. Dissenting information is mainly ignored." It seems clear that more must be done to understand and improve the impact of corrections, especially as applied to interactions in online groups globally and in different cultures (Ecker et al., 2022). Some researchers argue that information reliability is also a platform responsibility, and their algorithms ought to factor audience diversity (and thus higher news reliability) into their decisions to reinforce higher journalistic standards (Bhadani et al., 2022).

Informed and engaged citizens also need cognitive tools and knowledge to consider information critically. Several factors make this a challenge in the U.S. and internationally: the demise of local newspapers, the negative impact on schooling of the pandemic, the absence or poor quality of formal cognitive training (critical thinking skills), and formal civics, health, and science education in K-12 schools. New research strategies, using a range of evaluation approaches including high-quality random control experiments, are being created to teach people to reject false information, or to resist attempts at radicalization and violent extremism (Bilali, 2022). Other research examines better ways to communicate evidence and facts through easily comprehended visualizations (Franconeri et al., 2021) and "gist" communications connected to people's core values and groups (Reyna, 2021).

Conclusion

I have never forgotten a talk on technological change that Neil Postman gave in Denver in 1998, although I have indeed forgotten how I learned of it. The gist was that technological change is always sought in our society but comes with tradeoffs—benefits to some but costs to others—and carries with it cultural and ecological change, not just gains in productivity. It becomes mythic, as though technological progress arises naturally rather than is socially constructed. I highly recommend this talk in its entirety.[9] (See also the Wikipedia page on Neil Postman.) Postman didn't get everything right back in 1998, but he did understand that technological advance arrives not only with its intended improvements but also with unintended, sometimes indirect, consequences. I have focused in this brief essay on some of the unintended consequences of what we today experience as cyberspace. I have not been comprehensive, but it is my hope that group researchers join other scientists, professionals, journalists who are trying to tackle the challenges presented by

a cyberspace that is global and impactful, with spillover effects on the rest of our society. The aim we want is a world in which groups online and offline flourish while adhering to a system that allows people to express themselves while honoring the welfare and freedom of others.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Notes

1. The reason I ended up at third-choice Stanford? Harvard wrote that I was too young and a distinguished professor at Yale said the department did not accept women as graduate students or faculty, which partly explains my interest in perceptions and norms about women (e.g., Kiesler, 1975; Kiesler et al., 1985; Pheterson et al., 1971).
2. This article results in part from my participation in a U.S. interagency working group for "information integrity," which led in 2022 to a roadmap for future research (Information Integrity Interagency Working Group, 2022; Networking and Information Technology Research and Development Subcommittee, (2022). The working group was "sunsetted" in December 2024.
3. For an example of the difference access can make, see Kell et al. (2023).
4. See, for example, a special issue of Enserink (2022).
5. Some code words: aliens, big-pharma, black ops, bioengineering, chemtrails, corporate media, deep state, fake science, flat earth, false flag, George Soros, globalist, GMO, illuminati, media, New World order Cabal, nutritional supplements, pedophile rings, Rothchilds, vax, woke, Zionist, 9–11 truther, DEI hire.
6. "A QAnon con: How the viral Wayfair sex trafficking lie hurt real kids." Jessica Contrera, *Washington Post*, December 16, 2021.
7. Lorenz, T. How "Libs of TikTok" became a powerful presence in Oklahoma schools. *Washington Post*, February 4, 2024.
8. One recent example: A "fully-automated" Tesla crashed due to the system failing to see a stop sign, but the driver failed to collect damages in court because he was impaired (and that's why he was leaving the driving to the car). The system designers apparently did not consider the contexts in which drivers are likely to leave the driving to the car, and did not have a mechanism for assessing driver attention or competence.
9. https://www.cs.ucdavis.edu/~rogaway/classes/188/materials/postman.pdf

References

Acemoglu, D., Ozdaglar, A. E., & Siderius, J. (2021). *A model of online misinformation* (National Bureau of Economic Research Working Paper No. w28884). https://doi.org/10.3386/w28884

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Balakrishnan, A., Kiesler, S., Cummings, J., & Zadeh, R. (2011). *Research team integration: What it is and why it matters* [Conference session]. Proceedings of the ACM conference on computer supported cooperative work (pp. 523–532). ACM Press. https://doi.org/10.1145/1958824.1958905

Belew, K. (2018). *Bring the war home: The White Power movement and paramilitary America*. Harvard University Press. https://doi.org/10.4159/9780674984943

Belew, K. (2023). Insurrection. In K. M. Kruse & J. E. Zelizer (Eds.) *Myth America: Historians take on the biggest legends and lies about our past*. Basic Books.

Bhadani, S., Yamaya, S., Flammini, A., Menczer, F., Ciampaglia, G. L., & Nyhan, B. (2022). Political audience diversity and news reliability in algorithmic ranking. *Nature Human Behaviour*, *6*(4), 495–505. https://doi.org/10.1038/s41562-021-01276-5

Bilali, R. (2022). Fighting violent extremism with narrative intervention: Evidence from a field experiment in West Africa. *Psychological Science*, *33*(2), 184–195. https://doi.org/10.1177/09567976211031895

Diamond, L. L, Anderson, J., Batan, H., & Palen, L. (2022). *The polyvocality of online COVID-19 vaccine narratives that invoke medical racism* [Conference session]. CHI conference on human factors in computing systems (CHI '22), April 29–May 5, 2022, New Orleans, LA, USA. ACM Press. https://doi.org/10.1145/3491102.3501892

Enserink, M. (2022, March 24). Social media storm. *Science*, *375*(6587), 1334–1337. https://www.science.org/toc/science/375/6587

Dolata, U., & Schrape, J.-F. (2016). Masses, crowds, communities, movements: Collective action in the internet age. *Social Movement Studies*, *15*(1), 1–18. https://doi.org/10.1080/14742837.2015.1055722

Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The psychology of conspiracy theories. *Current Directions in Psychological Science*, *26*(6), 538–542. https://doi.org/10.1177/0963721417718261

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*, 13–29. https://doi.org/10.1038/s44159-021-00006-y

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. *Psychological Science in the Public Interest*, *22*(3), 110–161. https://doi.org/10.1177/1529100621105

Gallacher, J. D., Heerdink, M. W., & Hewstone, M. (2021). Online engagement between opposing political protest groups via social media is linked to physical

violence of offline encounters. *Social Media + Society*, *7*(1), 1–16. https://doi.org/10.1177/2056305120984445

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, *17*(10), 847–853. https://doi.org/10.1111/j.1467-9280.2006.01793.x

Hassan, G., Brouillette-Alarie, S., Alava, S., Frau-Meigs, D., Lavoie, L., Fetiu, A., Varela, W., Borokhovski, E., Venkatesh, V., Rousseau, C., & Sieckelinck, S. (2018). Exposure to extremist online content could lead to violent radicalization: A systematic review of empirical evidence. *International Journal of Developmental Science*, *12*(1–2), 71–88. https://doi.org/10.3233/DEV-170233

Hesse, B. W., Sproull, L. S., Kiesler, S. B., & Walsh, J. P. (1993). Returns to science: Computer networks in oceonography. *Communications of the ACM*, *36*(8), 90. https://doi.org/10.1145/163381.163409

Hiltz, R. S., & Turoff, M. (1978). *Network nation: Human communication via computer*. Addison-Wesley [Revised edition: MIT Press, 1993].

Hinds, P. J., & Mortensen, M. (2005). Understanding conflict in geographically distributed teams: The moderating effects of shared identity, shared context, and spontaneous communication. *Organization Science*, *16*(3), 290–307. https://doi.org/10.1287/orsc.1050.0122

Hoover, J., Atari, M., Mostafazadeh Davani, A., Kennedy, B., Portillo-Wightman Yeh, L, & Dehghani, M. 2021. Investigating the role of group-based morality in extreme behavioral expressions of prejudice. *Nature Communication*, *12*, 4585. https://doi.org/10.1038/s41467-021-24786-2

Hornsey, M. J. (2020). Why facts are not enough: Understanding and managing the motivated rejection of science. *Current Directions in Psychological Science*, *29*(6), 583–591. https://doi.org/10.1177/0963721420969364

Information Integrity Interagency Working Group. (2022). *Roadmap for researchers on priorities related to information integrity research and development*. Report available in the National Archives. https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/12/Roadmap-Information-Integrity-RD-2022.pdf

Janis, I. L. (2008). Groupthink. *IEEE Engineering Management Review*, *36*(1), 36. https://doi.org/10.1109/EMR.2008.4490137

Jiang, T., Li, J. P., Haq, A. U., Saboor, Q., & Ali, A. (2021) A novel stacking approach for accurate detection of fake news. *IEEE Access*, *9*, 22626–22639. https://doi.org/10.1109/ACCESS.2021.3056079

Kane, A., Kiesler, S., & Kang, R. (2018, April). *Inaccuracy blindness in collaboration persists, even with an evaluation prompt* [Conference session]. Proceedings of the 2018 CHI conference. April 21–26, 2018, Montreal, QC, Canada. ACM Press. https://doi.org/10.1145/3173574.317406

Kang, R., Kane, A., & Kiesler, S. (2014). *Teammate inaccuracy blindness: When information sharing tools hinder collaborative analysis* [Conference session]. CSCW ’14: Proceedings of the 17th ACM conference on computer supported cooperative work & social computing (pp. 797–806). ACM Press. https://doi.org/10.1145/2531602.2531681

Kell, T., Yousaf, H., Allen, S., Meiklejohn, S., & Juels, A. (2023). Forsage: Anatomy of a smart-contract pyramid scheme. In F. Baldimsti & C. Cahin (Eds.), *Financial cryptography and data security* (pp. 241–258). Springer. https://doi.org/10.1007/978-3-031-47751-5_14

Kiesler, S. (1975). Actuarial prejudice toward women and its implications. *Journal of Applied Social Psychology*, *5*, 201–216. https://doi.org/10.1111/j.1559-1816.1975.tb00676.x

Kiesler, S. (2007). Network nation: Human communication via computer. In T. Erickson & D. W. McDonald (Eds.), *HCI remixed: Reflections on works that have influenced the HCI community*. MIT Press. https://doi.org/10.7551/mitpress/7455.003.0017

Kiesler, S., Siegel, J. L., & McGuire, T. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, *39*, 1123–1134. https://doi.org/10.1037/0003-066X.39.10.1123

Kiesler, S., Sproull, L., & Eccles, J. (1985). Poolhalls, chips, and war games: Women in the culture of computing. *Psychology of Women Quarterly*, *9*, 451–462. https://doi.org/10.1111/j.1471-6402.1985.tb00895.x

Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity. *Journal of Communication*, *71*(6), 922–946. https://doi.org/10.1093/joc/jqab034

Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens versus the internet: Confronting digital challenges with cognitive tools. *Psychological Science in the Public Interest*, *21*(3), 103–156. https://doi.org/10.1177/1529100620946707

Kshetri, N., & Voas, J. (2017). The economics of "fake news." *IT Professional*, *19*(6), 8–12. https://doi.org/10.1109/MITP.2017.4241459

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, *37*(6), 822–832. https://doi.org/10.1037/0022-3514.37.6.822

Liu, M., Ke, W., & Xu, D. (2023). *Will humans be free-riders? The effects of expectations for AI on human-AI team performance* [Conference session]. 2023 Pacific Asia Conference on Information Systems (PACIS 2023). Nanchang, China, July, 2023. Association for Information Systems. https://aisel.aisnet.org/pacis2023/20

McCauley, C. (1989). The nature of social influence in groupthink: Compliance and internalization. *Journal of Personality and Social Psychology*, *57*(2), 250–260. https://doi.org/10.1037/0022-3514.57.2.250

McGuire, T., Kiesler, S., & Siegel, J. L. (1987). Group and computer-mediated discussion effects in risk decision making. *Journal of Personality and Social Psychology*, *52*(5), 917–930. https://doi.org/10.1037/0022-3514.52.5.917

Montrey, M., & Shultz, T. R. (2022). Copy the in-group: Group membership trumps perceived reliability, warmth, and competence in a social-learning task. *Psychological Science*, *33*(1), 165–174. https://doi/org/10.1177/09567976211032224

Moorijman, M., Hoover, J., Line, Y., Ji, H., & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behavior*, *2*, 289–396. https://doi.org/10.1038/s41562-018-0353-0

Mosleh, M., Martel, C., Eckles, D., & Rand, D. G. (2021). Shared partisanship dramatically increases social tie formation in a Twitter field experiment. *Proceedings of the National Academy of Sciences*, *118*(7), e2022761118. https://doi.org/10.1073/pnas.2022761118

Networking and Information Technology Research and Development Subcommittee. (2022). *Roadmap for researchers on priorities related to information integrity research and development*. https://www.whitehouse.gov/wp-content/uploads/2022/12/Roadmap-Information-Integrity-RD-2022.pdf

Ophir, Y., Tikochinski, R., Brunstein Klomek, A., & Reichart, R. (2022). The hitchhiker's guide to computational linguistics in suicide prevention. *Clinical Psychological Science*, *10*(2), 212–235. https://doi.org/10.1177/21677026211022013

Pheterson, G. I., Kiesler, S., & Goldberg, P. A. (1971). Evaluation of the performance of women as a function of their sex, achievement and personal history. *Journal of Personality and Social Psychology*, *19*, 114–118. https://doi.org/10.1037/h0031215

Porter, E., & Wood, T. (2021) The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proceedings of the National Academy of Sciences*, *118*(37), e2104235118. https://doi.org/10.1073/pnas.2104235118

Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience*, *38*, 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030

Reyna, V. F. (2021). A scientific theory of gist communication and misinformation resistance, with implications for health, education, and policy. *Proceedings of the National Academy of Sciences*, *118*(15), e1912441117. https://doi.org/10.1073/pnas.1912441117

Rosenthal-von der Pütten, A., & Abrams, A. M. H. (2020). Social dynamics in human-robot groups: Possible consequences of unequal adaptation to group members through machine learning in human-robot groups. In H. Degen & L. Reinerman-Jones (Eds.), *Artificial intelligence in HCI*. Springer. https://doi.org/10.1007/978-3-030-50334-5_27

Scherer, M. U. (2016). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law & Technology*, *29*(2), 353–400. http://dx.doi.org/10.2139/ssrn.2609777

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1961). *Intergroup conflict and cooperation: The Robbers Cave study*. University Book Exchange.

Silverman, C. (2015). *Lies, damn lies, and viral content* (pp. 4–169). Tow Center for Digital Journalism. https://doi.org/10.7916/D8Q81RHH

Sproull, L., & Kiesler, S. (1986). Reducing social context cues: Electronic mail in organizational communication. *Management Science*, *32*, 1492–1512. https://doi.org/10.1287/mnsc.32.11.1492

Stewart, L. G., Arif, A., & Starbird, K. (2018). *Examining trolls and polarization with a retweet network* [Paper presentation]. MIS2, 2018, Marina Del Rey, CA, USA. https://doi.org/10.475/123_4

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*, 149–178. https://doi.org/10.1002/ejsp.2420010202

van Prooijen, J.-W., & Douglas, K. M. (2017). Conspiracy theories as part of history: The role of societal crisis situations. *Memory Studies*, *10*(3), 323–333. https://doi.org/10.1177/1750698017701615

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Zannettou, S., Caulfield, T., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Blackburn, J. (2019). *Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web* [Conference session]. WWW '19: Companion proceedings of the 2019 world wide web conference (pp. 218–226). IEEE. https://doi.org/10.1145/3308560.331649

Zmigrod, L., Eisenberg, I. W., Bissett, P. G., Robbins, T. W., & Poldrack, R. S. (2021). The cognitive and perceptual correlates of ideological attitudes: A data-driven approach. *The Philosophical Transactions of the Royal Society*, *376*(1822), 20200424. https://doi.org/10.1098/rstb.2020.0424

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., Havlin, S., & Quattrociocchi, W. (2017). Debunking in a world of tribes. *PLoS One*, *12*(7), e0181821. https://doi.org/10.1371/journal.pone.0181821

Author Biography

Sara Kiesler is University Professor and Hillman Chair Emerita of Human-Computer Interaction and Computer Science at Carnegie Mellon University and Program Director at the National Science Foundation in the Social, Behavioral, and Economics Directorate of the National Science Foundation. She has a Ph.D. in social psychology. Her research has spanned multiple disciplines and domains including studies of online communication and group behavior.