International Journal of Communication 16(2022), 1–20 1932–8036/20220005

The SPIR Framework of Social Media and Polarization: Exploring the Role of Selection, Platform Design, Incentives, and Real-World Context

ELIZABETH HARRIS
New York University, USA

STEVE RATHJE
University of Cambridge, UK

CLAIRE E. ROBERTSON
JAY J. VAN BAVEL[1]
New York University, USA

Because of the rapid growth of social media, nearly 4 billion people now have online accounts where they engage with their social network, learn about the news, and share content with other people. The rapid growth of this technology has raised important questions about its potential impact on political action and polarization. We propose a framework to address how Selection, Platform Design, Incentives, and Real-World Context (the SPIR framework) might explain social media's role in exacerbating polarization and intergroup conflict. Rather than simply asking whether social media as a whole causes polarization, we examine how each of these processes can spur polarization in certain contexts. Specifically, we explain how these features of social media can act as an accelerant, amplifying divisions in society between social groups and spilling over into offline behavior. We discuss how interventions might target each of these factors to mitigate (or enhance) polarization.

Keywords: polarization, social media, intergroup conflict, Internet

Elizabeth Harris: eah561@nyu.edu
Steve Rathje: sr6276@nyu.edu
Claire E. Robertson: crobertson@nyu.edu
Jay J. Van Bavel: jay.vanbavel@nyu.edu
Date submitted: 2021-11-17

[1] We are grateful for support from the John Templeton Foundation to Jay J. Van Bavel, a Gates Cambridge Scholarship awarded to Steve Rathje (Grant No. OPP1144), and the Social Sciences and Humanities Research Council of Canada doctoral fellowship to Elizabeth Harris (Grant No. 752-2018-0213). Elizabeth Harris and Steve Rathje are co-first authors.

Because of the rapid growth of social media, nearly 4 billion people now have online accounts where they engage with their social network, learn about the news, and share content with other people (Statista, 2020). The rapid growth of this recent technology has raised important questions about its potential impact on political action, collective behavior, and polarization (see Bak-Coleman et al., 2021; Van Bavel, Rathje, Harris, Robertson, & Sternisko, 2021). Several scholars have argued that social media has democratized political discourse, fostered social justice, and facilitated revolution (Eltantawy & Wiest, 2011; Jost et al., 2018; Tufekci & Wilson, 2012). However, there is now a growing body of evidence that social media may contribute to polarization, political violence, and hate crimes (Allcott, Braghieri, Eichmeyer, & Gentzkow, 2020; Müller & Schwarz, 2020). The current article discusses how specific features of social media may spur polarization, and offers insights into potential solutions to reduce this influence.

Polarization

There are two core forms of polarization: (1) affective polarization and (2) ideological polarization. Although these two forms of polarization are often highly interwoven, they are conceptually distinct. Affective polarization refers to the divide between groups characterized by negative feelings toward out-group members (Iyengar, Lelkes, Levendusky, Malhotra, & Westwood, 2019). Over the past few decades, affective polarization, as measured by the number of Americans with "very unfavorable" attitudes toward their political out-party, has been increasing (Pew Research Center, 2014). Ideological polarization refers to the divide between groups in terms of values and beliefs (Iyengar et al., 2019), and has been increasing as well (Pew Research Center, 2014).

The literature has looked at both forms of polarization; therefore, we review evidence of both. However, affective polarization appears to be more pernicious than ideological polarization and is associated with democratic backsliding, or a decline in democratic principles of governance (Orhan, 2022), reduced support for democracy (Kingzette et al., 2021), and lower intellectual humility (Bowes, Blanchard, Costello, Abramowitz, & Lilienfeld, 2020). Therefore, our article focuses, wherever possible, on affective polarization and other pernicious forms of political conflict, such as political sectarianism (Finkel et al., 2020).[2]

It is important to note that social conflict and moral outrage can be the function of sincere political disagreements and, in some contexts, can foster important social change (Spring, Cameron, & Cikara, 2018); however, there is a growing trend in the United States and several other nations toward out-group hate and false polarization (i.e., misrepresentations of the beliefs of out-groups; see Brady & Crockett, 2019; Finkel et al., 2020). Although polarization research often focuses on political parties, it may also focus on conflicts between other ethnic, religious, sectarian, or national groups within society. In the current article, we discuss how social media platforms can exacerbate polarization and intergroup conflict within societies.

[2] Most of the literature studying the link between social media and polarization is correlational. More work needs to be done to explore and establish the causal relationship between social media and polarization.

Social Media

Here, we define "social media" as "a group of Internet-based applications that . . . allow the creation and exchange of User Generated Content" (Kaplan & Haenlein, 2010, p. 28). Different social media platforms have different features and norms, and certain aspects of our framework may therefore apply to some platforms but not others. For example, although WhatsApp may contribute to the spread of misinformation and polarization content (Machado, Kira, Narayanan, Kollanyi, & Howard, 2019), it does not have an algorithmic recommendation system like Facebook, Twitter, or Instagram. The platform LinkedIn shares many of the features of Facebook, but it is less relevant to the issue of polarization because of different platform norms (e.g., of professionalism) and its focus on career issues and networking as opposed to society and politics.

Additionally, the distinction between different social media platforms, as well as between social media and traditional forms of media, has become increasingly blurred in recent years. For instance, content from one social network site is often reposted to other social network sites and covered by legacy media sites. Many mainstream journalists also promote their work and find content for stories from social media. Although each unique social media site has its own idiosyncrasies, most of the literature to this point has focused on data from Twitter and Facebook (Blank & Lutz, 2016); therefore, most work we cite here focuses on those platforms. It will be important for future work to examine these processes across other platforms.

The SPIR Model

There is a growing interest in social media and polarization (see Kubin & von Sikorski, 2021; Prior, 2013; Tucker et al., 2018). We build on this prior work by developing a model that addresses how and when social media might increase polarization. Specifically, we review the role of four factors: Online *Selection*, *Platform Design*, *Incentive* Structures, and *Real-World* Context (SPIR) in polarizing the public and shaping offline behavior. Prior work has described how these factors can facilitate the spread of moral content online (see the MAD Model; Brady, Crockett & Van Bavel, 2020). We expand on this work by applying these concepts to political conflict and polarization. Specifically, we explain how these features of social media can act as an accelerant, amplifying divisions in society between social groups and spilling over into offline behavior. To help make sense of these processes, we propose a framework for understanding how social media can polarize people.

According to the *SPIR framework* (as seen in Figure 1), social media users often seek out content relevant to their identities and beliefs. These selections, in turn, increase the probability that users interact with people and content that amplifies or reinforces their identities, values, and beliefs—which can potentially trigger radicalization (Atari et al., 2021). Some platforms have algorithms that may amplify political content that is polarizing or hostile to increase the engagement of users (Protecting Kids Online, 2021). This provides an incentive for users (as well as political elites, news agencies, and foreign actors) to use language and other content that attracts attention and reinforcement on the platform (Simchon, Brady, & Van Bavel, 2022). All of this unfolds in a broader context, which includes both the norms specific to the social media platform—and, more narrowly, the social network of the user—and can influence offline behavior.

It is difficult to discern the causal impact of social media on polarization (Van Bavel, Rathje, et al., 2021). For example, polarization was increasing in many countries before the rise of social media and is currently not increasing in every country that does have access to social media (Boxell, Gentzkow, & Shapiro, 2022). Although one experiment found that deleting Facebook led to decreases in polarization in the United States (Allcott et al., 2020), other research has found that such results might depend on one's offline social network (Asimovic, Nagler, Bonneau, & Tucker, 2021). There are also many different ways of using social media, from professional to political, and the consequences of social media use may depend on features like a particular platform's design or one's offline context. Thus, rather than answering the simple question of whether social media as a whole causes polarization, we aim to examine the interrelated processes by which social media might spur polarization and intergroup conflict.

Our view is that social media is no longer distinct from other modes of communication; our different modes of communication are intertwined. According to a recent survey, reporters said their primary source for news was online (58%) and many others said Twitter (16%), with cable or TV news trailing in third place (7%; Frank, 2021). Political elites, news agencies, and journalists not only use social media to find content for reporting but also use these platforms to disseminate the news and build their professional profile, and therefore may be motivated to present the news in a way that elicits online engagement (e.g., appealing to the norms of a platform or leveraging the algorithms). This makes it difficult to fully disentangle the impact of mainstream media and social media on polarization. Moreover, actions in the real world can reinforce polarization on social media—creating a vicious cycle. Accordingly, it is important to study social media to understand the full picture of how media affects polarization and conflict.

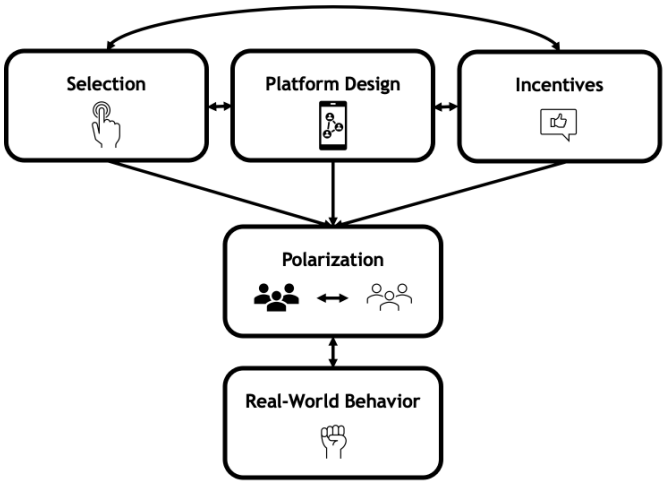

Figure 1. Our framework suggests that people Select identity-congruent news and social networks. The Platform Design and algorithms on social media influence people's online behavior and the type of content that people see. Social media's business model of rewarding viral content may provide Incentives for the creation of divisive content. All of these features interact with the Real-World Contexts and offline social networks that people are embedded within to facilitate polarization.

Selection and Sorting

With the advent of the Internet, and social media more specifically, there is now an overwhelming amount of information available to people at all times. On YouTube alone, users are uploading 500 hours of video per minute. It would take more than 80 years to watch a single day's worth of new video content (Hale, 2019). By one account, social media users scroll through 300 feet of newsfeed per day (Qin, n.d.). As such, it is critical to understand how people sort through this information and what consequence this sorting process has on their beliefs and behavior.

Experimental work suggests that people often seek out information that is congruent with what they already believe (Knobloch-Westerwick & Meng, 2009)—known as "selective exposure" (Frey, 1986; although see Nelson & Webster, 2017, who argue that selective exposure may be weaker than initially thought). In the realm of political (mis)information, this is the tendency for people to predominantly read news that is in agreement with their political beliefs (Knobloch-Westerwick & Meng, 2009). This bias further extends to the source of the message. For instance, when Americans seek out news online, they read news from sources aligned with their political identity and beliefs, and this tendency is increasing over time (Rodriguez, Moskowitz, Salem, & Ditto, 2017). Similarly, the political lean of media outlets is reflected in the political leaning of their Twitter account followers (Golbeck & Hansen, 2011). Crucially, increased selective exposure to political news is correlated with political polarization—and this relationship is potentially bidirectional (Stroud, 2010).

In addition to seeking out and selecting congruent information, people are also predisposed to believe congruent information (see Van Bavel & Pereira, 2018). In one series of experiments, researchers explored how participants differed in their belief of news stories that were partisanship congruent (i.e., positive about their political in-group or negative about their political out-group) or incongruent (Pereira, Harris, & Van Bavel, 2021). The results revealed greater belief in partisan congruent information (see also Jennings & Stroud, 2021). Additionally, when people see a correction to a piece of political misinformation online, they are more likely to update their belief (i.e., believe it less than before the correction) when the correction is politically congruent (i.e., asymmetric updating; Jennings & Stroud, 2021; Sunstein, Bobadilla-Suarez, Lazzaro, & Sharot, 2016). The impact of partisanship on belief may be largest for ambiguous information, and people likely behave more rationally when the information is unambiguous (see Xiao, Coppin, & Van Bavel, 2016).

Further, social media users may follow or self-select into online networks that are composed of other users with the same political identity. For example, people are three times more likely to follow back Twitter users who share their partisan identity (Mosleh, Martel, Eckles, & Rand, 2021). One study found that Facebook users in Italy are ideologically polarized and form two homophilous groups (Bessi et al., 2015). Given that people are also more inclined to share news that is congruent with their politics on social media platforms (Pereira et al., 2021; Shin & Thorson, 2017), this would increase exposure to identity-congruent content. Increasing polarization may be related to these tendencies toward increased seeking, sharing, belief in, and belief updating for politically congruent content, as well as tendencies to self-select into identity-congruent networks.

Platform Design and Algorithms

Social media platforms are not created equally—it is important to differentiate between social media platforms and their distinct features (e.g., populations, social norms, social-feedback dynamics, algorithms). Some platforms appear far more likely to increase polarization than others. For example, researchers observed polarizing social dynamics on Facebook and Twitter but not on Reddit, Gab, and WhatsApp (Cinelli, De Francisci Morales, Galeazzi, Quattrociocchi, & Starnini, 2021; Yarchi, Baden, & Kligler-Vilenchik, 2021). Different platforms also appear to foster different types of polarization. For example, Facebook has been linked to attitudinal polarization (the extremity of citizens' political opinions; Levendusky, 2013), whereas Twitter has been linked to both affective and attitudinal (i.e., how citizens feel about and evaluate political parties) polarization (Levy, 2021; Yarchi et al., 2021). As such, distinguishing between different platforms is likely critical to understanding their role in polarization.

One important platform design feature is the newsfeed algorithm, which determines the content users see when they use the platform. Multiple social media sites appear to operate by showing their users more content that is congruent with their political beliefs. For instance, watching algorithm-recommended YouTube videos on partisan issues increased participants' polarization, particularly when the algorithm was based on their own search preferences (Cho, Ahmed, Hilbert, Liu, & Luu, 2020). Similarly, Facebook appears to infer its users' political ideology and shapes their newsfeed to be politically congruent (Levy, 2021), and the TikTok algorithm appears to send people down "rabbit holes" on the app (Wall Street Journal Staff, 2021). For one investigation, researchers created TikTok "bot" accounts that would rewatch videos with specific hashtags. One bot account was assigned interests in "sadness" and "depression" and would rewatch videos that had any hashtags related to those topics. The algorithm quickly discovered these interests, and soon sent this bot down a "rabbit hole" of depression-related videos, to the point where 93% of this bot's recommended videos were about depression (Wall Street Journal Staff, 2021). As such, it is possible that these algorithms or "filter bubbles" have the capacity to amplify polarization (Spohr, 2017). Indeed, being part of morally homogenous social networks increases radical intentions and willingness to fight and die for one's group (Atari et al., 2021). Thus, social media's tendency to show people information congruent with their beliefs—especially their moral beliefs—may increase political and sectarian conflict.

There is debate, however, over the extent to which online polarization is platform-driven versus user-driven. In other words, it is difficult to determine whether and when people self-select into online networks that contain polarizing content, or whether social media algorithms amplify and encourage people to view polarizing content. For instance, one article found that while engagement with right-wing and "anti-woke" content (e.g., content critical of modern liberal activism) was increasing on YouTube, this did not appear to be driven by the YouTube algorithm recommending increasingly radical content. Instead, it seemed to be driven by several other forces, such as preferences of Internet users as a whole and demand on the broader Internet (Hosseinmardi et al., 2021). This contrasts with the common narrative that people fall down YouTube "rabbit holes" that show them increasingly extreme content. Another study found that while political polarization increased on Reddit, this was largely because of an influx of conservative Reddit users in 2016—in other words, this polarization appeared to be largely driven by users, as opposed to Reddit "polarizing" users (Waller & Anderson, 2021).

Although, without access to data about how social media algorithms operate, it is difficult to make strong claims about the extent to which algorithms play a role in polarizing individuals. Thus, many inferences about the internal dynamics of different algorithms are largely speculative or rely on indirect evidence, which is compounded by the problem that algorithms are continually being updated. More work in this area—in addition to further visibility into algorithm design—is needed to examine the content and impact of these algorithms over time.

While there is a broad consensus that exposure to too much political congruent content might be problematic, it is not clear how exposure to incongruent content might help. For instance, simply exposing individuals to diverse partisan sources of information does not necessarily reduce polarization. One field experiment paid Democrats and Republicans to follow Twitter accounts that retweeted messages by elected officials and opinion leaders with opposing political views for one month (Bail et al., 2018). Surprisingly, exposure to members of the other party increased ideological polarization (although this backlash effect was only significant among Republicans). This highlights another possible process by which social media can increase ideological polarization: as social media tends to amplify extreme viewpoints (Bail, 2021; Rathje, Van Bavel, & van der Linden, 2021), exposure to hyperpartisans from the in-group or out-group may lead people to become even more entrenched in their own viewpoint. However, other research suggests these "backlash" or "boomerang" effects are relatively rare (Casas, Menchen-Trevino, & Wojcieszak, 2022). Thus, more research is needed to examine the circumstances under which contact with opposing viewpoints is productive.

Social media's platform design also allows political actors to foment political conflict by deploying automated users—known as "bots." Bots are user accounts that present themselves as being real users, attempting to influence other users' opinions (Yan, Yang, Menczer, & Shanahan, 2020). They are present in online communities for various topics, such as the vaccination debate (Yuan, Schuchard, & Crooks, 2019) and discussion of international conflicts in India (Neyazi, 2020) on Twitter. Users can then be further exposed to hyperpartisan (mis)information through bots or trolls, which tend to use polarized rhetoric and content (Simchon et al., 2022). Indeed, research studying the influence of bot accounts suggests that they increase polarization on Twitter (Ozer, Yildirim, & Davulcu, 2019). This body of work suggests that social media algorithms (and other platform features) and bots may further amplify polarization.

Incentive Structures and Message Content

Social media platforms also seem to reward certain types of political rhetoric. For instance, divisive social media messages are more likely to succeed online. A recent analysis of 3 million social media posts found that posts about the political out-group (often reflecting out-group animosity) were much more likely to be shared than those about the political in-group. Each additional out-group word (e.g., "liberal," if the post came from a Republican) increased a posts' shares by approximately 67% and also strongly increased the likelihood of that post receiving "angry" reactions, "haha" reactions, and comments on Facebook (Rathjeet al., 2021). Relatedly, content expressing moral outrage is more likely to be shared on Twitter, especially within—and not between—partisan echo chambers (Brady, Wills, Jost, Tucker, & Van Bavel, 2017). Additionally, positive social feedback (e.g., likes, shares) on posts expressing outrage increases the likelihood that people will express outrage in the future (Brady, McLoughlin, Doan, & Crockett, 2021). Furthermore, the most popular content on

Facebook tends to consist of right-wing, hyperpartisan media sources (e.g., Ben Shapiro), which may be more likely to express outrage and out-group animosity (Thompson, 2020).

On some platforms, misinformation can receive more engagement than true information. For instance, one study found that false news was more likely to be shared than true news on Twitter (Vosoughi, Roy, & Aral, 2018), and this was especially true of political misinformation. The popularity of misinformation may be closely related to affective polarization (or out-party animosity). For instance, a recent study found that the strongest psychological predictor of sharing fake news on Twitter was affective polarization—perhaps because fake news often derogates the out-party (Osmundsen, Bor, Vahlstrup, Bechmann, & Petersen, 2021). Thus, the popularity of misinformation might be related to the general motivation to share content online that denigrates out-group members (Pereira et al., 2021; Rathje et al., 2021).

Divisive content may succeed online because it is particularly likely to capture our attention (Brady, Gantman, & Van Bavel, 2020). Since social media operates as an attention economy (Bak-Coleman et al., 2021), whereby users compete for the chances to go "viral," writing socially divisive social media posts that fulfill identity-based motivations (such as out-group derogation) may be an effective way for capturing attention and engagement. Indeed, one study found that the most politically extreme politicians have the most followers (Hong & Kim, 2016). In other words, the social media incentive structure may be creating social and economic incentives for producing and sharing polarizing content (Bail, 2021; Rathje et al., 2021).

While divisive posts and misinformation might generate engagement in the short term (and thus revenue for social media companies and enterprising users) they may have harmful side effects in the long term, including polarization. Researchers have proposed models through which misinformation increases ideological polarization (e.g., Au, Ho, & Chiu, 2021). Further, survey experiments find that people do not like the expression of partisan animus (Costa, 2020), even though this is what social media platforms appear to be incentivizing. Thus, social media platforms may be keeping people engaged by featuring content that they do not truly enjoy. Facebook recently chose to reduce the amount of political content in people's newsfeeds after discovering that, although it led to increased engagement, survey data revealed that people did not like it (Gupta, 2021).

Real-World Behavior

Social Media Activity Has Offline Consequences

Behavior on social media can have far-reaching offline consequences. Nefarious movements on the fringes of the political spectrum have originated online, giving voice to conspiracy theories and hate groups (Douglas et al., 2019). Recently, the U.S. conspiracy theory group QAnon has gained massive online popularity and may now encompass as many as 30 million followers (Russonello, 2021). QAnon's online rhetoric bled into offline spaces in January 2021, when there was an insurrection at the U.S. Capitol committed by people who believed the false claim propagated by conspiracy theorists that the U.S. presidential election had been fraudulent (Luke, 2021). Twitter use has also been linked to an increase in hate crimes at the community level (Hoover et al., 2021; Müller & Schwarz, 2020). For instance, anti-Muslim tweets from Donald Trump during his presidency were associated with an increase in hate crimes in the following days.

Unfortunately, examining causal ties between online and offline behavior is challenging. Because of the incredible number of social changes that occurred in parallel with the development and adoption of social media and the Internet broadly, most work relating online and offline behavior is correlational (Jost et al., 2018). Because the outcomes of interest are protests, civic engagement, violence, or even revolutions, the level of experimental control necessary for causal claims is difficult to achieve. Research using quasiexperiments, qualitative data, and archival research are useful to understand these important phenomena. However, more work using experimental manipulations and interventions is needed to draw causal conclusions.

Although our article has focused largely on the drawbacks of social media, it is important to note that social media can also have positive societal impacts. The ability to communicate critical protest information rapidly and broadly using social media has been associated with increases in democratic action and protest behavior across the world (González-Bailón, Borge-Holthoefer, Rivero, & Moreno, 2011; Jost et al., 2018). The Arab Spring, for example, relied heavily on social media's ability to rapidly coordinate protest information, call for aid, and amplify voices of dissent (Eltantawy & Wiest, 2011). Even adjusting for other factors such as age and sex, those who used social media were much more likely to attend the first day of protests than those who did not use social media (Tufekci & Wilson, 2012). Social media was also a source of information for the #MeToo and Black Lives Matter movements (Cox, 2017), and the subsequent Black Lives Matter protests in 2020 may have been the largest protests in U.S. history (Bolsover, 2020; Buchanan, Bui, & Patel, 2020). The impact of social media or other technologies likely hinges on the political context in which it exists. This is why we propose that social media is more of an accelerant for polarization, rather than its cause.

The Consequences of Social Media Activity Are Context Dependent

Social media's effect on polarization may also be moderated by one's offline social network. In Bosnia and Herzegovina, researchers randomly assigned participants to delete their Facebook accounts during genocide remembrance week (Asimovic et al., 2021). The effect of social media exposure on out-group ethnic regard depended on the diversity of participants' offline social networks. After a week without social media, participants reported lower ethnic out-group regard than those who had not deactivated, but only if their offline social network was homogenous. This finding suggests that offline behavior may improve because of social media when the political and ethnic makeup of one's online social networks are more diverse compared with their offline social networks.

Other research suggests that social media may have more beneficial outcomes in less established democracies, facilitating political protests and access to political news, but may have more harmful outcomes in more established democracies, such as the United States (Lorenz-Spreen, Oswald, Lewandowsky, & Hertwig, 2021). However, since most research on social media is conducted in the global North (Ghai, Magis-Weinberg, Stoilova, Livingstone, & Orben, 2022), more research is needed on the causal effect of social media usage around the globe. This should be a priority for future work in this area.

Many factors also appear to moderate the relationship between social media use and real-world behavior. The 2021 insurrection highlights the importance of examining individual differences, since not every person who is exposed to or believes in particular online content participates in related offline behavior (Arceneaux et al., 2021). As mentioned above, polls suggest that approximately 15% of Americans (~30 million

people) believe in the core tenets of QAnon (Russonello, 2021). However, only several thousand people were present during the capitol insurrection on January 6 (Doig, 2021). Although belief in QAnon conspiracies is relatively common, behavioral action related to QAnon has thus far been rare. However, not everyone in a society *needs* to be radicalized to create a polarized society. For example, the most hostile individuals online also tend to be similarly hostile offline, but because social media affords them more visibility, they have disproportionate influence online and can create a hostile or polarized environment for many people (Bor & Petersen, 2021).

Practical Applications and Interventions

In addition to informing theoretical accounts of social media and polarization, the *SPIR framework* can be applied practically to design interventions to mitigate polarization on social media. For instance, some interventions can target *Selection*. While social media recommendation algorithms are thought to contribute to polarization by sorting people into networks of like-minded others (Santos, Lelkes, & Levin, 2021), changes can be made to discourage this type of sorting. In fact, new laws have been recently introduced in U.S. Congress, such as the Filter Bubble Transparency Act (2021), with the goal of making social media algorithms more transparent or giving people the choice to disable algorithmic social media feeds. Researchers should also examine the impact of simply disabling algorithms (e.g., Twitter provides this option, while Facebook does not).

Other interventions can alter the social media *Platform Design* to reduce the spread of polarizing, false, or hostile content. For instance, an intervention deployed by Twitter asking people to revise potentially offensive content before they posted it reduced the overall amount of offensive Twitter users posted by 6% (Katsaros, Yang, & Fratamico, 2021). Similar interventions have found that empathic appeals (Hangartner et al., 2021) or warnings about the consequences of hate speech (Yildirim, Nagler, Bonneau, & Tucker, 2021) can reduce hate speech on Twitter. Additionally, interventions that correct misperceptions about the opposing party (Ruggeri et al., 2021) or create a sense of shared identity (Levendusky, 2018) have been effective in reducing polarization and can potentially be integrated into the social media platform design.

Interventions can also target *Incentives*. Providing people with financial incentives to accurately identify true and false news can substantially reduce the partisan divide in belief in political news (Rathje, Van Bavel, & van der Linden, 2022). However, incentivizing people to think about whether news headlines will be liked by one's in-group increases intentions to share politically congruent news—even if the news is false. These incentives do not necessarily need to be monetary. For instance, simply signaling that fellow in-group members find a post misleading can make people less likely to share it (Pretus et al., 2021). Thus, social media platforms might be able to shift incentive structures to decrease users' motivations to post polarizing or misleading content that may receive high engagement online, and instead increase users' motivations to post accurate content.

Additionally, interventions can target the incentive structures driving creators to post polarizing content. Some online content creators appear to have strong incentives to post false, polarizing, or conspiratorial content; for instance, prominent conspiracy theorist Alex Jones made more than 100 million dollars selling supplements to his followers (Dreisbach, 2021). More recently, many of the top shared

substack accounts were spreading anti-vax misinformation and profiting from doing so (Dwoskin, 2022). However, interventions can try to decrease incentives to post this kind of content. For instance, one field experiment found that sending state legislators letters warning them of the consequences of making false statements led these legislators to subsequently make fewer false statements (Nyhan & Reifler, 2015). Further, one case study looked at the deplatforming of three public figures involved in offensive speech (Alex Jones, Milo Yiannopoulos, and Owen Benjamin) and found that once the figures were deplatformed, they were discussed less on the platform and their supporters' activity and post toxicity decreased (Jhaver, Boylston, Yang, & Bruckman, 2021). Content moderation can also be highly effective for reducing the spread of false news online (Papakyriakopoulos, Serrano, & Hegelich, 2020), and the risk of content moderation can potentially discourage creators from producing false or hateful content.

Finally, social media interventions need to take the *Real-World* Context into account. Interventions likely have different effects—there is no obvious "one size fits all" solution. For instance, anti-misinformation interventions "nudging" social media users to share more accurate content are less effective for right-wing audiences (Rathje, Roozenbeek, Steenbuch, Van Bavel, & van der Linden, 2022) or participants highly aligned with Trump (Pretus et al., 2021). As such, some interventions may be effective only for certain demographic groups. Interventions aimed at decreasing the spread of false and polarizing content online should be rigorously tested with diverse samples and need to consider the Real-World Context before being effectively applied at scale.

Conclusion

Our article summarizes a growing literature on the impact of social media on intergroup polarization. Unfortunately, more work needs to be done to fully understand the impact of social media on polarization, especially in cultural contexts outside of the United States. In addition to polarization—and related threats to democracy—we have urgent concerns about the role of social media in the spread of misinformation (Van Bavel, Harris, et al., 2021; van der Linden et al., 2021). Given these widespread negative consequences, the study of social media should constitute a "crisis discipline," with a focus on providing actionable insight to policymakers and regulators for the stewardship of social systems (Bak-Coleman et al., 2021). Our *SPIR framework*, which highlights the role of *Selection*, *Platform Design*, *Incentive* Structures, and *Real-World* Context in polarization, can be used to inform both theoretical accounts of social media's relationship to polarization and real-world solutions for reducing polarization on social media.

References

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review, 110*(3), 629–676. doi:10.1257/aer.20190658

Arceneaux, K., Gravelle, T. B., Osmundsen, M., Petersen, M. B., Reifler, J., & Scotto, T. J. (2021). Some people just want to watch the world burn: The prevalence, psychology and politics of the "need for chaos." *Philosophical Transactions of the Royal Society B, 376*(1822), 1–9. doi:10.1098/rstb.2020.0147

Asimovic, N., Nagler, J., Bonneau, R., & Tucker, J. A. (2021). Testing the effects of Facebook usage in an ethnically polarized setting. *Proceedings of the National Academy of Sciences of the United States of America, 118*(25), 1–9. doi:10.1073/pnas.2022819118

Atari, M., Davani, A. M., Kogon, D., Kennedy, B., Ani Saxena, N., Anderson, I., & Dehghani, M. (2021). Morally homogeneous networks and radicalism. *Social Psychological and Personality Science,* 1–11. doi:10.1177/19485506211059329

Au, C. H., Ho, K. K., & Chiu, D. K. (2021). The role of online misinformation and fake news in ideological polarization: Barriers, catalysts, and implications. *Information Systems Frontiers,* 1–24. doi:10.1007/s10796-021-10133-9

Bail, C. (2021). *Breaking the social media prism*. Princeton, NJ: Princeton University Press.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., . . . Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(37), 9216–9221. doi:10.1073/pnas.1804840115

Bak-Coleman, J. B., Alfano, M., Barfuss, W., Bergstrom, C. T., Centeno, M. A., Couzin, I. D., . . . Weber, E. U. (2021). Stewardship of global collective behavior. *Proceedings of the National Academy of Sciences of the United States of America, 118*(27), 1–10. doi:10.1073/pnas.2025764118

Bessi, A., Petroni, F., Del Vicario, M., Zollo, F., Anagnostopoulos, A., Scala, A., . . . Quattrociocchi, W. (2015, May). Viral misinformation: The role of homophily and polarization. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 355–356). New York, NY: Association for Computing Machinery. doi:10.1145/2740908.2745939

Blank, G., & Lutz, C. (2016, July). The social structuration of six major social media platforms in the United Kingdom: Facebook, LinkedIn, Twitter, Instagram, Google+ and Pinterest. In A. Gruzd, J. Jacobson, P. Mai, E. Ruppert, & D. Murthy (Eds.), *Proceedings of the 7th 2016 International Conference on Social Media & Society* (pp. 1–10). New York, NY: Association for Computing Machinery. doi:10.1145/2930971.2930979

Bolsover, G. (2020). *Black Lives Matter discourse on US social media during COVID: Polarised positions enacted in a new event*. *SSRN*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3688909

Bor, A., & Petersen, M. B. (2021). The psychology of online political hostility: A comprehensive, cross-national test of the mismatch hypothesis. *American Political Science Review, 116*(1) 1–18. doi:10.1017/S0003055421000885

Bowes, S. M., Blanchard, M. C., Costello, T. H., Abramowitz, A. I., & Lilienfeld, S. O. (2020). Intellectual humility and between-party animus: Implications for affective polarization in two community samples. *Journal of Research in Personality, 88*, 1–12. doi:10.1016/j.jrp.2020.103992

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2022). Cross-country trends in affective polarization. *The Review of Economics and Statistics,* 1–60. doi:10.1162/rest_a_01160

Brady, W., & Crockett, M. J. (2019). How effective is online outrage? *Trends in Cognitive Sciences, 23*(2), 79–80. doi:10.1016/j.tics.2018.11.004

Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The MAD model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science, 15*(4), 978–1010. doi:10.1177/1745691620917336

Brady, W. J., Gantman, A. P., & Van Bavel, J. J. (2020). Attentional capture helps explain why moral and emotional content go viral. *Journal of Experimental Psychology: General, 149*(4), 746–756. doi:10.1037/xge0000673

Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances, 7*(33), 1–14. doi:10.1126/sciadv.abe5641

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America, 114*(28), 7313–7318. doi:10.1073/pnas.1618923114

Buchanan, L., Bui, Q., & Patel, J. K. (2020, July 3). Black Lives Matter may be the largest movement in U.S. history. *The New York Times*. Retrieved from https://www.nytimes.com/interactive/2020/07/03/us/george-floyd-protests-crowd-size.html

Casas, A., Menchen-Trevino, E., & Wojcieszak, M. (2022). Exposure to extremely partisan news from the other political side shows scarce boomerang effects. *Political Behavior,* 1–40. doi:10.1007/s11109-021-09769-9

Cho, J., Ahmed, S., Hilbert, M., Liu, B., & Luu, J. (2020). Do search algorithms endanger democracy? An experimental investigation of algorithm effects on political polarization. *Journal of Broadcasting & Electronic Media, 64*(2), 150–172. doi:10.1080/08838151.2020.1757365

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences of the United States of America, 118*(9), 1–8. doi:10.1073/pnas.2023301118

Costa, M. (2020). Ideology, not affect: What Americans want from political representation. *American Journal of Political Science, 65*(2), 342–358. doi:10.1111/ajps.12571

Cox, J. M. (2017). The source of a movement: Making the case for social media as an informational source using Black Lives Matter. *Ethnic and Racial Studies, 40*(11), 1847–1854. doi:10.1080/01419870.2017.1334935

Doig, S. (2021, January 8). It is difficult, if not impossible, to estimate the size of the crowd that stormed Capitol Hill. *The Conversation*. Retrieved from http://theconversation.com/it-is-difficult-if-not-impossible-to-estimate-the-size-of-the-crowd-that-stormed-capitol-hill-152889

Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., & Deravi, F. (2019). Understanding conspiracy theories. *Political Psychology, 40*(S1), 3–35. doi:10.1111/pops.12568

Dreisbach, T. (2021, March 24). Alex Jones still sells supplements on Amazon despite bans from other platforms. *NPR*. Retrieved from https://www.npr.org/2021/03/24/979362593/alex-jones-still-sells-supplements-on-amazon-despite-bans-from-other-platforms

Dwoskin, E. (2022, January 27). Conspiracy theorists, banned on major social networks, connect with audiences on newsletters and podcasts. *The Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2022/01/27/substack-misinformation-anti-vaccine/

Eltantawy, N., & Wiest, J. B. (2011). Social media in the Egyptian revolution: Reconsidering resource mobilization theory. *International Journal of Communication, 5*, 1207–1224. Retrieved from https://ijoc.org/index.php/ijoc/article/viewFile/1242/597

Protecting kids online: Testimony from a Facebook whistleblower: Hearings from the U.S. Subcommittee on Consumer Protection, Product Safety, and Data Security, 117th Cong. (2021). (Testimony of Frances Haugen).

Filter Bubble Transparency Act, S. 2024, 117th Cong. (2021).

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., . . . Druckman, J. N. (2020). Political sectarianism in America. *Science, 370*(6516), 533–536. doi:10.1126/science.abe1715

Frank, S. (2021, March 30). *How 2,400 journalists use social media for reporting* [and why PR should get serious about social] [Blog post]. Retrieved from https://www.swordandthescript.com/2021/03/journalists-social-media/

Frey, D. (1986). Recent research on selective exposure to information. *Advances in Experimental Social Psychology, 19*, 41–80. doi:10.1016/S0065-2601(08)60212-9

Ghai, S., Magis-Weinberg, L., Stoilova, M., Livingstone, S., & Orben, A. (2022). Social media and adolescent well-being in the Global South. *Current Opinion in Psychology, 46*, 1–7. doi:10.1016/j.copsyc.2022.101318

Golbeck, J., & Hansen, D. (2011, May). Computing political preference among Twitter followers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1105–1108). New York, NY: Association for Computing Machinery. doi:10.1145/1978942.1979106

González-Bailón, S., Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2011). The dynamics of protest recruitment through an online network. *Scientific Reports, 1*, 1–7. doi:10.1038/srep00197

Gupta, A. (2021, February 10). *Reducing political content in news feed* [About Facebook page]. Retrieved from https://about.fb.com/news/2021/02/reducing-political-content-in-news-feed/

Hale, J. (2019, May 7). More than 500 hours of content are now being uploaded to YouTube every minute. *Tubefilter*. Retrieved from https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute/

Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., . . . Jochum, M. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences of the United States of America, 118*(50), 1–3. doi:10.1073/pnas.2116310118

Hoover, J., Atari, M., Mostafazadeh Davani, A., Kennedy, B., Portillo-Wightman, G., Yeh, L., & Dehghani, M. (2021). Investigating the role of group-based morality in extreme behavioral expressions of prejudice. *Nature Communications, 12*, 1–13. doi:10.1038/s41467-021-24786-2

Hong, S., & Kim, S. H. (2016). Political polarization on Twitter: Implications for the use of social media in digital governments. *Government Information Quarterly, 33*(4), 777–782. doi:10.1016/j.giq.2016.04.007

Hosseinmardi, H., Ghasemian, A., Clauset, A., Mobius, M., Rothschild, D. M., & Watts, D. J. (2021). Examining the consumption of radical content on YouTube. *Proceedings of the National Academy of Sciences of the United States of America, 118*(32), 1–8. doi:10.1073/pnas.2101967118

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science, 22*(1), 129–146. doi:10.1146/annurev-polisci-051117-073034

Jennings, J., & Stroud, N. J. (2021). Asymmetric adjustment: Partisanship and correcting misinformation on Facebook. *New Media & Society,* 1–21. doi:10.1177/14614448211021720

Jhaver, S., Boylston, C., Yang, D., & Bruckman, A. (2021). Evaluating the effectiveness of deplatforming as a moderation strategy on Twitter. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW2), 1–30. doi:10.1145/3479525

Jost, J. T., Barberá, P., Bonneau, R., Langer, M., Metzger, M., Nagler, J., . . . Tucker, J. A. (2018). How social media facilitates political protest: Information, motivation, and social networks. *Political Psychology, 39*(S1), 85–118. doi:10.1111/pops.12478

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons, 53*(1), 59–68. doi:10.1016/j.bushor.2009.09.003

Katsaros, M., Yang, K., & Fratamico, L. (2021). *Reconsidering tweets: Intervening during tweet creation decreases offensive content*. Retrieved from https://arxiv.org/abs/2112.00773

Kingzette, J., Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2021). How affective polarization undermines support for democratic norms. *Public Opinion Quarterly, 85*(2), 663–677. doi:10.1093/poq/nfab029

Knobloch-Westerwick, S., & Meng, J. (2009). Looking the other way: Selective exposure to attitude-consistent and counterattitudinal political information. *Communication Research, 36*(3), 426–448. doi:10.1177/0093650209333030

Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association, 45*(3), 188–206. doi:10.1080/23808985.2021.1976070

Levendusky, M. (2013). Partisan media exposure and attitudes toward the opposition. *Political Communication, 30*(4), 565–581. doi:10.1080/10584609.2012.737435

Levendusky, M. S. (2018). Americans, not partisans: Can priming American national identity reduce affective polarization? *The Journal of Politics, 80*(1), 59–70. doi:10.1086/693987

Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review, 111*(3), 831–870. doi:10.1257/aer.20191777

Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2021). *Digital media and democracy: A systematic review of causal and correlational evidence worldwide*. Retrieved from https://osf.io/preprints/socarxiv/p3z9v/

Luke, T. W. (2021). Democracy under threat after 2020 national elections in the USA: "Stop the steal" or "give more to the grifter-in-chief?" *Educational Philosophy and Theory,* 1–8. doi:10.1080/00131857.2021.1889327

Machado, C., Kira, B., Narayanan, V., Kollanyi, B., & Howard, P. (2019, May). A study of misinformation in WhatsApp groups with a focus on the Brazilian Presidential Elections. In L. Liu & R. White (Eds.), *Companion Proceedings of the 2019 World Wide Web Conference* (pp. 1013–1019). New York, NY: Association for Computing Machinery. doi:10.1145/3308560.3316738

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 of the United States of America, 118*(7), 1–3. doi:10.1073/pnas.2022761118

Müller, K., & Schwarz, C. (2020). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association, 19*(4), 2131–2167. doi:10.1093/jeea/jvaa045

Nelson, J. L., & Webster, J. G. (2017). The myth of partisan selective exposure: A portrait of the online political news audience. *Social Media + Society, 3*(3), 1–13. doi:10.1177/2056305117729314

Neyazi, T. A. (2020). Digital propaganda, political bots and polarized politics in India. *Asian Journal of Communication, 30*(1), 39–57. doi:10.1080/01292986.2019.1699938

Nyhan, B., & Reifler, J. (2015). The effect of fact-checking on elites: A field experiment on US state legislators. *American Journal of Political Science, 59*(3), 628–640. doi:10.1111/ajps.12162

Orhan, Y. E. (2022). The relationship between affective polarization and democratic backsliding: Comparative evidence. *Democratization, 29*(4), 714–735. doi:10.1080/13510347.2021.2008912

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *American Political Science Review, 115*(3), 999–1015. doi:10.1017/S0003055421000290

Ozer, M., Yildirim, M. Y., & Davulcu, H. (2019). Measuring the polarization effects of bot accounts in the US gun control debate on social media. In *Proceedings of ACM Conference (Conference '17)*. New York, NY: Association for Computing Machinery. Retrieved from https://artisinternational.org/wp-content/uploads/2019/08/Measuring_the_Polarization_Effects_of_Bot_Accounts_in_the_U_S__Gun_Control_Debate_on_Social_Media.pdf

Papakyriakopoulos, O., Serrano, J. C. M., & Hegelich, S. (2020). The spread of COVID-19 conspiracy theories on social media and the effect of content moderation. *The Harvard Kennedy School (HKS) Misinformation Review, 1*, 1–19. doi:10.37016/mr-2020-034

Pereira, A., Harris, E. A., & Van Bavel, J. J. (2021). Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news. *Group Processes & Intergroup Relations*, 1–24. doi:10.1177/13684302211030004

Pew Research Center. (2014, June 12). *Political polarization in the American public: How increasing ideological uniformity and partisan antipathy affect politics, compromise and everyday life*. Retrieved from https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/

Pretus, C., Van Bavel, J. J., Brady, W. J., Harris, E. A., Vilarroya, O., & Servin, C. (2021). *The role of political devotion in sharing partisan misinformation*. Retrieved from https://psyarxiv.com/7k9gx/

Prior, M. (2013). Media and political polarization. *Annual Review of Political Science, 16*(1), 101–127. doi:10.1146/annurev-polisci-100711-135242

Qin, L. (n.d.). *How many miles will you scroll?* [Blog post]. Retrieved from https://www.leozqin.me/how-many-miles-will-you-scroll/

Rathje, S., Roozenbeek, J., Steenbuch, C., Van Bavel, J. J., & van der Linden, S. (2022). Letter to the editors of Psychological Science: Meta-analysis reveals that accuracy nudges have little to no effect for US conservatives: Regarding Pennycook et al. (2020). *Psychological Science*. doi:10.25384/SAGE.12594110.v2

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 of the United States of America, 118*(26), 1–9. doi:10.1073/pnas.2024292118

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2022, January 31). *Accuracy and social motivations shape judgements of (mis)Information*. Retrieved from https://psyarxiv.com/hkqyv/

Rodriguez, C. G., Moskowitz, J. P., Salem, R. M., & Ditto, P. H. (2017). Partisan selective exposure: The role of party, ideology and ideological extremity over time. *Translational Issues in Psychological Science, 3*(3), 254–271. doi:10.1037/tps0000121

Ruggeri, K., Većkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., . . . Bursalıoğlu, A. (2021). The general fault in our fault lines. *Nature Human Behaviour, 5*(10), 1369–1380. doi:10.1038/s41562-021-01092-x

Russonello, G. (2021, May 27). QAnon now as popular in U.S. as some major religions, poll suggests. *The New York Times*. Retrieved from https://www.nytimes.com/2021/05/27/us/politics/qanon-republicans-trump.html

Santos, F. P., Lelkes, Y., & Levin, S. A. (2021). Link recommendation algorithms and dynamics of polarization in online social networks. *Proceedings of the National Academy of Sciences of the United States of America, 118*(50), 1–9. doi:10.1073/pnas.2102141118

Shin, J., & Thorson, K. (2017). Partisan selective sharing: The biased diffusion of fact-checking messages on social media. *Journal of Communication, 67*(2), 233–255. doi:10.1111/jcom.12284

Simchon, A., Brady, W. J., & Van Bavel, J. J. (2022). Troll and divide: The language of online polarization. *PNAS Nexus, 1*(1), 1–12. doi:10.1093/pnasnexus/pgac019

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review, 34*(3), 150–160. doi:10.1177/2F0266382117722446

Spring, V. L., Cameron, C. D., & Cikara, M. (2018). The upside of outrage. *Trends in Cognitive Sciences, 22*(12), 1067–1069. doi:10.1016/j.tics.2018.09.006

Statista. (2020, July). *Number of social network users worldwide from 2017 to 2025 (in billions)* [Chart]. Retrieved from https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication, 60*(3), 556–576. doi:10.1111/j.1460-2466.2010.01497.x

Sunstein, C. R., Bobadilla-Suarez, S., Lazzaro, S. C., & Sharot, T. (2016). How people update beliefs about climate change: Good news and bad news. *Cornell Law Review, 102*(6), 1431–1444. doi:10.2139/ssrn.2821919

Thompson, A. (2020, September 26). Why the right wing has a massive advantage on Facebook. *Politico*. Retrieved from https://www.politico.com/news/2020/09/26/facebook-conservatives-2020-421146

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., . . . Nyhan, B. (2018). *Social media, political polarization, and political disinformation: A review of the scientific literature*. The Hewlett Foundation. doi:10.2139/ssrn.3144139

Tufekci, Z., & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of Communication, 62*(2), 363–379. doi:10.1111/j.1460-2466.2012.01629.x

Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K., & Tucker, J. A. (2021). Political psychology in the digital (mis) information age: A model of news belief and sharing. *Social Issues and Policy Review, 15*(1), 84–113. doi:10.1111/sipr.12077

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences, 22*(3), 213–224. doi:10.1016/j.tics.2018.01.004

Van Bavel, J. J., Rathje, S., Harris, E., Robertson, C., & Sternisko, A. (2021). How social media shapes polarization. *Trends in Cognitive Sciences, 25*(11), 913–916. doi:10.1016/j.tics.2021.07.013

van der Linden, S., Roozenbeek, J., Maertens, R., Basol, M., Kácha, O., Rathje, S., & Traberg, C. S. (2021). How can psychological science help counter the spread of fake news? *The Spanish Journal of Psychology, 24*, 1–9. doi:10.1017/SJP.2021.23

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. doi:10.1126/science.aap9559

Wall Street Journal Staff. (2021, July 21). Inside TikTok's algorithm: A WSJ video investigation. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/tiktok-algorithm-video-investigation-11626877477

Waller, I., & Anderson, A. (2021). Quantifying social organization and political polarization in online platforms. *Nature, 600*(7888), 264–268. doi:10.1038/s41586-021-04167-x

Xiao, Y. J., Coppin, G., & Van Bavel, J. J. (2016). Perceiving the world through group-colored glasses: A perceptual model of intergroup relations. *Psychological Inquiry, 27*(4), 255–274. doi:10.1080/1047840X.2016.1199221

Yan, H. Y., Yang, K.-C., Menczer, F., & Shanahan, J. (2020). Asymmetrical perceptions of partisan political bots. *New Media & Society, 23*(10), 3016–3037. doi:10.1177/1461444820942744

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2021). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication, 38*(1–2), 98–139. doi:10.1080/10584609.2020.1785067

Yildirim, M. M., Nagler, J., Bonneau, R., & Tucker, J. A. (2021). Short of suspension: How suspension warnings can reduce hate speech on Twitter. *Perspectives on Politics,* 1–13. doi:10.1017/S1537592721002589

Yuan, X., Schuchard, R. J., & Crooks, A. T. (2019). Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. *Social Media + Society, 5*(3), 1–12. doi:10.1177/2056305119865465