Title: The SPIR Model of Social Media and Polarization:
Exploring the role of Selection, Platform Design, Incentives, and Real World Context

Authors:
Elizabeth Harris*1
Steve Rathje*3
Claire E. Robertson1
Jay J. Van Bavel12

Affiliations:
1Department of Psychology, New York University
2Department of Psychology, New York University
[3]Department of Psychology, University of Cambridge, Cambridge CB2 3RQ, United Kingdom

Harris, E. A., Rathje, S., Robertson, C., & Van Bavel, J. J. (in press). The SPIR Model of Social Media and Polarization: Exploring the role of Selection, Platform Design, Incentives, and Real World Context. *International Journal of Communication.*

Abstract (91/150)
Due to 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 the SPIR Model to address how Selection, Platform Design, Incentives, and Real World Context 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. We discuss how interventions might target each of these factors to mitigate polarization.

Keywords: Polarization, social media, intergroup conflict

The SPIR Model of Social Media and Polarization: Exploring the role of Selection, Platform Design, Incentives, and Real World Context

Due to 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 technology has raised important questions about its potential impact on political action and polarization (see Van Bavel et al., 2021). A number of scholars have argued that social media has democratized political discourse, fostered social justice, and facilitated revolution (Jost et al. 2018 Eltantawy & Wiest, 2011;Tufekci & Wilson, 2012). Altneriatively, there is a growing body of evidence that social media has facilitated foreign political interference, protest violence, hate crimes, and genocide (Müller & Schwarz, 2020; Mooijman et al. 2018). The current paper discusses how specific features of social media may spur polarization (i.e., the division of a population into two sharply contrasting groups or sets of beliefs) and offers insights into potential solutions to reduce this influence.

Although social conflict and moral outrage can be the function of sincere political disagreements and foster important social change (Spring et al., 2018), there is a growing trend in the US and 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 polarized groups are often characterized as political parties, they may also be other ethnic, religious, sectarian, or national groups within society. In the current paper, we discuss how social media platforms can facilitate polarization and intergroup conflict within societies. Specifically, we review the role of Online *Selection*, *Platform* Design, *Incentive* Structures, and *Real-World* Context (SPIR) in polarizing the public and shaping offline behavior. We explain how these features of social media can create divisions in society between political groups and spillover 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 seek out content that reflects their identities and beliefs. These selections, in turn, increase the probability they interact with people and content that amplifies or reinforces their prior identities and beliefs. The design of the platform affords users the opportunity to efficiently signal their identity and beliefs to attract other followers. Some platforms also have algorithms that amplify political content that is polarizing or hostile to increase the engagement of users. 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. All of this unfolds in a broader social context, which includes both the norms specific to the social media platform--and, more narrowly, the social network of the user--as well as the real-world context.

It is difficult to discern the causal impact of social media on polarization (see Van Bavel et al., 2021). For example, polarization was increasing in many countries before the rise of social media, and is not increasing in every country (Boxell et al., 2020). While one experiment has found that deleting Facebook led to decreases in polarization in the United States (Allcott et al., 2020), other research has found that results like these might depend on one's social context or their offline social network (Asimovic et al., 2021). There are many different ways of using social media, and the consequences depend on features like its platform design or one's offline social context. Thus, rather than answering the question "does social media cause

polarization?" we aim to examine the processes by which social media can spur polarization and intergroup conflict in different contexts.

Our view is that social media is no longer distinct from other modes of communication. For instance, political elites, news agencies and journalists not only use social media to find content for reporting, but they also use these platforms to disseminate the news and build their professional profile. As such, they 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 nearly impossible 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. As such, understanding social media will henceforth be critical to understanding the dynamics of political polarization and intergroup conflict.
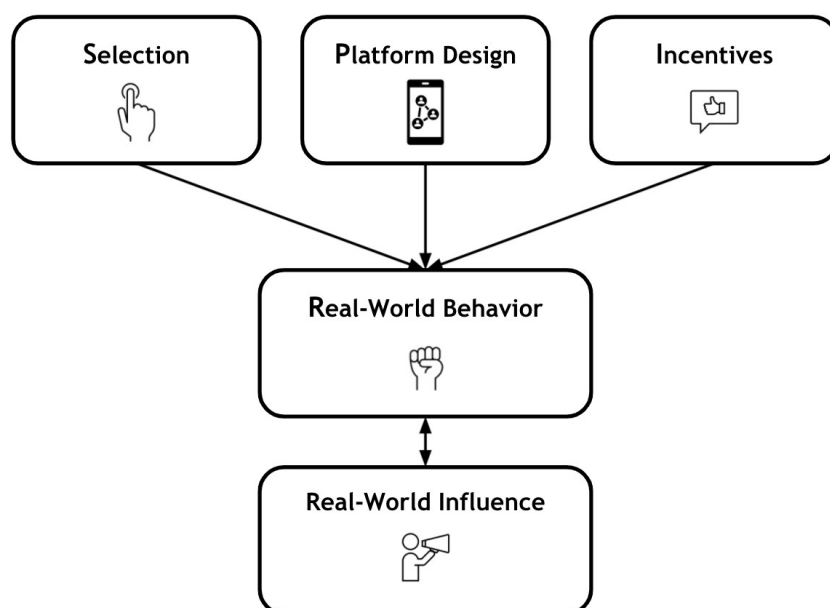
Figure 1. Our model suggests that people *Select* identity-congruent news and social networks. The *Platform-Design* and algorithms on social media influence people's online behavior and the type of content that people see. Social media's business model of rewarding viral content may provide *Incentives* 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 & Sorting

With the advent of the internet, and social media more specifically, there is an overwhelming amount of information available to people at all times. On Youtube alone, users are uploading 500 hours of video per minute, which means it would take over 80 years to watch one day's worth of new video content (Hale, 2019). By one account, social media users scroll through 300 feet of news feed per day. As such, it's critical to understand how people sort through this information.

Experimental work suggests that people often seek out information that is congruent with what they already believe (Knobloch-Westerwick & Meng, 2009; although see Nelson & Webster, 2017)--known as "selective exposure" (Frey, 1986). 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. When Americans seek out news online, for instance, they read news from sources aligned with their political identity and beliefs, and this tendency is only increasing over time (Rodriguez et al., 2017). Increased selective exposure to political news is correlated with political polarization (possibly in a bidirectional causal manner; Stroud, 2010). 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).

In addition to seeking out and selecting this information, people are also predisposed to believe it (i.e., believing political identity congruent information more than political identity incongruent information; See Van Bavel & Periera, 2018). In one series experiments, researchers explored how participants differed in their belief of news stories that were partisanship congruent (i.e., positive about their political ingroup or negative about their political outgroup) or incongruent (Pereira et al., 2021). The results suggest 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; Sunstein et al., 2016; Jennings & Stroud, 2021). People are also three times more likely to follow back Twitter users who share their partisan identity (Mosleh et al., 2021). Increasing polarization may be related to these tendencies for increased seeking, sharing, belief in, and belief updating for politically congruent content.

Platform Design & Algorithms

Social media platforms are not the same. We differentiate between social media platforms and their distinct features (e.g., populations, social norms, social-feedback dynamics, algorithms, etc.). Some platforms appear 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 et al., 2021; Yarchi et al., 2021). There are also different types of polarization across platforms. For example, Facebook has been linked to attitudinal polarization (the extremity of citizens' political opinions; Levendusky, 2013), whereas Twitter has been linked to both attitudinal and affective (how citizens feel about and evaluate the political parties) polarization (Levy, 2021; Yarchi et al., 2021).

One important platform design feature is the news feed 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 et al., 2020). Similarly, Facebook appears to infer its users' political ideology and shapes their newsfeed to be politically congruent (Levy, 2021). As such, these algorithms have the capacity to amplify polarization. Similarly, the TikTok algorithm appears to send people down "rabbit holes" on the app. For this 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 re-watch 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).

There is debate, however, over the extent to which online polarization is algorithm-driven versus user-driven. For instance, one paper found that while engagement with right-wing and "anti-woke" content was increasing on Youtube, this did not appear to be driven by the Youtube algorithm (Hosseinmardi et al., 2021). 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, any inferences about the internal dynamics of different algorithms is speculative.

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 polarization (although this backfire effect was only significant among Republicans). This highlights another possible process by which social media can increase polarization: as social media tends to amplify extreme viewpoints (Bail, 2021; Rathje et al., 2021), exposure to hyper-partisans from the out-group may lead people to become even more entrenched in their own viewpoint.

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 et al., 2020). They are present in online communities for a variety of 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 hyper-partisan (mis)information through bots (Simchon et al., 2021). 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 moral outrage, echo chambers, and polarization.

Incentive Structures & 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 outgroup (often reflecting out-group animosity) were much more likely to be shared than those about the political in-group. Each additional outgroup 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 (Rathje, Van Bavel, & Van der Linden, 2021). Relatedly, content expressing moral outrage is more likely to be shared on Twitter, especially within – and not between – partisan echo chambers (Brady et al., 2017). Additionally, positive social feedback (e.g., likes and shares) on posts expressing outrage increases the likelihood that people will express outrage in the future (Brady et. al,

2021). Additionally, the most popular content on Facebook tends to consist of right-wing, hyper-partisan 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 et al., 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 et al., 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, Van Bavel, & Van Der Linden, 2021).

One potential reason divisive content succeeds online may be because it is particularly likely to capture our attention (Brady et al., 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, Van Bavel, & Van Der Linden, 2021).

While divisive posts 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. Indeed, 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. Randomized control trials find that social media usage decreases well-being (Allcott et al., 2020; Asimovic et al., 2021); 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 news feeds after discovering that, while it led to increased engagement, survey data revealed that people did not enjoy it (Gupta, 2021).

Real World Behavior

Behavior on social media can have far-reaching offline consequences. Due to social media's ability to communicate critical protest information rapidly and broadly, social media has been associated with increases in democratic action and protest behavior across the world (Jost et al. 2018; Gonzalez-Bailon et al., 2011). 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 Black Lives Matter movement (Cox, 2017), and the subsequent BLM protests in 2020 may have been the largest protests in U.S. history (Bolsover, 2020; Buchanan et al. 2020).

Nonetheless, nefarious movements have also originated online, giving voice to conspiracy theories and hate groups (Douglas et al., 2019). Recently, the U.S. conspiracy theory group QAnon has gained 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 United States Capitol committed by people who believed the false claim propagated by QAnon that the U.S. Presidential Election had been fraudulent (Luke, 2021). Twitter use has also been linked to an increase in anti-Muslim hate crimes at the community level (Müller & Schwarz, 2020). Furthermore, the same authors found that 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. Due to the incredible number of social changes that occurred in parallel with the development and adoption of social media and the internet broadly, much 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. However, research using quasi-experiments, qualitative data, and archival research are useful to understand these important phenomena.

Real World Social Context

Additional challenges of examining the causal influence of social media on offline behavior are individual differences between users and the offline social contexts they occupy. Recent polls estimate that approximately 15% of Americants (~30 million people) believe in the core tenets of QAnon (Russonello, 2021), but only several thousand people were present during the capitol insurrection on January 6th (Doig, 2021). The insurrection highlights how individual differences are important to examine, because not every person who is exposed to or believes in this online content participates in related offline behavior (Arceneaux et al. 2018). 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 (Bor & Petersen, 2021).

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). Researchers found that the effect of social media on outgroup ethnic regard depended on the homogeneity of participants offline social networks. After a week without social media, participants reported lower ethnic outgroup regard than those who had not deactivated, but only if their offline social network was homogenous. This finding suggests that offline behavior may improve due to social media when the political and ethnic makeup of one's online social networks are more diverse compared to their offline social networks.

At the society level, political context interacts with the influence of internet and social media use on offline behavior (Zhuravskaya et al., 2020). For example, research has suggested that greater internet use is associated with lower government approval, but only where government corruption was already present (Guriev et al., 2020). This dip in government approval may in part be due to the increasing ease of online dissemination of information about genuinely corrupt actions. Furthermore, digitally native social movements like Black Lives Matter

had been active for several years before the unjust death of George Floyd sparked massive protests (Bolsover, 2020). Hence, outcomes related to the internet and social media may rely on specific features/events in society to catalyse.

Discussion

Our paper summarizes a growing vanguard of work on the impact of social media on political life. Unfortunately, more work needs to be done to fully understand the impact of social media. Given the rapid growth and global scale of this technology, we need to urgently support more work on the topic. In addition to polarization--and concomitant threats to democracy--we also have urgent concerns about the role of social media in the spread of misinformation. Given these widespread negative consequences--and lack of access to internal social media platform data--a recent paper has argued that 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).

We introduced the *SPIR* framework of social media and polarization, which illustrates the ways in which *Sorting*, *Platform-design*, *Incentive structures*, and *Real-world context* may contribute to social media's effect on polarization. Potential solutions can also target each aspect of this model to reduce online polarization. For instance, design-based interventions can be implemented within the social media platform design to discourage the sharing of false or polarizing content. Social media algorithms can be shifted to provide incentives for more constructive and reliable content rather than false or divisive content. Additionally, the real-world context in which social media operates can be considered in implementing potential solutions, since the effects of social media on polarization and inter-group conflict vary depending on social contextual factors and individual differences.

References

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629–676. https://doi.org/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), 20200147. https://doi.org/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*, *118*(25). https://doi.org/10.1073/pnas.2022819118

Bail, C. (2021). *Breaking the Social Media Prism*. Princeton University Press.

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

Bak-Coleman, J. B., Alfano, M., Barfuss, W., Bergstrom, C. T., Centeno, M. A., Couzin, I. D., Donges, J. F., Galesic, M., Gersick, A. S., Jacquet, J., Kao, A. B., Moran, R. E., Romanczuk, P., Rubenstein, D. I., Tombak, K. J., Van Bavel, J. J., & Weber, E. U. (2021). Stewardship of global collective behavior. *Proceedings of the National Academy of Sciences*, *118*(27). https://doi.org/10.1073/pnas.2025764118

Bolsover, G. (2020). *Black Lives Matter discourse on US social media during COVID: Polarised positions enacted in a new event*. arXiv. https://arxiv.org/search/?query=Black+Lives+Matter+Discourse+on+US+Social+Media+during+COVID%3A+Polarised+Positions+Enacted+in+a+New+Event&searchtype=all&source=header

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*, 1-18. https://doi.org/10.1017/S0003055421000885

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2020). *Cross-country trends in affective polarization*. National Bureau of Economic Research. http://www.nber.org/papers/w26669

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

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. https://psycnet.apa.org/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), Article eabe5641. https://doi.org/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*, *114*(28), 7313–7318. https://doi.org/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*. https://www.nytimes.com/interactive/2020/07/03/us/george-floyd-protests-crowd-size.html

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. https://doi.org/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*, *118*(9), Article e2023301118. https://doi.org/10.1073/pnas.2023301118

Clement, J. (2020). Number of social network users worldwide from 2017 to 2025. Retrieved June, 4, 2020.

Costa, M. (2020). Ideology, not affect: What Americans want from political representation. *American Journal of Political Science*, *65*(2), 342-358. https://doi.org/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. https://doi.org/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. 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. https://doi.org/10.1111/pops.12568

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

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, *370*(6516), 533–536. https://doi.org/10.1126/science.abe1715

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

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

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

Guriev, S., Melnikov, N., & Zhuravskaya, E. (2020). 3G internet and confidence in government. *Quarterly Journal of Economics*. https://doi.org/10.1093/qje/qjaa040

Hale, J. (2019, May 7). *More Than 500 Hours Of Content Are Now Being Uploaded To YouTube Every Minute*. Tubefilter. https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute/

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. https://doi.org/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*, *118*(32), Article e2101967118. https://doi.org/10.1073/pnas.2101967118

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

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

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. https://doi.org/10.1177/0093650209333030

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

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

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. https://doi.org/10.1080/00131857.2021.1889327

Mooijman, M., Hoover, J., Lin, Y., Ji, H., & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour*, *2*(6), 389–396. 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), Article e2022761118. https://doi.org/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. https://doi.org/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. https://doi.org/10.1177/2056305117729314

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

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. https://doi.org/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. *Proceedings of ACM*

Conference.
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

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*. PsyArXiv. https://doi.org/10.31234/osf.io/7vc5d

Rathje, S. (2021). Why Facebook really, really doesn't want to discourage extremism. *The Washington Post*. https://www.washingtonpost.com/politics/2021/07/13/why-facebook-really-really-doesnt-want-discourage-extremism/

Rathje, S, Roozebeek, J, Traberg, Van Bavel, J.J., & van der Linden, S. (2021). Matters arising: Partisan differences in the effect of priming accuracy on misinformation sharing. *Nature* (under review).

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, *118*(26), Article e2024292118. https://doi.org/10.1073/pnas.2024292118

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. https://doi.org/10.1037/tps0000121

Roose, K. (2021, July 14). Inside Facebook's data wars. *The New York Times*. https://www.nytimes.com/2021/07/14/technology/facebook-data.html

Roose, K., Isaac, M., & Frenkel, S. (2020, November 24). Facebook struggles to balance civility and growth. *The New York Times*. https://www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html

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

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. https://doi.org/10.1111/jcom.12284

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

Statista. (2020, July). *Number of social network users worldwide from 2017 to 2025 (in billions)* [Chart]. Statista. 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. https://doi.org/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*, 1431-1444. https://dx.doi.org/10.2139/ssrn.2821919

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

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. https://doi.org/10.1111/j.1460-2466.2012.01629.x

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

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

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

Yan, H. Y., Yang, K.-C., Menczer, F., & Shanahan, J. (2020). Asymmetrical perceptions of partisan political bots. *New Media & Society*, Article 1461444820942744. https://doi.org/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. https://doi.org/10.1080/10584609.2020.1785067

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), Article 205630511986546. https://doi.org/10.1177/2056305119865465

Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, *12*(1), 415–438. https://doi.org/10.1146/annurev-economics-081919-050239