Science & Society

How social media shapes polarization

Jay J. Van Bavel [1,2,*]
Steve Rathje [3,*]
Elizabeth Harris [1,*]
Claire Robertson [1,*] and
Anni Sternisko [1,*]

This article reviews the empirical evidence on the relationship between social media and political polarization. We argue that social media shapes polarization through the following social, cognitive, and technological processes: partisan selection, message content, and platform design and algorithms.

With nearly 4 billion people now on social media, understanding the impact it has on individual cognition and society at large is increasingly important. There is a growing body of work examining the transmission of polarizing content and divisive information on online platforms. However, the impact of social media on political polarization (i.e., the division into two sharply contrasting groups or sets of beliefs) remains a contentious issue. According to one perspective, social media platforms increase polarization by amplifying divisive partisan content, moving people into echo chambers, and incentivizing intergroup conflict [1]. Other scholars argue that social media does not play a major role in polarization or sort us into echo chambers [2][i].

We aim to reconcile this debate by reviewing the empirical evidence in multiple literatures and highlighting controversies and gaps in our understanding of these issues. We focus on both observational and experimental evidence of online behavior. Although social media is unlikely to be the main driver of polarization, we posit that it is often a key facilitator. Here, we highlight processes by which social media might shape polarization (i.e., partisan selection, message content, platform design, and algorithms). Our article provides a framework for understanding the impact of social media on polarization.

Controversies and misconceptions

Some of the most widely discussed aspects of social media's role in polarization may be the least influential. A popular notion is that social media increases political polarization by creating online political echo chambers (i.e., exposing individuals to increasingly partisan and polarizing content over time due to selective exposure and fringe content sites). The number of Americans who consume highly polarized content online is dwarfed by the number who consume more centrist media, or the number of Americans who simply do not care about politics [3], although these data should be treated with caution since partisans may selectively share or consume content that aligns with their identity or ideology.

Another source of debate is whether social media exposure causes polarization at all. Facebook, for instance, has argued that the narrative that social media is contributing to polarization is 'not supported by research'[ii]. Additionally, one prominent paper found that the most polarized segment of American society is older adults and yet this group is the least likely to use social media [2]. The authors concluded that social media is therefore unlikely to be the main cause of polarization. However, this paper relies on observational data and did not analyze the causal effect of social media exposure on people's political beliefs (Box 1).

Experimental evidence of polarization

A pair of recent experiments found that social media platforms appear to cause greater polarization. One large field study randomly assigned Americans to deactivate their Facebook account (or not) for 4 weeks prior to the 2018 US election. Deactivating Facebook reduced issue (i.e., policy preferences) polarization and marginally reduced affective (i.e., feelings about the other party) polarization, largely by decreasing people's exposure to news that facilitated a better understanding of their own party relative to the other party [4]. With the caveat that its sample was not nationally representative (e.g., it tended to be younger, more educated, and willing to delete their Facebook accounts), the study provides causal evidence that Facebook promotes polarization.

This effect might not be generalizable to all intergroup contexts, especially when people have homogeneous offline social networks. A randomized control trial in Bosnia and Herzegovina found that deactivating Facebook during Genocide Remembrance Week lowered ethnic outgroup regard compared with those who did not deactivate, but only among individuals who had a homogeneous offline social network [5]. The political and ethnic makeup of some online social networks may be more diverse than offline social networks, and this can shape the impact of social media on polarization.

Another field experiment paid Democrats and Republicans to follow Twitter bots that retweeted messages by elected officials and opinion leaders with opposing political views for 1 month [6]. Exposure to members of the other party increased political polarization (although this backfire effect was significant only among Republicans). This highlights a possible mechanism for how social media may increase polarization. As social media tends to amplify extreme viewpoints [7,8], exposure to hyper-partisans from the outgroup may lead people to become even more entrenched in their own viewpoint. Next, we turn to the potential processes

---

Box 1. Evidence against polarization suffers from the ecological fallacy

Boxell and colleagues [2] found that greater Internet use is not associated with faster growth in political polarization among US demographic groups. They found that polarization had increased the most among the demographic groups least likely to use the Internet and social media (i.e., elderly people). Between 1996 and 2016, polarization grew by 0.23, 0.23, and 0.47 index points, respectively, among those aged 18–39, 40–64, and 65+ years. However, the elderly (65+ years) have substantially lower levels of Internet use. The study did not measure individual social media usage.

A limitation with the study is that it was correlational and appears to suffer from the 'ecological fallacy' – a logical fallacy in the interpretation of statistical data that occurs when inferences about the nature of individuals are deduced from inferences about the group to which those individuals belong. It is impossible to know from these data whether individuals were unaffected by social media exposure. It is possible that social media experience increased polarization in every age group. Subsequent studies have used random assignment to determine whether social media shapes polarization. These studies have found evidence that there is a causal link between social media exposure and polarization at the individual level [3,4].

---

by which social media may contribute to polarization.

Partisan selection

While the internet and social media offer an overwhelming amount of information, people tend to seek out information that affirms their beliefs or identities. Thus, they often prefer news that is congruent, rather than incongruent, with their political views [9]. Research suggests that greater selective exposure to congruent political news is correlated with (possibly being both caused by and causing) polarization [4].

Not only do people seek out politically congruent information, they also update their beliefs more when that information supports what they already believe (asymmetric updating [10]). Finally, they are more willing to share politically congruent information online [11]. These cognitive biases in information seeking, belief updating, and sharing may all increase polarization. We posit that these effects are user driven, although these cognitive biases may also interact with platform features to amplify the effect (e.g., seeking out and engaging with congruent information results in the platform's algorithm exposing the individual to more similar content).

Message content

Divisive social media messages tend to receive more engagement, which might contribute to polarization. An analysis of nearly 3 million social media posts found that posts about the political outgroup (often reflecting outgroup animosity) were more likely to be shared than those about the political ingroup. Each additional outgroup word (e.g., 'liberal' or 'Democrat' for a Republican post) increased the odds of that post being shared by 67% and increased the volume of 'angry' reactions on Facebook (Figure 1) [8]. The rhetoric about outgroup members was usually negative. Content expressing moral outrage is also more likely to be shared on Twitter, especially within – and not between – partisan social networks [12].

Another reason why divisive content might be more prominent online is because it captures our attention [1]. Social media operates as an attention economy, whereby individuals, organizations, and politicians try to go 'viral'. Sharing socially divisive content may be an effective strategy for message diffusion. Indeed, the most politically extreme American politicians have the most followers [13]. Thus, social media platforms may be incentivizing moral outrage and divisive content, especially towards outgroups.

Platform design and algorithms

The relationship between social media and polarization is also likely to depend on the features of the different social media platforms (e.g., populations, social norms, social-feedback dynamics). Such differences may explain why there seems to be greater polarization on certain platforms. For instance, researchers in Israel found more polarization on Twitter than Facebook and WhatsApp [14]. Similarly, different platforms may facilitate different types of polarization. Facebook algorithms may increase affective polarization while Twitter has been linked to both affective and attitudinal polarization [14,15]. Social media algorithms may further compound these cognitive processes. Some platforms' algorithms seem to amplify content that affirms one's social identity and pre-existing beliefs. For instance, Facebook's newsfeed seems to increasingly align content with cues about users' political ideology (e.g., following partisan news sources increases exposure to similar content) [15].

Concluding remarks

The massive growth of social media underscores the need for more work on the impact these platforms are having on human cognition and behavior. Here, we highlight how social media may contribute to polarization (although social media use and polarization may be mutually reinforcing). There are numerous limitations in this literature, including the reliance on observational studies and the small number of social media platforms, as well as relatively little research involving countries outside the USA. Little is known about the underlying recommendation algorithms due to lack of transparency on the part of social media companies. Given the significant costs to society of increasing misinformation, polarization, and other conflict, it is critical that scholars gain access to data from these platforms. Without access, it is difficult to know which platforms, design features, and algorithms are likely to foment social conflict. More work also needs to examine the potential long-term costs and benefits of social media for political discourse.
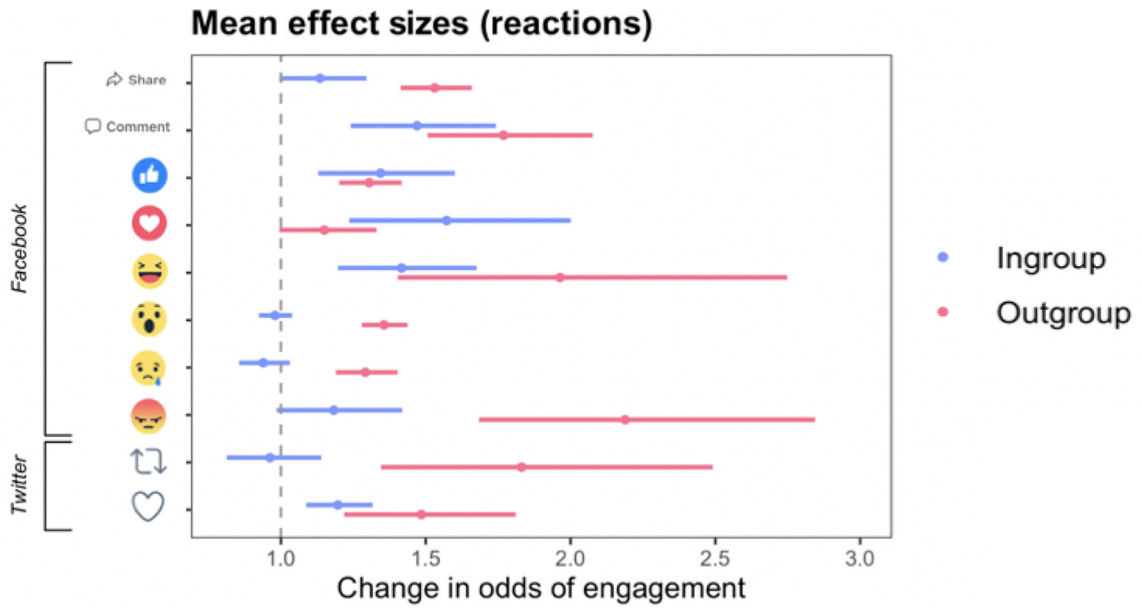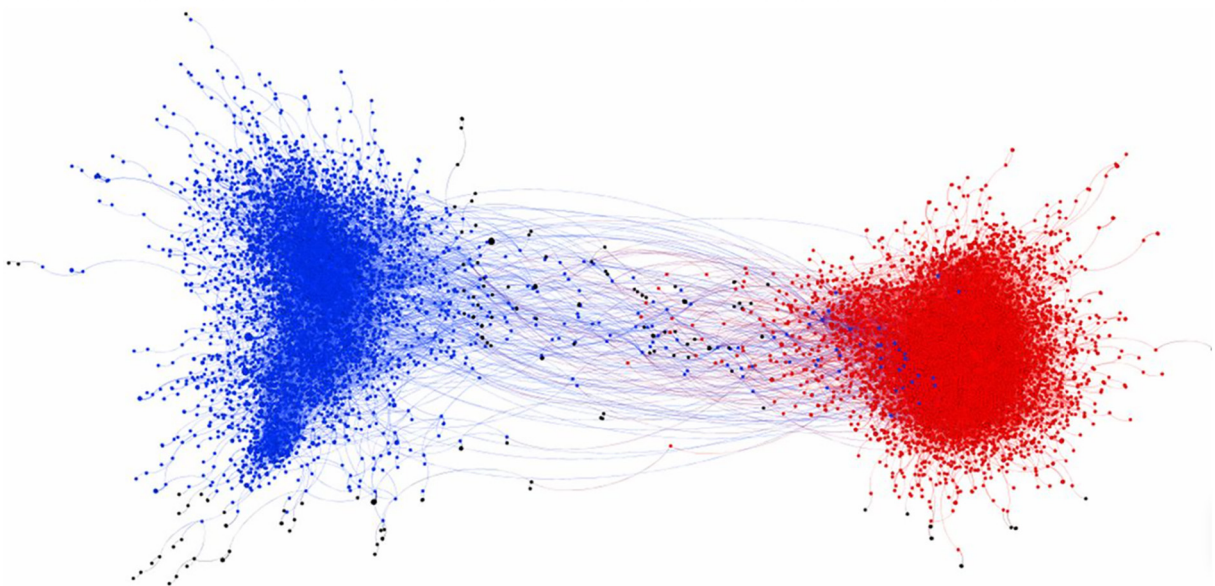
Figure 1. Partisan rhetoric on social media. (A) Shows the effects of ingroup and outgroup language on retweets, shares, and 'reactions' on Facebook (from [8]). (B) Shows the retweet networks of liberals (in blue) and conservatives (in red) when they include moral emotional language (from [12]).

Declaration of interests

No interests are declared.

[1]Department of Psychology, New York University, New York, NY, USA
[2]Center for Neural Science, New York University, New York, NY, USA
[3]Department of Psychology, University of Cambridge, Cambridge CB2 3RQ, UK

*Correspondence:
jay.vanbavel@nyu.edu (J.J. Van Bavel),
sjr@cam.ac.uk (S. Rathje), eah561@nyu.edu (E. Harris),
cer493@nyu.edu (C. Robertson), and
sternisko@nyu.edu (A. Sternisko).

https://doi.org/10.1016/j.tics.2021.07.013

References

1. Brady, W.J. *et al.* (2019) The MAD model of moral contagion: the role of motivation, attention and design in the spread of moralized content online. *Perspect. Psychol. Sci.* 15, 978–1010
2. Boxell, L. *et al.* (2017) Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proc. Natl. Acad. Sci. U. S. A.* 114, 10612–10617
3. Guess, A. *et al.* (2018) *Avoiding the echo chamber about echo chambers*, Knight Foundation
4. Allcott, H. *et al.* (2020) The welfare effects of social media. *Am. Econ. Rev.* 110, 629–676
5. Asimovic, N. *et al.* (2021) Testing the effects of Facebook usage in an ethnically polarized setting. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2022819118
6. Bail, C.A. *et al.* (2018) Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci. U. S. A.* 115, 9216–9221
7. Bail, C. (2021) *Breaking the social media prism*, Princeton University Press
8. Rathje, S. *et al.* (2021) Out-group animosity drives engagement on social media. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2024292118
9. Knobloch-Westerwick, S. and Meng, J. (2009) Looking the other way: selective exposure to attitude-consistent and counterattitudinal political information. *Commun. Res.* 36, 426–448
10. Sunstein, C.R. *et al.* (2016) How people update beliefs about climate change: good news and bad news. *Cornell Law Rev.* 102, 1431
11. Shin, J. and Thorson, K. (2017) Partisan selective sharing: the biased diffusion of fact-checking messages on social media. *J. Commun.* 67, 233–255
12. Brady, W.J. *et al.* (2017) Emotion shapes the diffusion of moralized content in social networks. *Proc. Natl. Acad. Sci. U. S. A.* 114, 7313–7318
13. Hong, S. and Kim, S.H. (2016) Political polarization on twitter: implications for the use of social media in digital governments. *Gov. Inf. Q.* 33, 777–782
14. Yarchi, M. *et al.* (2021) Political polarization on the digital sphere: a cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Polit. Commun.* 38, 98–139
15. Levy, R. (2021) Social media, news consumption, and polarization: evidence from a field experiment. *Am. Econ. Rev.* 111, 831–870