Neural polarization and routes to depolarization

Samantha L. Moore-Berg[a,b] , Jacob M. Parelman[a], Yphtach Lelkes[a] , and Emily B. Falk[a,c,d,1]

Political polarization has intensified in the lead-up to the 2020 US presidential election, with liberal and conservative politicians hurling insults at one another, journalists highlighting ways in which Americans are deeply divided, and parts of the general American public condoning violence if their side does not win the upcoming election. Likewise, in countries around the world, political extremists are gaining political power, the media has zeroed in on division, and individuals perceive deepening political divides. Within this context, Leong et al. (1) report evidence in PNAS for "neural polarization," or divergent brain activity between self-identified liberals and conservatives, while watching media clips about a salient political issue (immigration). The team focuses on the DMPFC (dorsomedial prefrontal cortex), a brain region that helps people navigate the social world, interpret narratives, and understand others' mental states and attributions (2). Those with stronger partisan identities had more similar DMPFC activity in response to media coverage of immigration, relative to other members of their own group, and were subsequently more likely to change their attitudes to even more partisan positions. Thus, these findings highlight ways in which strongly identifying with a particular group (such as a political party or position) might color the ways we interpret new information, and the ways we process incoming information may reinforce our existing identities.

Is this polarization inevitable? Or is it possible to find common ground to move forward? Inspired by Leong et al.'s (1) findings, we argue that polarization is not inevitable and suggest several points where intervention might be fruitful.

Leong et al. (1) start with the idea that people interpret information based on important parts of their identities, with political affiliations as a particularly strong example (e.g., ref. 3). In contrast with work that focuses on the possibility that the brains of liberals and the brains of conservatives process information differently, Leong et al. (1) demonstrate that partisans from

both sides of the political spectrum exhibited highly similar neural responses in a variety of brain regions involved in information processing. It was only in a single higher-order brain region (DMPFC) that is involved in social thought and narrative interpretation that the authors observed neural polarization, regardless of party. One possible explanation for this finding is that people respond to ideas and words they are familiar with. Neural polarization may be driven by conservatives recognizing words and ideas they have heard on Fox News, for example, or liberals recognizing words and ideas they have heard on MSNBC, and each responding to ideas that political elites call most attention to. Then, when exposed to new ideas and content, even if it is identical [as was the case in Leong et al.'s (1) paradigm], different points may draw attention and be interpreted differently.

Understanding the way that information is filtered during such exposure is key to understanding the ways that partisan divides may be reinforced or mitigated. In particular, when learning new information, partisans often seek out information that is consistent with their existing political biases (4) and interpret the world in ways that are consistent with their worldviews. In turn, people whom we believe are similar to us hold particularly strong influence on our beliefs and preferences (5). Further, people tend to adopt the ideological positions of others they identify with (e.g., ref. 6).

Within this self-reinforcing cycle, how we process incoming information not only affects our own views but it also affects how we influence and communicate with others. For instance, brain activation within the DMPFC and other regions used to understand other people (i.e., mentalize) tracks with decisions about what news articles people share online (e.g., ref. 7), and synchrony in mentalizing regions of interlocutors is associated with successful communication (e.g., ref. 8). This happens even more so when people have similar starting assumptions (9). Likewise, people who are closer together in social networks use their brains
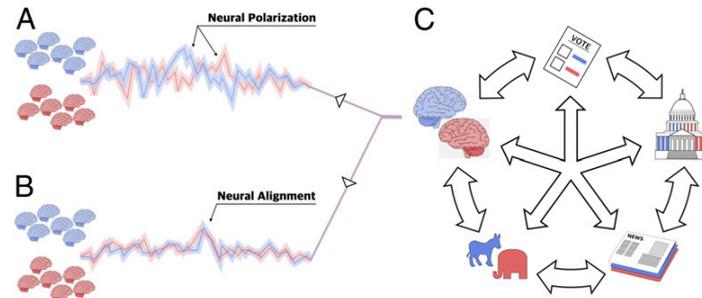
Fig. 1. Leong et al. (1) find that (A) neural patterns diverged between liberals and conservatives at key moments of emotional language in media narratives about immigration. Other work suggests that, although common, such polarization is not inevitable, and targeted interventions may be able to encourage (B) greater alignment between partisans. (C) Such interventions may target the many broader institutions and factors (e.g., elected officials, news media, partisan group identities, individual psychology, and partisan behavior) that reciprocally influence one another.

similarly when watching media clips (e.g., ref. 10) and people whose brains respond more similarly are more likely to become friends (10). The current findings (1) suggest that similar relationships may exist in the case of political media. The findings also highlight ways that starting differences in worldview may become even more entrenched as we not only remain close to people only within our own echo chambers, but we also seek out attitude-consistent media and interpret the same media through partisan lenses.

How then can we make it more likely that people break out of their own echo chambers to learn about the other party's point of view? What makes it harder or easier to get on the same page? First, a long history of research in communication shows that the media play a strong role in determining what issues people pay attention to (11) and the way issues are covered matters (12). Likewise, Leong et al. (1) found that heightened neural polarization in the DMPFC was particularly strong in response to the use of moral–emotional or risk-related words in the political content that participants viewed. Neural polarization increased when such language was present, suggesting that emotional content (i.e., risk and moral language) is most likely to result in biased narrative interpretation via the DMPFC. Unfortunately, the current design of many online social media platforms amplifies and increases our exposure to expressions of exactly this type of language [e.g., moral outrage (13)].

Despite a media environment that highlights and amplifies polarization, not all participants' brains showed the same level of neural polarization. People whose brain responses are more similar to outgroup responses may be more open to changing their minds about social issues such as immigration policy (1). This offers a possible clue into ways that we might hope to steer our collective raft toward core, shared ideals of open dialogue and thoughtful compromise. Given our understanding of empathy, and the malleability of neural and behavioral responses, polarization need not be inevitable.

For example, the key brain region whose activity reflected neural polarization (DMPFC) is also a region that is activated by empathy inductions (e.g., ref. 14), and narrative interventions can increase empathy toward outgroup members (15). Yet, empathizing

with others' mental states might be challenging, as people view the world through the lens of their ingroup and have brain processing that responds accordingly. Although often reserved for those closest to us (14), empathy can be actively cultivated and expanded to include a wider range of targets.

However, one barrier to extending empathy to other groups could be inaccurate perceptions about the other side. For instance, research suggests that we hold biased views about what others think about us (16, 17). That is, people tend to think that the other side views them much more negatively than in reality, which in turn predicts support for policies that threaten democratic norms (e.g., ref. 16). However, research in this area also offers hope: Correcting inaccurate beliefs about what other groups think of us improves intergroup interactions (17). Likewise, positive intergroup contact can also help mitigate some of these harmful perceptions about the other group's view of us (18). Finally, priming people to think of their superordinate identities (e.g., Americans) rather than their narrow identities (e.g., partisans) decreases polarization (19).

If we want to effectively reduce political polarization, we need to recognize the biases that our brains impose in processing and the ways that broader institutions (e.g., media and political systems) may shape our thoughts and feelings (Fig. 1). Journalists and media makers can more mindfully use language that mitigates mass polarization perceptions, and media and educational institutions can teach individuals about how context can shape their interpretation and communication of social issues. Only once we realize that we are all subject to many layers of influence that our brains seamlessly convince us constitute "reality" will we then be able to successfully reduce political polarization.

1 Y. C. Leong, J. Chen, R. Willer, J. Zaki, Conservative and liberal attitudes drive polarized neural responses to political content. *Proc. Natl. Acad. Sci. U.S.A.*, 10.1073/pnas.2008530117 (2020).

2 U. Frith, C. D. Frith, Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 459–473 (2003).

3 S. Iyengar, M. Krupenkin, The strengthening of partisan affect. *Polit. Psychol.* **39**, 201–218 (2018).

4 C. S. Taber, M. Lodge, Motivated skepticism in the evaluation of political beliefs. *Am. J. Pol. Sci.* **50**, 755–769 (2006).

5 R. B. Cialdini, N. J. Goldstein, Social influence: Compliance and conformity. *Annu. Rev. Psychol.* **55**, 591–621 (2004).

6 B. N. Bakker, Y. Lelkes, A. Malka, Understanding partisan cue receptivity: Tests of predictions from the bounded rationality and expressive utility perspectives. *J. Polit.* **82**, 1530–1544 (2020).

7 C. Scholz et al., A neural model of valuation and information virality. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2881–2886 (2017).

8 G. J. Stephens, L. J. Silbert, U. Hasson, Speaker-listener neural coupling underlies successful communication. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14425–14430 (2010).

9 Y. Yeshurun et al., Same story, different story: The neural representation of interpretive frameworks. *Psychol. Sci.* **28**, 307–319 (2017).

10 C. Parkinson, A. M. Kleinbaum, T. Wheatley, Similar neural responses predict friendship. *Nat. Commun.* **9**, 332 (2018).

11 M. McCombs, D. Shaw, The agenda-setting function of mass media. *Public Opin. Q.* **36**, 176–187 (1972).

12 D. Chong, J. N. Druckman, Framing public opinion in competitive democracies. *Am. Polit. Sci. Rev.* **101**, 637–655 (2007).

13 M. J. Crockett, Moral outrage in the digital age. *Nat. Hum. Behav.* **1**, 769–771 (2017).

14 J. Zaki, Empathy: A motivated account. *Psychol. Bull.* **140**, 1608–1647 (2014).

15 E. G. Bruneau, M. Cikara, R. Saxe, Minding the gap: Narrative descriptions about mental states attenuate parochial empathy. *PLoS One* **10**, e0140838 (2015).

16 S. L. Moore-Berg, L. O. Ankori-Karlinsky, B. Hameiri, E. Bruneau, Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 14864–14872 (2020).

17 J. Lees, M. Cikara, Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nat. Hum. Behav.* **4**, 279–286 (2020).

18 E. Bruneau, B. Hameiri, S. L. Moore-Berg, N. Kteily, Intergroup contact reduces dehumanization and meta-dehumanization: Cross-sectional, longitudinal, and quasi-experimental evidence from 16 samples in five countries. *Pers. Soc. Psychol. Bull.*, 10.1177/0146167220949004 (2020).

19 M. S. Levendusky, Americans, not partisans: Can priming American national identity reduce affective polarization? *J. Polit.* **80**, 59–70 (2018).