

Modeling the power of polarization

With assists from politicians and social media, people are increasingly dividing themselves into social and political factions. Models can hint at how it happens—and maybe offer ways to mitigate it.

M. Mitchell Waldrop, *Science Writer*

In 2016, when Vicky Chuqiao Yang first started to work on computer simulations of US politics, she was fascinated by the realization that the left–right standoff widely described as “polarization” is not one thing. “There are two kinds of polarization that the media and the public often get confused,” says Yang, an applied mathematician at the Santa Fe Institute in New Mexico. One type is issue polarization: “how much people disagree on policies like ‘What should be the tax rates?’ or ‘What should be the laws to regulate guns?’.”

Those divisions have been widening of late. But they aren’t nearly as incendiary as social or “affective” polarization, which is about anger, distrust, resentment,

tribal identity, and mutual loathing (see Fig. 1). As a team of prominent social scientists warned last year, social polarization in conjunction with legislative gridlock and hyper-partisan media have created an “American sectarianism” that threatens democracy itself (1).

Researchers are trying to understand why social polarization is on the rise and—perhaps more importantly—what we can do about it. Can we find solutions by focusing on racial anxieties, conspiracy theories, and social media echo chambers that endlessly reinforce a single viewpoint? Or do we also need to look for more fundamental forces at work?



Americans have always been polarized in some ways. But using models, a cadre of researchers is trying to understand why social polarization is on the rise and—perhaps more importantly—what we can do about it. Image credit: Dave Cutler (artist).

Published under the [PNAS license](#).

Published September 8, 2021.

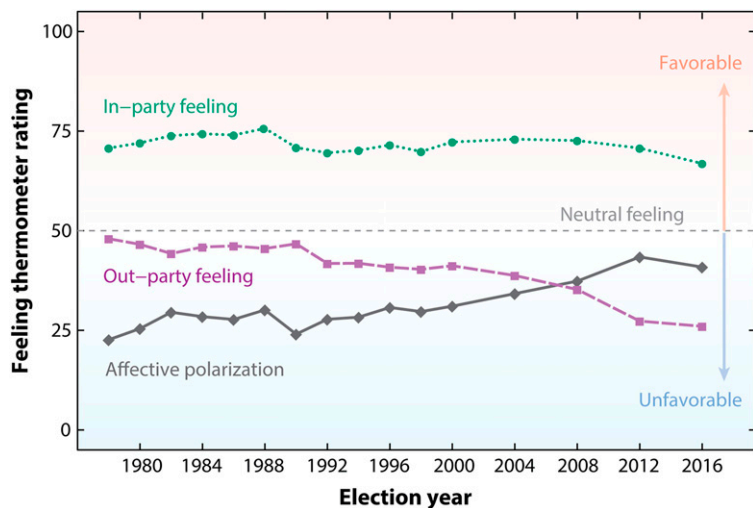


Fig. 1. Recent years have seen a marked rise in “affective” polarization, a feeling of mutual dislike and mistrust between the two sides. The trend is illustrated in data from the American National Election Survey: People’s feeling of warmth toward members of their own party (green) has held steady since 1980, whereas their feelings toward members of the other party (purple) have dropped. The difference (black) is a measure of affective polarization. Reprinted with permission from ref. 22.

These are the kinds of questions that have brought Yang and a host of other modelers into the long-established field of opinion dynamics: the study of how people’s viewpoints form and change as they interact. From a research perspective, the timing couldn’t be better, says Antonio Sirianni, a postdoc in computational social science at Dartmouth University in Hanover, NH. Thanks to the ever-increasing pace of technological advances, researchers now have the computational power to run complex simulations and models, as well as access to an unprecedented amount of real-world data on political opinion.

Their results to date are intriguing, if incomplete. Researchers have begun to map out some of the ways that geography, psychology, and group dynamics shape polarization. They’re starting to understand why some of the most intense social polarization seems to be driven by relatively small cadres of highly politicized individuals who mainly talk to each other. And researchers found hints that the often-discouraging effort to improve communication across the divide might in fact be the best way forward. They’ve shown that some of the most frequently proposed technical fixes—for example, trying to break up social media echo chambers by tinkering with the algorithms on Facebook, Twitter, and the like—could easily backfire and produce more antagonism, not less.

“We always try to be very careful with drawing conclusions about how to intervene because these models are based on assumptions,” says Michael Mäs, a sociologist at the Karlsruhe Institute of Technology in Germany. Society as a whole, and social media platforms in particular, are highly complex systems in which the outcomes are very sensitive to initial conditions, he says. “So if these assumptions happen to be false, even in small ways, the predictions can change dramatically.”

Root Causes

Modeling is one of at least four basic strategies for studying any form of polarization, says Michael Macy, a sociologist at Cornell University in Ithaca, NY, who uses all four. The classic method is observation: using surveys and historical data to track how polarization has increased or decreased over time and which issues have been the most divisive. “There are some great studies using surveys going back to the late 1990s, when people first started worrying about polarization,” says Macy.

A second, newer strategy is to analyze the tsunami of data now available from the Internet. “It’s sort of like the survey research,” says Macy, “except that it’s actual behavior observed in online communities and social media. And that has proven to be very useful for studying things that you cannot learn from a survey,” like exactly who listens to whom and how ideas spread through the resulting social network like a contagion.

Then there is the experimental approach: watching how polarization develops among volunteers in a laboratory setting. These experiments allow you to control the conditions, separate the signal from the noise, and tease out what’s cause and what’s effect, he says—all of which is hard to do with survey data.

And finally, says Macy, models in the form of mathematical equations or computer simulations can help researchers explore the sometimes surprising outcomes of simple starting conditions or assumptions.

In the polarization studies that have been done to date, says Macy, one of the most striking insights is how much of it can be explained by the interplay of just two sociological forces. One of them is the assimilation, or “influence” effect: People who interact a lot will eventually start to think and act alike.

This effect is so strong, and so well documented in the literature, that social scientists spent decades trying to figure out why polarization exists at all—or why, for that matter, humans are divided by language, fashion, cuisine, music, folkways, and a host of other differences. Why do these divisions often endure for centuries, instead of gradually fading away as the assimilation effect seemed to predict? As psychologist Robert Abelson famously lamented in 1964, after his every attempt to model what he called “community cleavage” ended in yet more consensus, “we are naturally led to inquire what on earth one must assume in order to generate [the observed cleavages]” (2).

A big part of the answer turned out to be the second force, homophily: people’s preference for hanging out with others like themselves. One influential study of the power of homophily was Robert Axelrod’s 1997 model of culture formation (3). This model turned out to anticipate today’s rural–urban split between Republicans and Democrats, as well as the self-reinforcing echo chambers that have now become familiar on Twitter and Facebook. But at the time, says Axelrod, a political scientist at the University of Michigan in Ann Arbor, “I wasn’t interested in the left-to-right kind of differences, so I treated ‘culture’ simply as a list of arbitrary features that were observable, like ‘What kind of hat do you wear?’ or ‘What ethnicity are you?’” Next, Axelrod

modeled people as independent snippets of code, or “agents,” that could move around a simulated landscape (4), and gave each agent some initial set of cultural features. Then he set them to interacting with their neighbors according to two simple rules. First, the more items of culture agents share, the more likely they are to interact. And second, if agents do interact, they adopt some feature of the agent they’re interacting with.

In sum, says Axelrod, the model was nothing but assimilation plus homophily: “Like gravity, it’s all pulling together, right? There’s nothing but attraction.” Yet the result wasn’t anything like global consensus. Instead, says Axelrod, the model consistently locked itself into a patchwork resembling the multiple language regions of Europe—or those filter bubbles on Facebook. “It’s what I call local convergence and global polarization,” he says: Cultures do indeed tend toward consensus within a finite region. But at the boundaries, the differences eventually become so stark that the agents on either side quit interacting at all. “So they never talk to each other again,” says Axelrod, “and that’s why it freezes.”

The Emotional Connection

Recent modeling work has also yielded a second key insight about polarization: namely, the crucial role played by negative emotions, which can turn both influence and homophily inside out. Just as people can be drawn together by the influence effect, says Macy, “they can also become more different from each other through negative influence,” also known as “repulsion.” And the flip side of homophily is xenophobia, he says, “which is the tendency to run away from those who are different.”

Negative emotion is obviously crucial for understanding the intergroup venom we’re seeing today. But Noah Friedkin, a sociologist at the University of California, Santa Barbara, points out that efforts to model its effects actually date back to the birth of “balance theory” in the 1940s and 1950s (5, 6).

Balance theory describes how people’s opinions and feelings reinforce one another through a feedback loop, explains Friedkin. If you like the people you interact with, then you will start to think and act like them. But if you don’t, the dynamic gets flipped: You’ll increasingly tend to avoid the people you dislike and to reject their views. Meanwhile, you’re constantly adjusting how you feel about each person based on the views they espouse: You start to like them more if their beliefs jibe with yours, and vice versa.

The result is a complicated evolution of feelings and opinions that continues until this theoretical society achieves a stable equilibrium, or balance. In the theory’s original, simplest form, says Friedkin, only two such equilibria are possible, researchers found. The first is consensus, namely “one big happy clique where everyone’s friends with everyone else,” he says. In the second, “the group splits into two cliques that are at each other’s throats,” he says, “the scenario in which we’re now living.”

Intriguingly, says Friedkin, the same kind of bifurcation tends to show up even in the more sophisticated

forms of balance theory used today. In 2020, for example, a trio of researchers from Austria and Switzerland devised a balance-theory model in which the agents could have opinions on many issues, not just one (7). Their model also incorporated a parameter encoding the overall level of social polarization—or more precisely, how strongly the agents adhered to the famous fallacy, “if you’re not with me, you’re against me.” Lower that value, so that the agents are easily influenced by other views, and they would settle into a global consensus. But whenever the parameter was cranked up high enough to make the agents into intolerant absolutists, the artificial community would fracture into two antagonistic cliques.

“We’re very bad at reading the opinions and views of others. So we tend to exaggerate the ideological extremity on the other side, and minimize the ideological extremity of our own side.”

—Christopher Bail

Again, this kind of mutual loathing seems to echo today’s reality. But for David Garcia, a computer scientist at the Graz University of Technology in Vienna, Austria, and a coauthor of the 2020 article, among the model’s most fascinating findings is that the agents on one side end up agreeing with each other on a whole suite of issues—and despising everything the other side stands for. This, too, is strikingly similar to the real world, where you can reliably predict someone’s stance on a divisive issue like, say, climate change by knowing their position on similarly fraught issues such as gun control or abortion. Yet it’s happening in the model with “issues” that are just abstract labels, with none of the emotional resonance that people attach to real-world controversies—and that may very well end up grouped with a different set of labels in subsequent runs of the model.

This arbitrariness might not be as unrealistic as you’d think, says Garcia. Many researchers have tried to explain the observed partisan alignment on various issue by appealing to factors such as personal morality (8) or cognitive hardwiring (9). But then, he wonders, why are the alignments so different in other parts of the world? “In some countries ecology is highly correlated with being conservative,” he says, “as if it’s somehow defending the nature of the nation.” So how much of our left–right division isn’t about issues at all but is just the result of random chance?

Maybe a lot, says Macy. In 2019, he and three of his Cornell colleagues recruited more than 4,000 self-identified Republicans and Democrats and asked for their opinions on a number of “emerging controversies” (10). No one had a preexisting opinion because the issues had been made up for the experiment. But whenever the participants were allowed to see what earlier subjects had said before making their own choice, Republicans would unite on one side of the

issue whereas Democrats would unite just as passionately in opposition.

No surprise there—except that in subsequent trials asking different subjects about the same made-up issue, the two parties would often end up with their positions reversed. As long as people could see their fellow partisans' stances, says Macy, the first person to voice an opinion would tilt things in one or the other direction, and a self-reinforcing sectarian cascade would expand the divide from there. Or, as he and his coauthors put it, "what appear to be deep-rooted partisan divisions in our own world may have arisen through a tipping process that might just as easily have tipped the other way."

That said, the Cornell study also had a control group: subjects who were asked to give their opinions without knowing where anyone else came down. They generally took stances scattered in the middle—which is consistent with real-world data showing that the general population mostly couldn't care less about politics and isn't nearly as polarized on the issues as the two major US parties. As Yang points out, "there has been good evidence that a majority of the US voting public doesn't even know important facts about the candidates or what policies they propose."

Looking for Interventions

In last year's article on American sectarianism (1), the 15 social-scientist authors not only surveyed the rise of political hatred but also tackled the question of what to do about it.

They concluded that one of the most promising approaches is easy to state but hard to implement: Find ways to correct our misperceptions of people on the other side and become more open to their perspective (11). "We're very bad at reading the opinions and views of others," says Christopher Bail, a sociologist at Duke University in Durham, NC, and one of the article's coauthors. "So we tend to exaggerate the ideological extremity on the other side, and minimize the ideological extremity of our own side."—and this can make our differences seem much bigger than they actually are.

This bridge-building approach is certainly consistent with modeling results. In the multi-issue balance-theory model developed by Garcia and his colleagues, for example, polarization decreased markedly as the researchers lowered their for-us-or-against-us parameter. In a model of social media behavior developed by Sirianni and others last year (12), polarization went down when they cranked up the model's "open-mindedness" parameter, which measures how willing their agents were to consider other points of view. And in a new model (13) that Axelrod and two colleagues posted to ArXiv in March, polarization went away as the researchers raised a "tolerance" parameter that made the agents more accepting of other opinions.

This reach-across-the-aisle approach also seems to work in the real world—albeit in highly structured settings. Over the past decade, for example, there has been a surge of interest in various forms of deliberative democracy, in which people of multiple political persuasions are brought together for face-to-face

discussions in small groups (14–16). In most cases the participants are able to overcome their partisan antipathy and even reach some degree of consensus on hot-button issues such as abortion (17).

Antisocial Media

But just about everyone in this field is considerably less optimistic about proposals to reform social media. For one thing, it's not clear how effective any such reforms would be. Even though Facebook, Twitter, YouTube, and other platforms are widely viewed as vectors for misinformation and employed as partisan echo chambers, researchers are still arguing about how much they actually contribute to polarization (18). According to some studies, in fact, the algorithms that determine what users see in their feeds are just bit players; most of the online divisions come from people sorting themselves the way they always have, through "birds-of-a-feather" homophily (19).

For another thing, the reforms could easily backfire. In 2018, for example, Bail led a team that tested a frequently proposed idea for opening up the echo chambers (20). They paid more than 1,600 Republican and Democratic Twitter users to follow bots that would periodically show them tweets from figures in the opposite party. "The hope was that this would lead to moderation," says Bail. But in fact, he says, people mostly just recoiled from the discordant information. "Nobody became more moderate," he says. "And Republicans, in fact, became significantly more conservative."

And the real threats might not be the obvious ones. In the social media model (12) that Sirianni worked on last year, for example, he and his colleagues found that too much political advertising or campaigning might actually undermine a candidate. A campaign that pushes too hard may end up radicalizing its base, explains Sirianni. And because people tend to listen to political opinions that are somewhat similar to theirs, "those extreme voters might not be able to convince their centrist friends to vote for their candidate," he says. "Whereas if they were a little bit less extreme, maybe they could have brought people over."

Mäs and Karlsruhe sociologist Marijn Keijzer found essentially the same result earlier this year in a model of online bots (21). "Bots that have many followers and that are very aggressively posting falsehoods," says Mäs, "are less effective than bots that are not having many followers and only from time to time spread their content."

Although this finding has yet to be confirmed in real social media, write Mäs and Keijzer, it does suggest that policymakers and social-media engineers trying to limit the malicious spread of misinformation should avoid the temptation to focus only on the shrillest online voices and most active bots. Softer-spoken users and sparsely connected bots may be much harder to detect, yet far from innocent—and considerably more effective.

Conversely, though, the same lesson—avoid loud and dogmatic virtue-signaling—could be worth

remembering for people trying to promulgate truthful information about, say, vaccines or climate change.

Either way, says Mäs, it's crucial to remember that modelers are playing catch-up with Facebook, Twitter, YouTube, and other platforms. "They haven't understood the consequences of their technology," he says, "yet they provide a service to many, many individuals—who also influence each other in these

systems." In effect, the platforms are running a huge experiment all over the world, he says—with consequences that have yet to be determined.

Models can help researchers understand those consequences, says Mäs—if they are fed with a lot of top-quality empirical data. But tiny things can matter. "So before we have interventions, or before we play around with the algorithms," he says, "we have to be very, very careful."

- 1 E. J. Finkel *et al.*, Political sectarianism in America. *Science* **370**, 533–536 (2020).
- 2 R. P. Abelson, "Mathematical models of the distribution of attitudes under controversy" in *Contributions to Mathematical Psychology*, N. Fredericksen, H. Gullicksen, Eds. (Holt, Rinehart & Winston, 1964), pp. 1–54.
- 3 R. Axelrod, The dissemination of culture: A model with local convergence and global polarization. *J. Conflict Resolut.* **41**, 203–226 (1997).
- 4 M. M. Waldrop, News feature: Special agents offer modeling upgrade. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7176–7179 (2017).
- 5 F. Heider, Attitudes and cognitive organization. *J. Psychol.* **21**, 107–112 (1946).
- 6 D. Cartwright, F. Harary, Structural balance: A generalization of Heider's theory. *Psychol. Rev.* **63**, 277–293 (1956).
- 7 S. Schweighofer, F. Schweitzer, D. Garcia, A weighted balance model of opinion hyperpolarization. *J. Artif. Soc. Soc. Simul.* **23**, 5 (2020).
- 8 J. Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (Vintage, 2012).
- 9 S. Pinker, *The Blank Slate: The Modern Denial of Human Nature*, Reprint edition (Penguin Books, 2003).
- 10 M. Macy, S. Deri, A. Ruch, N. Tong, Opinion cascades and the unpredictability of partisan polarization. *Sci. Adv.* **5**, eaax0754 (2019).
- 11 J. Lees, M. Cikara, Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nat. Hum. Behav.* **4**, 279–286 (2020).
- 12 X. Wang, A. D. Sirianni, S. Tang, Z. Zheng, F. Fu, Public discourse and social network echo chambers driven by socio-cognitive biases. *Phys. Rev. X* **10**, 041042 (2020).
- 13 R. Axelrod, J. J. Daymude, S. Forrest, Preventing extreme polarization of political attitudes. *arXiv [cs.MA]* (2021).
- 14 N. Curato, J. S. Dryzek, S. A. Ercan, C. M. Hendriks, S. Niemeyer, Twelve key findings in deliberative democracy research. *Daedalus* **146**, 28–38 (2017).
- 15 America Talks (June 18, 2021). <https://americatalks.us/>. Accessed 27 August 2021.
- 16 Participedia (June 18, 2021). <https://participedia.net/>. Accessed 27 August 2021.
- 17 D. M. Farrell, J. Suiter, C. Harris, "Systematizing" constitutional deliberation: the 2016–18 citizens' assembly in Ireland. *Ir. Polit. Stud.* **34**, 113–123 (2019).
- 18 P. Barberá, "Social media, echo chambers, and political polarization" in *Social Media and Democracy: The State of the Field, Prospects for Reform*, N. Persily, J. A. E. Tucker, Eds. (Cambridge University Press, 2020), pp. 34–55.
- 19 E. Bakshy, S. Messing, L. A. Adamic, Political science. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
- 20 C. A. Bail *et al.*, Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9216–9221 (2018).
- 21 M. A. Keijzer, M. Mäs, The strength of weak bots. *Online Soc. Netw. Media* **21**, 100106 (2021).
- 22 S. Iyengar, Y. Lelkes, M. Levendusky, N. Malhotra, S. J. Westwood, The origins and consequences of affective polarization in the United States. *Annu. Rev. Polit. Sci.* **22**, 129–146 (2021).