AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Personality and Social Psychology:
Attitudes and Social Cognition

ISSN: 0022-3514

https://doi.org/10.1037/pspa0000469

Unnecessarily Divided: Civil Conversations Reduce Attitude Polarization More Than People Expect

Michael Kardas[1], Loran Nordgren[2], and Derek Rucker[2]
[1] Wisconsin School of Business, University of Wisconsin–Madison
[2] Kellogg School of Management, Northwestern University

People with opposing attitudes can learn from one another through civil discourse and debate. Yet, people routinely avoid discussing their differences of opinion, preferring instead to discuss their attitudes with like-minded others. We propose that people lack interest in discussing their differences of opinion, in part, because they expect such conversations are unlikely to change their own and others' attitudes. Importantly, we find these expectations are systematically miscalibrated: Civil conversations reduce attitude polarization more than people anticipate. Participants with opposing attitudes toward cats and dogs (Study 1 and Supplemental Study S1), cancel culture (Studies 2 and 4), and Joe Biden's performance as president (Study 5) underestimated how much their own and others' attitudes would depolarize in spoken conversations. Moreover, participants retained somewhat less polarized attitudes 1 week later. Participants underestimated attitude change, because they misunderstood why their attitudes differed: Whereas participants inferred their attitudes differed, because they fundamentally disagreed; their attitudes actually differed, because they were focused on different aspects of these topics (Study 3). As such, having a conversation surfaced unexpected areas of agreement (Studies 2, 4, and 5). Importantly, participants became more interested in discussing their differences of opinion, when they were informed that their own and others' attitudes might depolarize in a conversation (Study 6 and Supplemental Study S2). In total, the current work reveals that miscalibrated expectations can create an unnecessary barrier to civil discourse, leaving people with diverse points of view more divided, more polarized, and less informed than they otherwise could be.

Statement of Limitations

Our research has three main limitations. First, because participants completed the studies online, between 5% and 9% of participants dropped out after learning their partner's attitude and before having the conversation. Supplemental analyses, however, find that differences between predicted and actual depolarization—the focal outcome of interest—remain significant in all studies under explicitly conservative assumptions about the participants who dropped out. Second, although participants' attitudes remained somewhat less polarized 1 week after their conversations in all studies, we did not measure longer term outcomes. Finally, we recruited participants only from the United States and United Kingdom. Prior research suggests people from eastern cultures exhibit more dialectical thinking than those from western cultures—more tolerance for ideas that appear to contain contradictions (Peng & Nisbett, 1999)—and so people from eastern cultures might be more likely to recognize that an individual could hold either positive or negative attitudes toward an issue, depending on which aspects of the issue come to mind. If so, people from eastern cultures might be more likely to anticipate that conversations between people with opposing attitudes will reduce attitude polarization, compared to the western participants we recruited in these studies.

Keywords: attitude polarization, disagreement, conversation, accuracy, subjective construals

Supplemental materials: https://doi.org/10.1037/pspa0000469.supp

Paul Conway served as action editor.

The surveys, data, analysis script, Supplemental Materials, and preregistrations are available at https://osf.io/tgnjk/?view_only=81cedf8e423042e5ac2fb98d68e37bb0. Parts of this research have been presented at the annual conferences for the Society for Personality and Social Psychology, the Society for Judgment and Decision Making, and the Keeping the Republic Conference. The authors thank Nicholas Epley, Alex Shaw, and Anna Wisniewski for providing especially helpful feedback and Northwestern University's Kellogg School of Management for financial support.

Michael Kardas played a lead role in conceptualization, formal analysis, methodology, and writing–original draft. Loran Nordgren played a lead role in conceptualization and writing–review and editing and an equal role in methodology. Derek Rucker played a lead role in conceptualization and writing–review and editing and an equal role in methodology.

Correspondence concerning this article should be addressed to Michael Kardas, Wisconsin School of Business, University of Wisconsin–Madison, Grainger Hall, Room 4250B, Madison, WI 53715, United States. Email: mkardas@wisc.edu

> What indulgence should we not have … for opinions different from ours, when this difference often depends only upon the various points of view where circumstances have placed us! Let us enlighten those whom we judge insufficiently instructed; but first let us examine critically our own opinions and weigh with impartiality their respective probabilities. (Laplace, 1814/1956, p. 1328)

Society is becoming increasingly polarized. People in the United States, for example, are more likely to hold consistently liberal or consistently conservative positions today than in the past (Fiorina & Abrams, 2008; Jocker et al., 2024; Levendusky, 2009). They tend to live with, work with, and prioritize friendships with others who share their worldviews, creating "echo chambers" in which their views are more likely to be reinforced than to be challenged by others (Boutyline & Willer, 2017; Levy & Razin, 2019; McPherson et al., 2001). Rising levels of polarization may also be contributing to divisions in society. For example, people feel negatively toward others who do not share their political views, and this animosity toward the opposing party has risen over time (Iyengar et al., 2012). When people do encounter differences of opinion, they often keep their views to themselves rather than discuss them openly, maintaining mistrust across personal and political divides (Cowan & Baldassarri, 2018; Hutchens et al., 2019).

Does the polarization that exists in our society need to be a source of division? As argued by LaPlace, differences of opinion can serve as a source of enlightenment by enabling people to learn from diverse viewpoints. Civil conversations enable people to instruct one another, learn from one another, and form a more complete understanding of an issue (Fishkin et al., 2021). They build trust, mutual understanding, and are a crucial part of a healthy democracy (Mill, 1859). As such, this tension between the personal and political divisions that exist in our society, and the lack of civil conversations that could bridge these divisions, raises two questions that we investigate in this research. First, to what extent do people lack interest in discussing their differences of opinion because they expect such conversations are unlikely to bridge these differences? Second, to what extent are these expectations warranted?

In this article, we propose that people lack interest in discussing their differences of opinion partly because they expect that discussing them will not produce meaningful changes in their own or others' attitudes. Importantly, we further propose these expectations are systematically miscalibrated: People's attitudes depolarize more than they expect during conversations. We also explore the psychological mechanism underlying this misunderstanding, and in doing so, put forth means to encourage civil discourse and reduce polarization.

Psychological Barriers to Discussing Differences of Opinion

Social scientists have established that people interact more often with others who are similar to them than with others who are different, a tendency known as homophily. For example, people tend to interact with others who are similar to them in terms of age, ethnicity, and education (McPherson et al., 2001). People also interact more often with others who share their political attitudes (Oosterhoff et al., 2022), their religious beliefs (Liao & Stevens, 1994), and their personal preferences (Bainbridge & Stark, 1981) than with those who do not. Such interactions cause people's attitudes to become more polarized, because they selectively expose people to information that supports their views (Bishop & Myers, 1974; Levy & Razin, 2019; Myers & Lamm, 1976), and because people give more weight to information that supports their views than to information that challenges those views (Lord et al., 1979).

Research suggests this preference for interacting with similar others arises partly from an expectation that interactions with dissimilar others lead to a variety of unfruitful, and even negative, outcomes. For example, people expect that listening to opposing views will be unpleasant (Dorison et al., 2019), that others who do not share their views will respond negatively to them (Wald et al., 2024), and that they will not feel heard by others during a conversation (Teeny & Petty, 2022). Most relevant to our research, people are less interested in discussing their differences of opinion when they expect these interactions to produce little attitude change. In one study, for example, participants were less interested in advocating on behalf of their attitudes toward the death penalty, when they expected that another person's attitude was unlikely to change than when they expected that the person's attitude was likely to change (Akhtar & Wheeler, 2016).

Although pessimistic expectations of attitude change could reduce people's interest in discussing their differences of opinion, research has not investigated whether these pessimistic expectations are well calibrated to the actual changes people experience in a conversation. In this research, we hypothesize that people systematically underestimate how much their own and others' attitudes are likely to depolarize during a conversation, creating a potentially unnecessary barrier to discussing diverse points of view.

Miscalibrated Expectations Across Personal and Political Divides

To illustrate why expected and actual attitude change may diverge in conversations, consider two people with opposing attitudes toward Joe Biden's presidency, with one person approving and the other disapproving of his job performance. When these people learn that they have opposing attitudes, they may expect that a conversation is unlikely to narrow their difference of opinion, because they make what, at first blush, appears to be a reasonable inference: that their attitudes differ because they disagree. They might, for instance, infer that they disagree about Biden's domestic policies, his foreign policies, or his leadership abilities. They might draw the correspondent inference that they have different values, different beliefs, and different priorities (Gilbert & Malone, 1995; Jones & Harris, 1967), or that their counterpart must be stubborn, narrow-minded, and irrational (Haslam & Loughnan, 2014; Kennedy & Pronin, 2008; Pronin, Lin, & Ross, 2002). Because genuine disagreements about a contentious issue may be difficult, if not impossible, to resolve (Graham et al., 2009; Skitka et al., 2005), people may expect that a conversation will not draw their attitudes closer together, and they may choose not to discuss their differences as a result (Akhtar & Wheeler, 2016; Rattan & Georgeac, 2017).

Whereas people with different attitudes may readily infer that they disagree, we propose they underestimate how much their attitudes will depolarize because they fail to consider another explanation of why their attitudes differ. Specifically, people may hold different attitudes not only because they disagree, but largely because they are focused on different aspects of an issue—that is, they may be construing the issue differently (Brundage et al., 2024; Enke, 2020; Larrick et al., 2012; Lord & Lepper, 1999; Wilson &

Hodges, 1992). For example, one person might approve of Biden's presidency because they are focused on the nation's low unemployment rate or the passage of a bipartisan infrastructure bill. The other person might disapprove of Biden's presidency because they are focused on the nation's high inflation rate or the rise in undocumented immigration across the nation's southern border. Whereas these people may assume they have opposing attitudes because they have a point-by-point disagreement about Biden's entire presidency, they may actually have opposing attitudes because they brought to mind different aspects of Biden's presidency. In the words of the psychologist Solomon Asch (1952), they might assume they have fundamentally different "judgments of the object"—a deep-seated disagreement about the topic—when in fact they may be focused on different "objects of judgment"—different aspects of the topic. If these people were to discuss their attitudes, they might discover they largely agree about which areas of Biden's presidency they approve of and which they disapprove of, such that the conversation dissolves much of their apparent disagreement and draws their attention to different facets of his presidency that they had not previously considered.

Synthesizing these ideas, we hypothesized that people with opposing attitudes have miscalibrated expectations about discussing their differences of opinion. Whereas people may lack interest in discussing their differences of opinion partly because they expect their own and others' attitudes to change relatively little, having a conversation may reveal unexpected areas of agreement and cause their attitudes to depolarize more than they expected. This should occur not because people underestimate the potential of conversations to *resolve* disagreements about an issue as a whole, but because people underestimate the potential of conversations to *dissolve* disagreements by revealing they were focused on different aspects of the issue.

Three lines of research offer potential support for this hypothesis. First, research finds that the "false consensus effect"—in which people overestimate how much others will share their preferences and beliefs—is stronger for topics that can be construed in multiple ways, because people overlook alternative construals that cause their own and others' evaluations to diverge (Gilovich, 1990; see also Griffin et al., 1990). Second, research finds that people expect others to be less likely to comply with requests for help than others actually are, suggesting people may underestimate their capacity to influence others' behavior (Bohns, 2016; Zhao & Epley, 2022). Third, research on "false polarization" finds that people expect members of opposing groups, such as political parties, to have more polarized attitudes than they actually do (Fernbach & Van Boven, 2022; Robinson et al., 1995; Westfall et al., 2015). People are especially likely to overestimate how much others will disagree with them about values that are central to their own ideology (Chambers et al., 2006; Chambers & Melnyk, 2006).

We advance prior research by measuring predicted and actual attitude change in spoken conversations between participants who are accurately informed of the extremity of each other's attitudes. By assessing both expected and actual change, our research offers the first test of whether participants underestimate how much their attitudes will depolarize, as well as whether this occurs because they attribute differences in their attitudes to fundamental disagreements as opposed to differences in how they are construing an issue. In doing so, our research also tests a mechanism with potentially wide-ranging implications for understanding misperceptions of conflict, polarization, and group interactions.[1]

Overview of Studies

We conducted eight preregistered studies investigating whether people with opposing attitudes may be unnecessarily divided because they underestimate how much their attitudes will depolarize during spoken conversations. We measured predicted and actual attitude change in conversations about cats and dogs (Study 1 and Supplemental Study S1), cancel culture (Studies 2 and 4), and Joe Biden's job performance as president (Study 5). We also tested whether participants underestimate how much their own and others' attitudes will depolarize because they attribute differences in their attitudes to disagreements rather than to differences in how they are construing an issue, such that they might underestimate how much they will agree both in the static context of a survey and in back-and-forth conversations (Studies 2–5). Finally, we tested whether participants become more interested in discussing their differences of opinion when they are informed that their own and others' attitudes may depolarize during a conversation (Study 6 and Supplemental Study S2).

Study 1: Personal Preferences

Method

Transparency and Openness

Because our studies are the first to measure both predicted and actual attitude change in spoken conversations, and the first to use our conversation topics and procedures, we had no data available to conduct a priori power analyses. We therefore targeted at least 100 participants in each study with spoken conversations, and at least 100 participants per condition in our experiments. We then performed sensitivity power analyses using SIMR for studies with mixed linear models (Green & MacLeod, 2016), and using G*Power for all other studies (Version 3.1.9.4: Faul et al., 2007), to estimate the minimum effect sizes that our samples could detect with 80% probability. All analyses were performed using R, Version 4.3.1 (R Core Team, 2023), except mediational analyses which were performed using IBM SPSS Statistics, Version 29 (IBM Corp., 2023). The Supplemental Materials indicate which results, if any, change in their statistical significance when analyzing all participants who completed our studies without technical difficulties, whether or not they met our preregistered inclusion criteria. The surveys, data, analysis script, Supplemental Materials, and preregistrations for all studies are available on the Open Science Framework at https://osf.io/tgnjk/?view_only=81cedf8e423042e5ac2fb98d68e37bb0 (Kardas et al., 2025).

Our research follows the American Psychological Association's journal article reporting standards for quantitative research in psychology (see Appelbaum et al., 2018). We report all measures, manipulations, and data exclusions throughout the article. All

[1] We preregistered a different hypothesis in Study 1—namely, that people on both sides of an issue would expect to persuade their counterpart more than they would expect their counterpart to persuade them, consistent with naive realism (Ross, 2013). This alternative hypothesis was not supported in Study 1 or any of the following studies, however, leading us to conduct preregistered tests of the theory described throughout the introduction in Studies 2–6. We revisit research on naive realism in the General Discussion section.

studies were approved by the university's institutional review board. We obtained informed consent from all participants.

Recruitment

For studies with spoken conversations, we recruited participants from Prolific (n = 742, combined across Studies 1, 2, 4, 5, and Supplemental Study S1) and Amazon's Mechanical Turk (n = 8 in Study 1), with the restriction that no participant was allowed to complete more than one study. In Studies 1, 2, 4, and Supplemental Study S1, we posted the study online and allowed any eligible participants to enroll in real time. In Studies 1, 5, and Supplemental Study S1, we additionally used prescreen surveys in which participants verified that they could connect their camera and microphone to the survey and indicated which of several times they would be available for a study session. We then opened the study at the specified times only for participants who had signed up in the prescreen. All procedures took place in the Qualtrics survey software. Participants were matched with a partner using SMARTRIQS (Molnar, 2019) and interacted through a video feed embedded directly in the survey using SurvConf (Brodsky et al., 2022). Conducting our research online enabled us to recruit large groups of participants in each session, ensuring that enough participants would fall on both sides of the issue to be matched for conversations. It also enabled us to reach a more demographically and politically diverse sample of adults from around the United States (Studies 1, 2, 5, and Supplemental Study S1) and United Kingdom (Study 4) than we could recruit in a university research laboratory.

One concern about conducting our research online is that participants might be more likely to drop out of a study being conducted online, compared to a study being conducted in the laboratory, upon learning that their partner's attitude toward the conversation topic differs from their own. In our studies with spoken conversations, between 5% and 9% of participants dropped out between learning their partner's attitude and having the conversation. Supplemental analyses find that differences between predicted and actual depolarization remain significant in all studies under the null hypothesis that these participants who dropped out would have predicted as much depolarization as their partners predicted and would then have had conversations in which both people's attitudes depolarized exactly as much as they both predicted (i.e., no differences between predicted and actual attitude change—ts > 3.85, ps < .001). They also remain significant in all studies under the explicitly conservative hypothesis that these participants who dropped out would have predicted as much depolarization as their partners predicted and would then have had conversations in which neither person's attitude depolarized (i.e., overestimating attitude change—ts > 2.34, ps < .020; see Supplemental Materials for details). Our results are thus robust to even conservative assumptions about the participants who dropped out after learning their partner's attitude.

Participants

In Study 1, we recruited participants from the U.S. participant pools on Prolific and Amazon's Mechanical Turk to complete the main session in exchange for $5.00 and the follow-up survey in exchange for $1.00. One hundred fifty-four participants completed the main session without technical difficulties (n = 146 from Prolific, n = 8 from Amazon's Mechanical Turk). We excluded four participants from analyses: Two because they or their partner were not sufficiently proficient in English to have the conversation, and two because they had already completed the study earlier in data collection. This left a final sample of 150 participants after data exclusions (M_{age} = 40.05, SD_{age} = 13.27; 44.67% female, 52.67% male, 2.67% other gender; 74.67% White, 4.00% Black, 3.33% Hispanic, 9.33% Asian, 0.67% American Indian, 8.00% other ethnicity), which provided about 80% power to detect differences between predicted and actual attitude change of size b = 0.22.

Procedure

Participants connected to the survey using their laptop or desktop computers. They were informed that they would be asked to have a spoken conversation with another participant, were instructed to connect their camera and microphone to the survey, and provided informed consent. Participants who could not connect their camera or microphone were not allowed to continue.

Participants reported the extent to which they currently believe that cats or dogs are the better pets (−3 = *cats are much better*, −2 = *cats are somewhat better*, −1 = *cats are slightly better*, 0 = *both are equally good*, 1 = *dogs are slightly better*, 2 = *dogs are somewhat better*, 3 = *dogs are much better*, *I don't know*). Participants were then told that they would have a live, 10-min conversation about cats and dogs with another participant. As required by our institutional review board, they were told that they should stay on-task throughout the conversation, should not request or share personally identifying information, and should not use profanity, insults, or bully one another. In each study, we provided this information at the start of the procedure to ensure that participants were fully informed of what they would be asked to do before reporting predictions or having the conversation.

Participants were then matched with one another on a first-come-first-served basis, such that each pair included one participant who thought that cats were better and one who thought that dogs were better, regardless of the extremity of their attitudes. Participants who did not receive a match within 3 min, and those who selected "both are equally good" or "I don't know," were dismissed from the study in exchange for partial payment.

After receiving a match, participants viewed both their own and their partner's attitudes in the survey. They then read that they would be asked to discuss the following prompts during their 10-min conversation:

- Why you believe that [cats are much better]
- Why the other participant believes that [dogs are much better]
- What you think of each other's attitudes toward cats and dogs

After reading the prompts, participants completed comprehension checks in which they selected their own attitude and the other participant's attitude. Participants who responded incorrectly to either comprehension check were prompted to try again until they selected the correct attitudes.[2]

Next, participants made several predictions about the conversation. They first predicted what attitude they would report toward cats

[2] We included these comprehension checks in all studies but omit them from the following procedure sections for brevity.

and dogs after the conversation, and what attitude the other participant would report after the conversation, on separate scales (−3 = *cats are much better*, −2 = *cats are somewhat better*, −1 = *cats are slightly better*, 0 = *both are equally good*, 1 = *dogs are slightly better*, 2 = *dogs are somewhat better*, 3 = *dogs are much better*). The participant's original attitude, the partner's original attitude, and the conversation prompts were displayed directly above these items, ensuring that participants were fully informed of what they would be asked to discuss while making these predictions.

Because we sought to understand the causes of any systematic differences between predicted and actual attitude change, participants then completed several exploratory measures. They predicted how strong they would rate their own reasons to be after the conversation, and how strong the other participant would rate these reasons to be. Participants then predicted how strong they would rate the other participant's reasons to be after the conversation, and how strong the other participant would rate these reasons to be (0 = *not strong at all*, 3 = *somewhat strong*, 6 = *very strong*). Participants then indicated the extent to which they think they currently understand or misunderstand the other participant's perspective about cats and dogs (−3 = *I completely misunderstand their perspective*, 0 = *equally both*, 3 = *I completely understand their perspective*), and the extent to which they think the other participant currently understands or misunderstands one's own perspective (−3 = *the other participant completely misunderstands my perspective*, 0 = *equally both*, 3 = *the other participant completely understands my perspective*).

Participants then indicated the extent to which they think the difference between their own attitude and their partner's attitude represents an objective disagreement or a subjective difference of opinion (−3 = *completely an objective disagreement*, 0 = *equally both*, 3 = *completely a subjective difference of opinion*). To measure how certain participants were of their attitudes (Petrocelli et al., 2007), participants indicated how certain they were that the attitude they reported was really their true attitude toward cats and dogs, and how certain they were that the attitude they reported was the right way to think and feel about cats and dogs (0 = *not certain at all*, 3 = *somewhat certain*, 6 = *very certain*).

After completing these items, participants advanced the page and connected to the video feed, which was embedded in the survey. The conversation prompts were displayed directly above the video feed. After both participants connected to the video feed, a timer in the corner of the feed began counting down from 10 min. Participants were allowed to end the conversation at any time by clicking a checkbox at the bottom of the screen and then advancing the page, but 70 of 75 pairs talked for all 10 min.[3] When the timer expired, the video call ended and the survey automatically advanced to the next page. The conversations were not monitored in real time by the researchers in any of our studies, but video recordings were uploaded to a secure webpage immediately after each conversation ended. We manually reviewed the video recordings to verify that both participants explained their attitudes toward cats and dogs (as indicated in our preregistration).

After the video call ended, participants reported their own attitude toward cats and dogs, and estimated their partner's current attitude, using the same scales on which they had reported predictions. They then completed the remaining measures that they had completed before the conversation, phrased in the past tense to refer to the conversation they just had. We did not include measures of attitude certainty after the conversation, because our intention was to explore whether participants' certainty in their original attitudes (before the conversation) would moderate differences between predicted and actual attitude change.

After completing these measures, participants indicated how much personal experience they had with cats, how much personal experience they had with dogs (0 = *no personal experience*, 3 = *some personal experience*, 6 = *very much personal experience*), and completed the 10-item personality inventory (Gosling et al., 2003). Participants then indicated whether they had any issues during their conversation, reported demographic information, and were paid for their participation.

Follow-Up Survey

To assess whether any changes in the participants' attitudes from before to after the conversation persisted over time, we contacted participants 1 week after each session and asked them to complete a follow-up survey, without reminding them of the attitudes they reported in the main session. If a participant did not complete the follow-up survey within 7 days, we sent them another reminder. In the follow-up survey, participants first reported their current attitude toward cats and dogs. After submitting their response, participants were asked to think about the other participant with whom they had a conversation, and they predicted this other participant's current attitude. Finally, participants were debriefed about the purpose of the research and were paid for the follow-up survey.

Results and Discussion

Predicted and Actual Attitude Change

For each participant, we computed four difference scores: how much the participant *expected* their own attitude to change from before to after the conversation, how much the participant *expected* their partner's attitude to change, how much the participant's attitude *actually* changed from before to after the conversation, and how much their partner's attitude *actually* changed. Positive numbers represented changes in the direction of each other's initial attitudes, and negative numbers represented changes away from each other's initial attitudes. We used these difference scores as dependent measures in the mixed linear models described below.

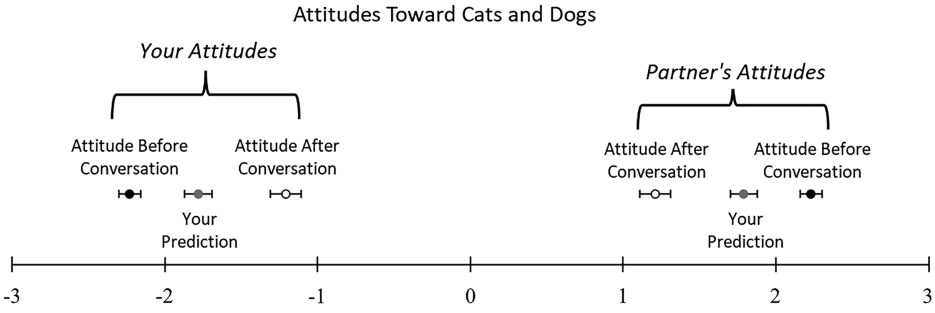
As seen in Figure 1, participants' predictions significantly underestimated how much their own and others' attitudes would depolarize.[4]

[3] Combined across the five studies with spoken conversations, 93% of participants talked for the entire 10-min interval. The degree to which participants' initial attitudes differed at baseline (before the conversation) did not differ significantly between participants who talked for all 10 min and those who ended their conversations early in any studies (all ts < 1.29, ps > .204). Differences between predicted and actual depolarization were not qualified by significant two-way interactions with the length of the conversation (talked for 10 min vs. fewer than 10 min) in any studies (all ts < 1.12, ps > .266). Finally, although all participants reported in the text met our preregistered inclusion criteria, differences between predicted and actual depolarization remain statistically significant in all studies when excluding participants who talked for fewer than 10 min (all ts > 4.02, ps < .001).

[4] As preregistered, we also compared participants' predictions of how much their own attitude and the other person's attitude would depolarize. Although participants predicted that their own and others' attitudes would depolarize slightly (b = 0.46, SE = 0.05), t(76.15) = 8.43, p < .001, 95% CI [0.35, 0.57], they did not predict that the other person's attitude would depolarize significantly more than their own attitude (b = 0.04, SE = 0.09), t(150.00) = 0.46, p = .649, 95% CI [−0.13, 0.21].

Figure 1

Mean Preconversation Attitudes, Predicted Attitudes, and Postconversation Attitudes for Oneself and One's Conversation Partner in Study 1



Note. We reverse coded the attitudes of participants who initially thought dogs were the better pets, so that negative numbers represent attitudes consistent with one's own position and positive numbers represent attitudes inconsistent with one's own position for all participants. Error bars represent ±1 standard error.

We constructed a mixed linear model with attitude change as the dependent variable, with fixed-effect terms for measurement type (predicted, actual), target (self, other), side (cats are better, dogs are better), and their interactions, with each of these variables centered around zero, and with random-intercept terms for pair number and for participants nested within pairs.[5] Whereas participants expected the conversations to narrow the divide between their initial attitudes by 20%, the conversations actually narrowed this divide by 46% on average, resulting in a significant effect of measurement type ($b = 0.57$, $SE = 0.08$), $t(523.05) = 7.40$, $p < .001$, 95% confidence interval (CI) [0.42, 0.72]. Differences between predicted and actual depolarization did not vary for one's own attitude or the partner's attitude, as indicated by a nonsignificant Measurement Type × Target interaction effect ($b = -0.03$, $SE = 0.15$), $t(523.05) = -0.22$, $p = .829$, 95% CI [−0.34, 0.27], and did not vary between participants who initially thought that cats or dogs were the better pets, as indicated by a nonsignificant Measurement Type × Side interaction effect ($b = -0.07$, $SE = 0.15$), $t(523.05) = -0.48$, $p = .634$, 95% CI [−0.38, 0.23], and a nonsignificant three-way interaction with target ($b = -0.41$, $SE = 0.31$), $t(523.05) = -1.34$, $p = .181$, 95% CI [−1.02, 0.19].[6]

Participants not only underestimated how much their own and their partner's attitudes would depolarize, but they also underestimated how much they personally would perceive their partner's attitude to have depolarized after the conversation. Specifically, participants estimated that their partner's attitude had depolarized significantly more after the conversation than they had predicted before the conversation ($Ms_{depolarization} = 0.87$ vs. 0.48, respectively; SDs = 1.03 vs. 0.97, $b = 0.39$, $SE = 0.10$), $t(150.00) = 3.88$, $p < .001$, 95% $CI_{difference}$ [0.19, 0.59], providing convergent evidence that participants underestimated changes in each other's attitudes before the conversation. Participants' estimates of how much their partners' attitudes had depolarized after the conversation did not differ significantly from how much their partners' attitudes had actually depolarized (Ms = 0.87 vs. 1.03, respectively; SDs = 1.03 vs. 1.15, $b = 0.16$, $SE = 0.11$), $t(150.00) = 1.51$, $p = .134$, 95% $CI_{difference}$ [−0.05, 0.37], likely because the conversations provided feedback that helped to calibrate the participants' inferences about each other's attitudes.

Exploratory analyses of our secondary measures did not find evidence of why participants underestimated how much their own and others' attitudes would depolarize (see Supplemental Materials). We continue to investigate potential explanations in the following studies. Our findings were not consistently moderated by the strength of the participants' attitudes in any of our studies, and so we report analyses of attitude strength in the Supplemental Materials.

Follow-Up Survey

Nearly all participants (149 of 150) completed the follow-up survey 1 week after their session. Participants' attitudes in the follow-up survey were more polarized than those they reported immediately after their conversations ($b = 0.63$, $SE = 0.08$), $t(149.34) = 7.62$, $p < .001$, 95% CI [0.46, 0.79], but were significantly less polarized than those they reported at baseline ($b = -0.39$, $SE = 0.07$), $t(149.36) = -5.44$, $p < .001$, 95% CI [−0.53, −0.25]. Importantly, the more that participants' own attitudes depolarized in the main session, and the more they underestimated how much their own attitudes would depolarize in the main session, the more moderate their attitudes remained 1 week later compared to the attitudes they reported at baseline, suggesting the conversations brought about changes in the participants' attitudes that remained detectable at least a week later (correlation between actual change in main session and sustained change at follow-up: ($b = 0.41$, $SE = 0.05$), $t(150.55) = 7.64$, $p < .001$, 95% CI [0.30, 0.51]; correlation between underestimation of change in main session and sustained change at follow-up: ($b = 0.27$, $SE = 0.06$), $t(151.12) = 4.28$, $p < .001$, 95% CI [0.14, 0.39]).

One alternative interpretation of the results of Study 1 is that the within-participants design—in which the same participants reported predicted attitudes (before the conversation) and actual attitudes (after

[5] Our preregistered models did not include fixed-effect terms for which side of the issue the participants were on. We include these terms in the main text, because we recognized their importance after conducting our studies. The preregistered models similarly find significant differences between predicted and actual depolarization in all studies and are reported in the Supplemental Materials.

[6] Very few results are qualified by significant interactions with self versus other or with the side of the issue that the participants were on. In the following analyses, we therefore report these interaction effects only when they are statistically significant to streamline the text.

the conversation)—could have induced a demand characteristic whereby participants presumed they should report a different attitude after the conversation than they had predicted before the conversation. Such a demand characteristic could create a pull toward more moderate attitudes, not because the participants' attitudes genuinely depolarized, but rather because they could have felt pressure to deviate from their initial predictions.

Although multiple observations in Study 1 are consistent with genuine attitude change—including the agreement between self-reported and partner-reported attitude change after the conversation, and the persistence of attitude change 1 week later—we conducted a preregistered, Supplemental Study S1 to test this alternative explanation directly (see Supplemental Materials for the full method and results). In this supplemental experiment (N = 200), we experimentally manipulated whether participants reported predictions before having a conversation. Pairs in the *predictions-and-experiences* condition followed a similar procedure as Study 1, in that they reported predicted attitudes before the conversation and actual attitudes after the conversation. Pairs in the *experiences-only* condition followed the same procedure except that they did not report predictions before the conversation.

The results of this supplemental experiment supported our hypotheses. Participants in the predictions-and-experiences condition anticipated significantly less depolarization before the conversation than they reported after the conversation, replicating the within-participants finding of Study 1, (b = 0.39, SE = 0.08), $t(350.00) = 4.66$, $p < .001$, 95% CI [0.22, 0.55]. Critically, the hypothesis was also supported in the between-participants comparisons: Participants in the experiences-only condition—who did not report predictions before the conversation—also reported significantly less polarized attitudes after the conversation than participants predicted in the predictions-and-experiences condition, (b = 0.46, SE = 0.11), $t(100.00) = 4.16$, $p < .001$, 95% CI [0.24, 0.67]. Participants' attitudes depolarized to a similar extent regardless of whether they reported predictions before the conversation, ($b = 0.07$, $SE = 0.14$), $t(200.00) = 0.50$, $p = .616$, 95% CI [−0.20, 0.34], suggesting that repeated measurements of predicted and actual attitudes do not systematically alter the attitudes participants report after a conversation.

Secondary analyses further suggest that participants' attitudes genuinely depolarized more than expected. Participants in the predictions-and-experiences condition not only reported that *their own* attitudes depolarized more than expected, but they also judged *each other's* attitudes to have depolarized more after the conversation than they predicted before the conversation ($b = 0.27$, SE = 0.10), $t(100.00) = 2.80$, $p = .006$, 95% CI [0.08, 0.46]. Moreover, after the conversation, participants' estimates of how much their partners' attitudes had depolarized did not differ significantly from how much the partners self-reported that their attitudes had actually depolarized ($b = 0.03$, $SE = 0.09$), $t(200.00) = 0.28$, $p = .778$, 95% CI [−0.15, 0.20], providing convergent evidence of attitude change from both self-report and partner-report data. Finally, participants' attitudes remained somewhat less polarized in a follow-up survey 1 week later (b = −0.48, $SE = 0.06$), $t(199.56) = -7.47$, $p < .001$, 95% CI [−0.61, −0.35], with no differences in this result across experimental conditions, $|t\text{s}| < 0.09$, $p\text{s} > .931$. These findings suggest participants' attitudes genuinely depolarized more than expected during their conversations, and that the act of reporting predictions did not systematically inflate their reported attitude change.

Although this supplemental experiment helps to rule *out* an alternative interpretation of our results, Study 1 and Supplemental Study S1 do not rule *in* the psychological mechanisms that explain the miscalibration between expected and actual attitude change. Studies 2–4 investigate potential mechanisms in the context of a divisive social issue: cancel culture (Mueller, 2021).

Study 2: Social Divides

Method

Participants

We recruited participants from the U.S. participant pool on Prolific to complete the main session in exchange for $5.00 and the follow-up survey in exchange for $1.00. One hundred fourteen participants completed the main session without technical difficulties. We excluded 14 participants from analyses: 10 because they or their partner did not state their attitude during the 30-s video call (which preceded the 10-min conversation), and four because they or their partner stated a different attitude during the 30-s video call than they had selected in the survey. This left a final sample of 100 participants after data exclusions (M_{age} = 41.98, SD_{age} = 14.40; 47.00% female, 53.00% male; 79.00% White, 6.00% Black, 4.00% Hispanic, 5.00% Asian, 6.00% other ethnicity; 61.00% liberal, 31.00% conservative, 8.00% moderate), which provided about 80% power to detect differences between predicted and actual attitude change of size b = 0.26.

Procedure

After providing informed consent, participants read that "cancel culture is a movement to remove celebrity status or esteem from a person, place, or thing in response to objectionable behavior." They rated the extent to which they support or oppose cancel culture (−3 = *strongly oppose*, −2 = *somewhat oppose*, −1 = *slightly oppose*, 0 = *neither support nor oppose*, 1 = *slightly support*, 2 = *somewhat support*, 3 = *strongly support*, *I don't know*). Participants were then matched with one another, such that each pair included one supporter and one opponent of cancel culture.

Unlike Study 1, participants did not learn each other's attitudes by reading them in the survey. Instead, participants were instructed to tell each other their attitudes in a 30-s video call that took place several minutes before the 10-min conversation. We made this change to ensure that participants would know their study partner was real before they made predictions about the 10-min conversation, and to test whether we would replicate our findings when participants have already seen and heard from their partner before making predictions. Participants connected to the video feed told each other their attitudes without explaining their reasoning (e.g., "I strongly oppose cancel culture" and "I strongly support cancel culture") and immediately advanced the page to end the video call whether or not all 30 s had passed.

After this exchange, participants were informed that they would have a 10-min conversation about cancel culture with this other participant. They were told that they should discuss the following prompts:

- Why you [strongly oppose] cancel culture
- Why the other participant [strongly supports] cancel culture
- What you think of each other's attitudes toward cancel culture

Participants then made a series of predictions about the conversation. They first predicted what attitude they would report toward cancel culture after the conversation, and what attitude their partner would report, on separate scales (−3 = *strongly oppose*, −2 = *somewhat oppose*, −1 = *slightly oppose*, 0 = *neither support nor oppose*, 1 = *slightly support*, 2 = *somewhat support*, 3 = *strongly support*). The definition of cancel culture, both participants' original attitudes, and the conversation prompts were displayed directly above these items, ensuring that participants were fully informed of what they would be asked to discuss while making these predictions.

Participants then completed measures of several potential mediators. Participants predicted how much they would disagree or agree with the other participant's reasons, and how much the other participant would disagree or agree with one's own reasons, on separate scales (−3 = *strongly disagree*, −2 = *somewhat disagree*, −1 = *slightly disagree*, 0 = *equally both*, 1 = *slightly agree*, 2 = *somewhat agree*, 3 = *strongly agree*). They then predicted how hard they would try to persuade the other participant (0 = *not hard at all*, 3 = *somewhat hard*, 6 = *very hard*), how hard they would try to defend their own point of view (0 = *not hard at all*, 3 = *somewhat hard*, 6 = *very hard*), how hard they would try to understand the other participant's point of view (0 = *not hard at all*, 3 = *somewhat hard*, 6 = *very hard*), and how open and receptive they would be to the other participant's point of view (0 = *not at all open and receptive*, 3 = *somewhat open and receptive*, 6 = *very open and receptive*). Participants then completed similar items about the other person, predicting how hard the other participant would try to persuade oneself, defend their point of view, try to understand oneself, and how open and receptive the other participant would be to one's own point of view.

Participants then indicated how much they think the difference between their own attitude and their partner's attitude represents an objective disagreement or a subjective difference of opinion (−3 = *completely an objective disagreement*, 0 = *equally both*, 3 = *completely a subjective difference of opinion*). To measure the strength of the participants' attitudes, participants rated how important they consider the topic of cancel culture to be (0 = *not important at all*, 3 = *somewhat important*, 6 = *very important*), how certain they are that the attitude they reported is really their true attitude toward cancel culture, and how certain they are that the attitude they reported is the right way to think and feel about cancel culture (0 = *not certain at all*, 3 = *somewhat certain*, 6 = *very certain*).

Participants then had their 10-min conversations, with the definition of cancel culture and the conversation prompts displayed directly above the video feed. After the conversation, participants rated their attitude toward cancel culture using the same scale on which they had reported predictions, estimated their partner's current attitude, and completed the other items analogous to those they had completed before the conversation. We did not include measures of attitude strength after the conversation. Participants then rated their political orientation (−3 = *very liberal*, −2 = *somewhat liberal*, −1 = *slightly liberal*, 0 = *moderate*, 1 = *slightly conservative*, 2 = *somewhat conservative*, 3 = *very conservative*, *other*), indicated whether they had any issues during their conversation, reported demographic information, and were paid for their participation.

Follow-Up Survey

We contacted participants 1 week after their session and asked them to complete a follow-up survey, without reminding them of the attitudes they reported in the main session. In this survey, the participants rated their current attitude toward cancel culture and predicted the other person's current attitude, using a similar procedure as the follow-up survey in Study 1. Finally, participants were debriefed about the purpose of the research and were paid for the follow-up survey.

Results and Discussion

Predicted and Actual Attitude Change

As seen in Figure 2, participants underestimated how much their own attitude and their partner's attitude would depolarize. Whereas participants expected the conversations to narrow the divide between their initial attitudes by 22%, the conversations actually narrowed this divide by 41% on average, resulting in a significant effect of measurement type using the same mixed linear model as Study 1 ($b = 0.37$, $SE = 0.09$), $t(350.00) = 4.00$, $p < .001$, 95% CI [0.19, 0.55]. Differences between predicted and actual depolarization did not vary for one's own attitude versus the partner's attitude or for supporters versus opponents, as indicated by nonsignificant interactions, $|t|s < 1.63$, $ps > .104$.

Participants not only underestimated how much their own and their partner's attitudes would depolarize, but they also underestimated how much they personally would perceive their partner's attitude to have depolarized after the conversation. Specifically, participants estimated that their partner's attitude had depolarized significantly more after the conversation than they had predicted before the conversation ($Ms_{\text{depolarization}} = 0.98$ vs. 0.44, respectively; $SDs = 1.11$ vs. 0.77, $b = 0.54$, $SE = 0.12$), $t(100.00) = 4.57$, $p < .001$, 95% $CI_{\text{difference}}$ [0.31, 0.77], providing convergent evidence that participants underestimated changes in each other's attitudes before the conversation. Participants' estimates after the conversation did not differ significantly from how much their partners' attitudes had actually depolarized ($b = -0.18$, $SE = 0.12$), $t(100.00) = -1.49$, $p = .140$, 95% $CI_{\text{difference}}$ [−0.42, 0.06], suggesting the conversations provided feedback that helped to calibrate the participants' inferences about each other's attitudes.

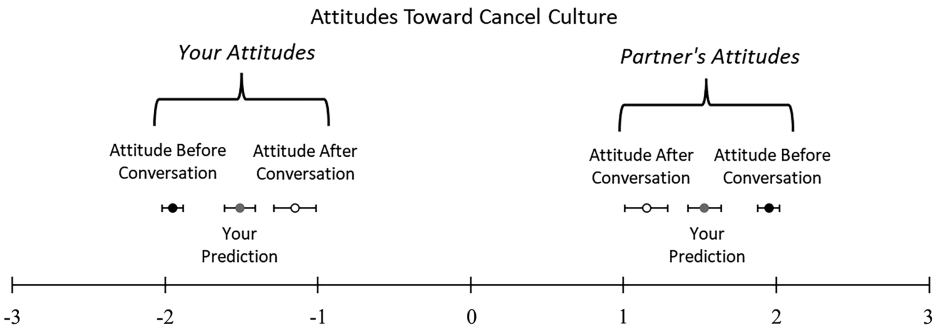
Mediators

To understand why participants underestimated how much their own and others' attitudes would depolarize, we analyzed the potential mediators. As seen in Table 1, participants significantly underestimated how much they would agree with each other's reasons, underestimated how hard they would try to understand one another, underestimated how receptive they would be to one another, underestimated how hard they would try to defend their own perspective, overestimated how hard their partner would try to defend their perspective, and overestimated how hard their partner would try to persuade them.

Because participants mispredicted several outcomes of the conversation, each of which could potentially explain why they underestimated

Figure 2

Mean Preconversation Attitudes, Predicted Attitudes, and Postconversation Attitudes for Oneself and One's Conversation Partner in Study 2



Note. We reverse coded the attitudes of participants who initially supported cancel culture, so that negative numbers represent attitudes consistent with one's own position and positive numbers represent attitudes inconsistent with one's own position for all participants. Error bars represent ±1 standard error.

changes in their own and others' attitudes, we conducted exploratory mediational analyses. Specifically, we performed within-participants mediational analyses using the MEMORE macro (Montoya & Hayes, 2017) with measurement type (predicted vs. actual) as the independent variable and attitude change as the dependent variable, separately for changes in one's own attitude and changes in the partner's attitude, and separately for each potential mediator. The extent to which participants underestimated how much their own attitude would depolarize was mediated by underestimating how much they would agree with the other person's reasons (indirect effect: $b = -0.14$, $SE = 0.08$, 95% CI [−0.34, −0.003]; direct effect: $b = -0.24$, $SE = 0.15$, 95% CI [−0.53, 0.06]), and by underestimating their own receptiveness to the other person (indirect effect: $b = -0.23$, $SE = 0.11$, 95% CI [−0.45, −0.02]; direct effect: $b = -0.15$, $SE = 0.16$, 95% CI [−0.46, 0.16]), but was not significantly mediated by miscalibration on the other measures. The extent to which participants underestimated how much the other person's attitude would depolarize was mediated by underestimating how much the other person would report agreeing with one's own reasons (indirect effect: $b = -0.28$, $SE = 0.11$, 95% CI [−0.55, −0.10]; direct effect: $b = -0.07$, $SE = 0.15$, 95% CI [−0.38, 0.22]), but was not significantly mediated by miscalibration on the other measures (see Supplemental Materials for the nonsignificant mediators). Thus, we found consistent evidence that participants underestimated how much their own and their partner's attitudes would depolarize because they underestimated how much they would agree with each other's reasons.

Table 1

Potential Mediators in Study 2

	Predictions before the conversation		Evaluations after the conversation	
Measure	Self	Other	Self	Other
Agreement	2.48_a (1.46)	2.43_a (1.42)	4.05_b (1.66)	3.88_b (1.48)
Try to persuade	2.55_a (1.63)	3.28_b (1.17)	2.36_a (1.63)	2.35_a (1.57)
Try to defend	2.71_a (1.60)	3.99_b (1.24)	3.09_c (1.58)	3.14_c (1.56)
Try to understand	4.10_a (1.69)	3.11_b (1.40)	4.96_c (1.41)	4.54_d (1.24)
Receptiveness	4.35_a (1.49)	3.31_b (1.51)	5.42_c (0.88)	5.06_d (1.20)

Note. For the agreement measure, the "self" columns refer to the participant's agreement with the partner's reasons, and the "other" columns refer to the partner's agreement with the participant's reasons. We added three points to the agreement measures so that they are presented on the same 0–6 scale as the other measures. Numbers outside parentheses represent means. Numbers inside parentheses represent standard deviations. Within each row, means that differ significantly ($p < .050$) are indicated by different subscripts.

Follow-Up Survey

Nearly all participants (97 of 100) completed the follow-up survey. Participants' attitudes 1 week after the study session did not differ significantly from the attitudes they reported immediately after their conversations ($b = 0.22$, $SE = 0.13$), $t(97.63) = 1.60$, $p = .113$, 95% CI [−0.05, 0.48], and were therefore significantly less polarized than those they reported at baseline ($b = -0.58$, $SE = 0.11$), $t(99.27) = -5.30$, $p < .001$, 95% $CI_{difference}$ [−0.80, −0.37]. Consistent with Study 1, the more that participants' attitudes depolarized in the main session, and the more they underestimated how much their attitudes would depolarize in the main session, the more moderate their attitudes remained 1 week later compared to the attitudes they reported at baseline, suggesting the conversations brought about somewhat lasting attitude change (correlation between actual change in main session and sustained change at follow-up: ($b = 0.27$, $SE = 0.09$), $t(99.00) = 2.94$, $p = .004$, 95% CI [0.09, 0.45]; correlation between underestimation of change in main session and sustained change at follow-up: ($b = 0.22$, $SE = 0.10$), $t(99.78) = 2.28$, $p = .025$, 95% CI [0.03, 0.41]).

In Study 2, participants with opposing attitudes toward cancel culture underestimated how much their own and others' attitudes would depolarize in a spoken conversation. Importantly, we found initial evidence that this occurred because participants underestimated how much they would agree about the content of the conversation (see also Dorison et al., 2019), rather than because they misjudged social dynamics of the conversation such as how hard they would try to understand or persuade each other.

Study 3: Misunderstanding Social Divides

Study 2 found that people underestimate how much their attitudes depolarize because they underestimate how much they will agree during a conversation. Study 3 goes one layer deeper by investigating

why people with opposing attitudes underestimate their agreement. We hypothesized that people underestimate their agreement because they misunderstand why their attitudes differ: People presume their attitudes differ because they fundamentally disagree about an issue, but their attitudes actually differ because they are focused on different aspects of the issue. For example, supporters and opponents of cancel culture may presume their attitudes differ because they fundamentally disagree about which public figures should be cancelled and which should not, but their attitudes may actually differ because they are focused on different examples: Supporters bring to mind public figures whose misconduct was severe (e.g., sexual assault), whereas opponents bring to mind public figures whose misconduct was mild (e.g., expressing unpopular political views).

We tested this possibility by asking supporters and opponents of cancel culture to rate their attitudes toward, and to predict another person's attitudes toward, 16 real examples of cancel culture. If participants *presume* they fundamentally disagree about cancel culture, then they should expect to disagree nearly as much about specific examples of cancel culture as they do about cancel culture as a whole. Importantly, if supporters and opponents are *actually* focused on different examples of cancel culture as we theorize, then they should agree more about specific examples of cancel culture than they anticipate.

Method

Participants

We recruited 300 participants from the U.S. participant pool on Prolific to complete the study in exchange for $5.00. We excluded 65 participants from analyses: 17 because they failed the attention check described below, and 48 because they neither supported nor opposed cancel culture and so could not be assigned a partner on the opposite side of the issue. After data exclusions, the final sample size included 235 participants (M_{age} = 39.64, SD_{age} = 13.44; 47.66% female, 51.91% male, 0.43% other gender; 73.62% White, 8.09% Black, 4.68% Hispanic, 7.66% Asian, 5.96% other ethnicity). This sample provided about 80% power to detect a minimum Measurement Type × Target interaction effect of size $\eta_p^2 = .02$.

Procedure

After providing informed consent, participants read the definition of cancel culture from Study 2 and reported their attitude. Participants who neither supported nor opposed cancel culture completed an abridged version of the survey that omitted several measures described below; per our preregistration, these participants were excluded from analyses.

To measure what examples naturally come to mind when supporters and opponents think of cancel culture—that is, to understand how they construe this issue—we asked participants to write down a typical example that comes to mind when they think of cancel culture. Because most examples of cancel culture include both a potentially objectionable action by a public figure and a reaction by the public to try to "cancel" the figure, we asked participants to describe their example in two free-response items. In the first item, they wrote down a typical example of an action—a potentially objectionable behavior—that would cause the cancel culture movement to try to cancel a person, place, or thing. In the second item, they wrote down a typical example of a reaction that the movement would exhibit to try to cancel this person, place, or thing. Participants were required to write at least 50 characters for each item.

After submitting their responses, participants viewed their responses and indicated how much they would support or oppose this example of cancel culture they had written down (−3 = *strongly oppose*, −2 = *somewhat oppose*, −1 = *slightly oppose*, 0 = *neither support nor oppose*, 1 = *slightly support*, 2 = *somewhat support*, 3 = *strongly support*). They also rated how mild or severe the actions they had written down were (1 = *relatively mild*, 7 = *relatively severe*), and how mild or severe the reactions they had written down were (1 = *relatively mild*, 7 = *relatively severe*).

Participants were then informed that they had been matched with another participant currently taking this survey. Unlike Studies 1 and 2, however, participants were not actually matched with another person because the procedure of Study 3 did not require social interaction. Instead, participants who supported cancel culture were randomly assigned to read that they had been matched with someone who slightly opposes, somewhat opposes, or strongly opposes cancel culture. Participants who opposed cancel culture were randomly assigned to read that they had been matched with someone who slightly supports, somewhat supports, or strongly supports cancel culture.[7]

After viewing the partner's overall attitude, participants provided an open-ended response in which they explained why they think their own attitude and the other person's attitude differ. To assess whether participants underestimate how much they will agree about specific examples of cancel culture even when they are informed of the extremity of another person's overall attitude, we then asked participants to read about and evaluate 16 real examples of cancel culture involving well-known public figures such as J. K. Rowling, Colin Kaepernick, Harvey Weinstein, Joe Rogan, and The Dixie Chicks, among others (see Supplemental Materials). Each example presented both the potentially objectionable actions of the public figure and the reactions of the public to try to cancel them. For instance, the example involving Joe Rogan read:

> ACTIONS: During the COVID-19 pandemic, podcaster Joe Rogan claimed that people who are young and healthy do not need to be vaccinated against COVID-19, and promoted the use of ivermectin contrary to FDA warnings. In one episode of the podcast, Rogan interviewed Dr. Robert Malone, who had previously been suspended from Twitter for spreading misinformation about COVID-19.
>
> REACTIONS: Musicians Neil Young, Joni Mitchell, David Crosby, Stephen Stills, and Graham Nash removed their music from Spotify to protest Rogan's presence on the platform. More than a thousand doctors, scientists, and health professionals signed an open letter asking Spotify to moderate misinformation from Rogan and other podcasters. Spotify then added content advisories to episodes of any podcast that discuss COVID-19.

Participants read the examples in randomized order. After reading each example, participants reported the extent to which they supported or opposed this example of cancel culture, predicted the

[7] Although participants were not matched with a specific individual in this procedure, they were nonetheless matched with an attitude that was held by many specific individuals in our sample. As we describe below, this enables us to compare the participants' predictions about their partners' judgments against the actual judgments that were provided by other participants in our sample who held the partner's overall attitude.

extent to which their partner would support or oppose this example, and predicted how much they would agree or disagree with each other's reasons if they were to discuss this example (−3 = *strongly disagree*, −2 = *somewhat disagree*, −1 = *slightly disagree*, 0 = *equally both*, 1 = *slightly agree*, 2 = *somewhat agree*, 3 = *strongly agree*). These items allow us to test whether supporters and opponents underestimate how much they will agree with each other's attitudes toward specific examples of cancel culture, despite being informed of the extremity of another person's overall attitude toward this issue.

After evaluating all 16 examples, participants reread the open-ended response in which they had previously explained why they thought their own overall attitude and their partner's overall attitude differed. They coded their response by selecting one or more of the following options: "We were thinking of different examples of cancel culture. We might have been thinking of different actions, different reactions, or different people, places, or things that the movement has tried to cancel" versus "We disagree about specific examples of cancel culture. We might disagree because we have different values, different political beliefs, different religious beliefs, different personal backgrounds, or different interpretations of the same examples of cancel culture" versus "Neither of the above." The first two response options were presented in randomized order. We asked participants to code their open-ended responses at the end of the procedure, rather than immediately after providing these open-ended responses, to ensure that reading the response options we provided in the survey would not contaminate their predictions of how their partner would evaluate the 16 examples of cancel culture described earlier.

Participants then completed an attention check in which they indicated the other person's overall attitude toward cancel culture. Finally, participants reported demographic information were debriefed about the purpose of the research and were paid for their participation.

Results and Discussion

Participants' overall attitudes toward cancel culture were highly correlated with their attitudes toward the specific examples of cancel culture that they wrote down in the survey ($r = .83$), $t(233) = 22.79$, $p < .001$, 95% CI [0.79, 0.87]. Supporters of cancel culture wrote down examples that they supported, $t(233) = 13.02$, $p < .001$, 95% CI [1.69, 2.29], $d = 1.60$, whereas opponents of cancel culture wrote down examples that they opposed, $t(233) = -12.29$, $p < .001$, 95% CI [−2.00, −1.44], $d = -0.95$. Supporters and opponents differed 91% as much in their attitudes toward these self-generated examples as they did in their attitudes toward cancel culture as a whole, suggesting concrete thinking by itself did not meaningfully reduce polarization (Alper, 2020; Trope & Liberman, 2010).

Supporters and opponents might be polarized around their own examples either because they brought to mind comparable examples of cancel culture and disagreed about how to evaluate them—they might have different "judgments of the object"—or because they brought to mind systematically different examples of cancel culture to begin with—they might be focused on different "objects of judgment" (Asch, 1952). Consistent with thinking of different examples, supporters and opponents differed dramatically less when they evaluated the 16 researcher-generated examples of cancel culture. A 2 (attitude type: self-generated, researcher-generated) × 2 (side: supporter, opponent) analysis of variance revealed that supporters and opponents differed only 38% as much in their average evaluations of the researcher-generated examples as they did in their evaluations of their own examples, resulting in a significant attitude Type × Side interaction effect, $F(1, 233) = 161.22$, $p < .001$, $\eta_p^2 = .41$. Supporters of cancel culture were less supportive of canceling the public figures from the researcher-generated examples than they were supportive of canceling the public figures from their own examples, $F(1, 233) = 45.26$, $p < .001$, $\eta_p^2 = .37$. Likewise, opponents of cancel culture were less opposed to canceling the public figures from the researcher-generated examples than they were opposed to canceling the public figures from their own examples, $F(1, 233) = 131.30$, $p < .001$, $\eta_p^2 = .46$. Item-level analyses revealed that supporters and opponents differed significantly less for all 16 examples than they did for their self-generated examples, Fs(1, 233) > 13.19, ps < .001, η_p^2s > .04. Participants' attitudes toward cancel culture thus differed not only because they disagreed about specific examples of cancel culture, but largely because they brought to mind systematically different examples to begin with. Having supporters and opponents evaluate the same examples dissolved much of their apparent disagreement.[8]

Importantly, as seen in Figure 3, participants failed to appreciate the extent to which evaluating specific examples of cancel culture would reveal underlying areas of agreement. A 2 (measurement type: predicted, actual) × 2 (target: supporter, opponent) analysis of variance on participants' average evaluations of the researcher-generated examples revealed that differences in predicted attitudes were significantly larger than differences in actual attitudes, as indicated by a significant Measurement Type × Target interaction effect, $F(1, 466) = 51.46$, $p < .001$, $\eta_p^2 = .10$.[9] Supporters predicted that opponents would be more opposed to the researcher-generated examples than they actually were on average, $F(1, 466) = 39.56$, $p < .001$, $\eta_p^2 = .10$, whereas opponents predicted that supporters would be more supportive of these examples than they actually were on average, $F(1, 466) = 15.75$, $p < .001$, $\eta_p^2 = .12$. Item-level analyses revealed that differences in predicted attitudes differed significantly from differences in actual attitudes for 15 of the 16 examples (see Figure 4), Fs > 4.42, ps < .036, η_p^2s > .008, suggesting the participants consistently underestimated their common ground.[10]

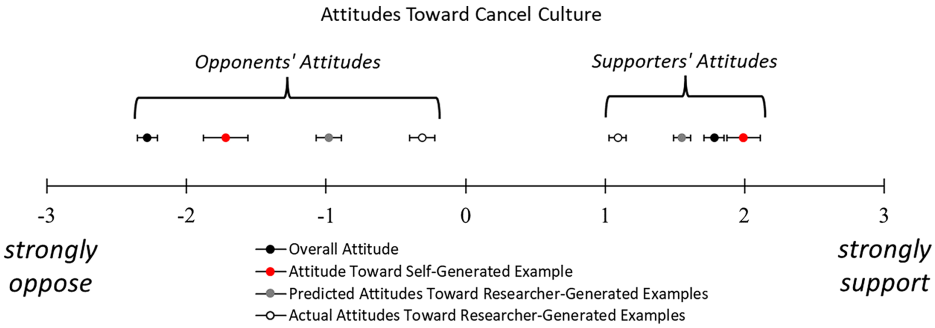
[8] We preregistered several secondary hypotheses that we report in the Supplemental Materials. All hypotheses were supported.

[9] Because the partner's overall attitude was determined by random assignment in Study 3, the number of partners with each overall attitude differed from the number of participants in our sample with the same overall attitude, $\chi^2(5, N = 235) = 19.87$, $p = .001$. We therefore compared predicted and actual attitudes toward the researcher-generated examples using weighted analysis of variance, in which participants' predictions are weighted more heavily if the randomly assigned attitude of the partner was common in our sample, and are weighted less heavily if the randomly assigned attitude of the partner was uncommon. We obtain the same results in unweighted analyses (see Supplemental Materials).

[10] For one example about Colin Kaepernick, supporters were unexpectedly less supportive of canceling Kaepernick than were opponents, $F(1, 466) = 19.54$, $p < .001$, $\eta_p^2 = .08$. This reversal produced a significant main effect of target in the opposite direction as the other examples, $F(1, 466) = 16.13$, $p < .001$, $\eta_p^2 = .03$, qualified by the hypothesized Measurement Type × Target interaction effect in the same direction as the other 14 examples described in the text, $F(1, 466) = 5.00$, $p = .026$, $\eta_p^2 = .01$. We therefore found significant support for our hypothesis in 15 of the 16 examples, and a nonsignificant difference interaction in the remaining example, $F(1, 466) = 2.08$, $p = .149$, $\eta_p^2 = .004$.

Figure 3

Mean Attitudes Toward Cancel Culture Among Supporters and Opponents in Study 3



Note. In the left side of the figure, predicted attitudes refer to supporters' predictions of opponents' attitudes toward the researcher-generated examples. In the right side of the figure, predicted attitudes refer to opponents' predictions of supporters' attitudes toward the researcher-generated examples. Predicted attitudes use weighted means as described in Footnote 9. Error bars represent ±1 standard error. See the online article for the color version of this figure.

The open-ended responses raise the possibility that participants underestimated how much they would agree because they failed to consider that their overall attitudes might be based on different examples of cancel culture. Significantly more than half the participants (63%) reported inferring that their overall attitudes differed because they "disagree about specific examples" of cancel culture, $\chi^2(1, N = 235) = 14.81$, $p < .001$. Significantly fewer than half the participants (39%) reported inferring that their overall attitudes differed because they were "thinking of different examples" of cancel culture, $\chi^2(1, N = 235) = 11.95$, $p < .001$. Thus, many participants assumed that their attitudes differed because they fundamentally disagreed about cancel culture, potentially helping to explain why they failed to appreciate that evaluating specific examples would reveal underlying areas of agreement.

Study 3 found that participants with opposing attitudes toward cancel culture underestimated how much they would agree about specific examples of this issue. This occurred at least partly because participants attributed differences in their overall attitudes to disagreements rather than to differences in the examples they brought to mind. These findings matter for our theory because they may help explain why people underestimate their agreement and underestimate attitude change in spoken conversations. Whereas people with opposing attitudes may assume they have a point-by-point disagreement about an issue, and so expect their own and others' attitudes to change relatively little, having a conversation may surface unexpected areas of agreement and so cause their attitudes to depolarize more than they expected. We continue testing this explanation with spoken conversations in Studies 4 and 5.

Study 4: Misunderstanding Social Divides in Conversation

In Study 4, we asked participants to write down an example that came to mind when they thought of cancel culture and to discuss their example in a spoken conversation with someone on the other side of this issue. We measured not only predicted and actual attitudes toward cancel culture as a whole but also predicted and actual attitudes toward the specific examples the participants wrote down and discussed.

We hypothesized that participants would underestimate how much their own and others' overall attitudes would depolarize, consistent with Studies 1 and 2. Novel to this study, the design allowed us to disentangle two explanations of the misprediction. If participants underestimate how much their overall attitudes will depolarize because they fail to appreciate that they are focused on different examples of cancel culture (as suggested by Study 3), then participants should underestimate how much they will agree with *each other's* examples during the conversation. If, however, participants underestimate how much their overall attitudes will depolarize because they overlook social dynamics of conversation that lead to compromise, then they should instead underestimate how much they will change their minds about *their own* examples.

Method

Participants

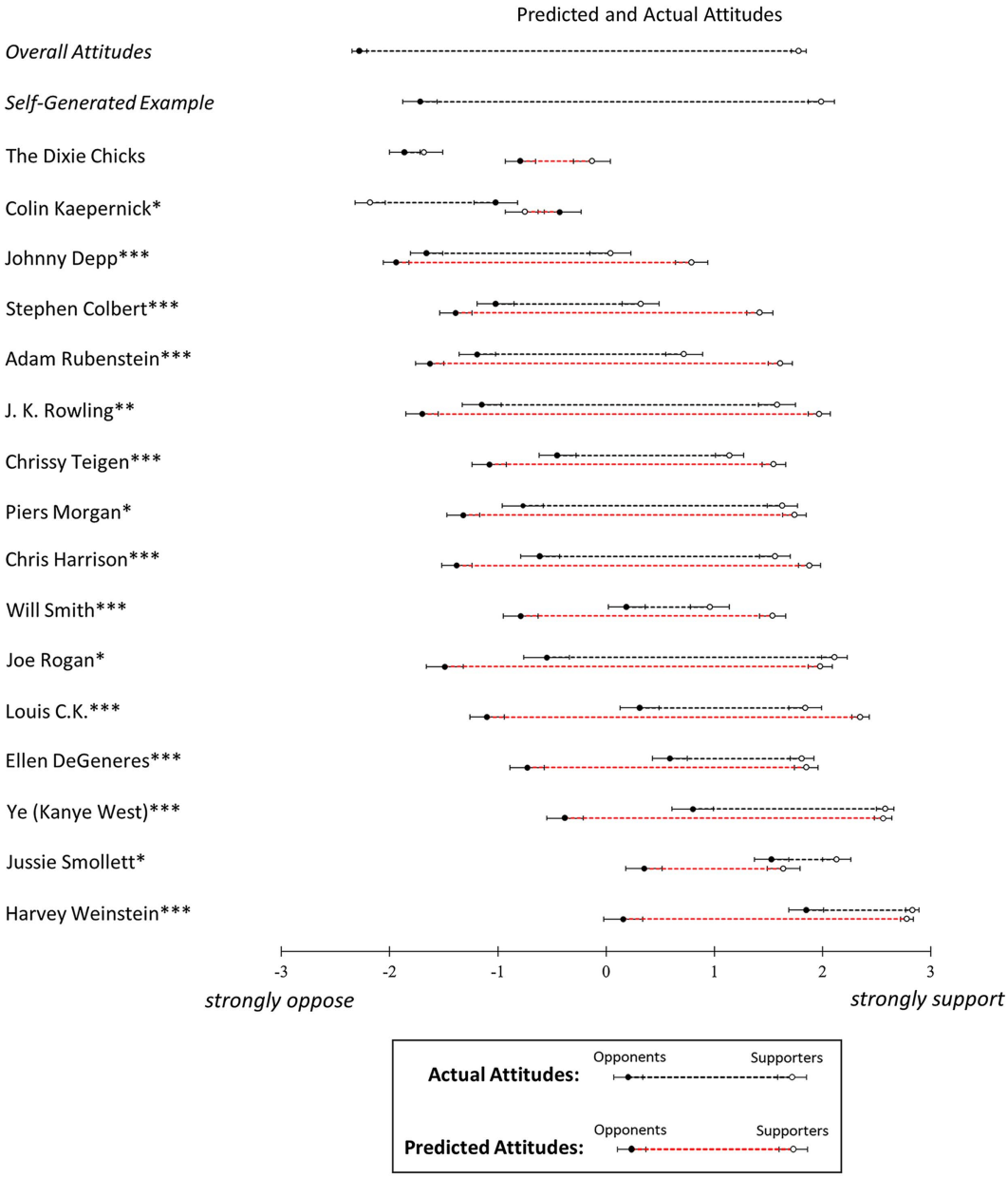
We recruited participants from the U.K. participant pool on Prolific to complete the main session in exchange for $6.00 and the follow-up survey in exchange for $1.00. One hundred eight participants completed the main session without technical difficulties. We excluded eight participants from analyses: four because they or their partner did not state their attitude during the 30-s video call, two because they or their partner did not explain their reasoning during the 10-min video call, and two because they or their partner provided nonsensical responses in the survey. This left a final sample size of 100 participants after data exclusions ($M_{age} = 39.71$, $SD_{age} = 12.25$; 38.00% female, 61.00% male, 1.00% other gender; 74.00% White, 9.00% Black, 15.00% Asian, 2.00% other ethnicity; 53.00% liberal, 26.00% conservative, 21.00% moderate), which provided about 80% power to detect differences between predicted and actual attitude change of size $b = 0.25$.

Procedure

After consenting to the study, participants read the definition of cancel culture from Studies 2 and 3, and rated their overall attitude (−3 = *strongly oppose*, −2 = *somewhat oppose*, −1 = *slightly*

Figure 4

Mean Attitudes Toward Specific Examples of Cancel Culture Among Supporters and Opponents in Study 3



Note. Actual attitudes are connected by black dotted lines. Predicted attitudes are connected by red dotted lines. Black dots represent opponents' actual attitudes and supporters' predictions of opponents' attitudes. White dots represent supporters' actual attitudes and opponents' predictions of supporters' attitudes. Predicted attitudes are represented using weighted means as described in Footnote 9. Error bars represent ±1 standard error. See the online article for the color version of this figure.
* $p < .05$. ** $p < .01$. *** $p < .001$.

oppose, 0 = *neither support nor oppose*, 1 = *slightly support*, 2 = *somewhat support*, 3 = *strongly support*, *I don't know*). Participants then completed exploratory measures in which they rated the extent to which they support cancel culture, and the extent to which they oppose cancel culture, on separate unipolar scales (0 = *not at all*, 1 = *slightly*, 2 = *somewhat*, 3 = *strongly*). Supporters reported how much they support cancel culture before they reported how much they oppose it, whereas opponents completed the items in the reverse order.

Participants were matched with a partner using the same procedure as Studies 1 and 2, such that each pair included one supporter and one opponent of cancel culture. To assess how participants

construed the topic of cancel culture, we asked them to write down an example of cancel culture that was congruent with the overall attitude they had already submitted. For example, participants who slightly supported cancel culture were asked to write down an example that they slightly supported, those who somewhat supported cancel culture were asked to write down an example that they somewhat supported, and so on.

As in Study 2, participants learned their partner's overall attitude toward cancel culture by hearing their partner state their attitude in a 30-s video call that took place several minutes before the 10-min conversation. Participants connected to the video feed and were instructed to tell each other their overall attitudes (e.g., "I strongly oppose cancel culture" and "I strongly support cancel culture") without explaining their reasons.

After this exchange, the survey reiterated the participants' overall attitudes and indicated that both participants had written down an example of cancel culture that was congruent with their overall attitude. For example, participants who "strongly opposed" and "strongly supported" cancel culture would have read: "You reported [strongly opposing] cancel culture, and you wrote down an example that you [strongly oppose]. The other participant reported [strongly supporting] cancel culture, and they wrote down an example that they [strongly support]."

Participants were then told that they and their partner would have a 10-min conversation about cancel culture. We modified the conversation prompts slightly to ensure that participants would discuss the examples of cancel culture that they had written down in the survey. Participants were told that they would be asked to discuss the following prompts:

- Why you [strongly oppose] your example of cancel culture
- Why the other participant [strongly supports] their example of cancel culture
- What you think of each other's attitudes toward these examples

Although participants were informed of each other's attitudes in the survey, they were not informed of the content of each other's examples before the conversation. Thus, participants in Study 4 possessed similar information before the conversation as did participants in Studies 1 and 2: They were aware that their own attitude and another person's attitude differed, but they were not aware of *why* their attitudes differed. As we describe below, this will allow us to test whether participants with opposing attitudes naturally underestimate how much they will agree with each other's examples during a conversation, and if so, whether this helps to explain why participants underestimate how much their own and others' overall attitudes will depolarize.

After participants read the conversation prompts, they reported a series of predictions about the conversation. They first predicted what overall attitude they would report toward cancel culture after the conversation, and what overall attitude their partner would report, on separate scales (−3 = *strongly oppose*, −2 = *somewhat oppose*, −1 = *slightly oppose*, 0 = *neither support nor oppose*, 1 = *slightly support*, 2 = *somewhat support*, 3 = *strongly support*, *I don't know*). Participants then completed several items to test two potential explanations of our hypothesized results on these measures of overall attitudes. As noted, participants might underestimate how much their overall attitudes will depolarize because they underestimate how much they will agree with each other's examples of cancel culture, or because they underestimate how much they and their partner will change their minds about their respective examples of cancel culture. To disentangle these explanations, participants were first asked to consider the example of cancel culture that they personally would share during the conversation. They predicted what attitude they would report toward this example after the conversation, and predicted what attitude their partner would report toward this example, on separate scales. Participants were then asked to consider the example of cancel culture that their partner would share during the conversation. Participants predicted what attitude their partner would report toward this example after the conversation, and predicted what attitude they personally would report toward this example, on separate scales (−3 = *strongly oppose*, −2 = *somewhat oppose*, −1 = *slightly oppose*, 0 = *neither support nor oppose*, 1 = *slightly support*, 2 = *somewhat support*, 3 = *strongly support*).

Participants then predicted how much they would disagree or agree with their partner's reasons and predicted how much their partner would disagree or agree with one's own reasons (−3 = *strongly disagree*, −2 = *somewhat disagree*, −1 = *slightly disagree*, 0 = *equally both*, 1 = *slightly agree*, 2 = *somewhat agree*, 3 = *strongly agree*). They completed the same measures of attitude strength from Study 2.

Participants then had their 10-min conversation, in which they discussed their own example and their partner's example of cancel culture. After the conversation, participants rated their overall attitude toward cancel culture and predicted their partner's overall attitude, using the same scales on which they had reported predictions (−3 = *strongly oppose*, −2 = *somewhat oppose*, −1 = *slightly oppose*, 0 = *neither support nor oppose*, 1 = *slightly support*, 2 = *somewhat support*, 3 = *strongly support*).

Participants were then asked to consider their own example of cancel culture. They rated their current attitude toward this example and predicted their partner's attitude toward this example, on separate scales. Next, participants were asked to consider their partner's example of cancel culture. Participants estimated their partner's current attitude toward this example, and they rated their own attitude toward this example,[11] on separate scales (−3 = *strongly oppose*, −2 = *somewhat oppose*, −1 = *slightly oppose*, 0 = *neither support nor oppose*, 1 = *slightly support*, 2 = *somewhat support*, 3 = *strongly support*).

Participants then indicated how much they agreed or disagreed with their partner's reasons, and estimated how much their partner

[11] After conducting the study, manual inspection of the data led us to suspect that 23 of 100 participants may have responded backward to these two items about the partner's example before the conversation, and nine of 100 participants may have responded backwards to the corresponding items after the conversation, because these participants judged their own attitude toward the partner's example and their partner's attitude toward the partner's example to differ in the opposite direction as their overall attitudes. This may have occurred because the previous survey items had asked participants to evaluate their own attitudes before their partners' attitudes, whereas these items were sequenced in the reverse order, potentially causing several participants to misread the items. We analyze the uncorrected data in the text, and analyze the corrected data in the footnotes below. Both sets of analyses produce similar results. Study 5 maintains a consistent order of items throughout the survey to reduce confusion, and replicates the results of this study.

would report having agreed or disagreed with the participant's own reasons (−3 = *strongly disagree*, −2 = *somewhat disagree*, −1 = *slightly disagree*, 0 = *equally both*, 1 = *slightly agree*, 2 = *somewhat agree*, 3 = *strongly agree*). Participants then completed exploratory items in which they rated the extent to which they support cancel culture, and the extent to which they oppose cancel culture, on separate unipolar scales (0 = *not at all*, 1 = *slightly*, 2 = *somewhat*, 3 = *strongly*). They rated their political orientation (−3 = *very liberal*, −2 = *somewhat liberal*, −1 = *slightly liberal*, 0 = *moderate*, 1 = *slightly conservative*, 2 = *somewhat conservative*, 3 = *very conservative*, *other*), indicated whether they had any issues during their conversation, reported demographic information, and were paid for their participation.

Follow-Up Survey

One week after each session, we contacted participants and asked them to complete a follow-up survey, without reminding them of the attitudes they reported in the main session. Participants first reported their overall attitude toward cancel culture, and after submitting their response, rated their attitudes toward cancel culture on the separate unipolar scales described earlier. Finally, participants were debriefed about the purpose of the research and were paid for the follow-up survey.

Results and Discussion

Predicted and Actual Attitude Change

Study 4 replicated the findings of the previous studies and found further evidence of why participants underestimate how much their own and others' overall attitudes depolarize. As seen in Figure 5, participants expected the conversations to narrow the divide between their initial attitudes by 23%, yet the conversations actually narrowed this divide by 48% on average, resulting in a significant effect of measurement type using the same mixed linear model as Studies 1 and 2 ($b = 0.46$, $SE = 0.09$), $t(350.00) = 5.02$, $p < .001$, 95% CI [0.28, 0.63]. Differences between predicted and actual depolarization did not vary for one's own attitude versus the partner's attitude or for supporters versus opponents, as indicated by nonsignificant interactions, $|t|s < 1.06$, $ps > .294$.

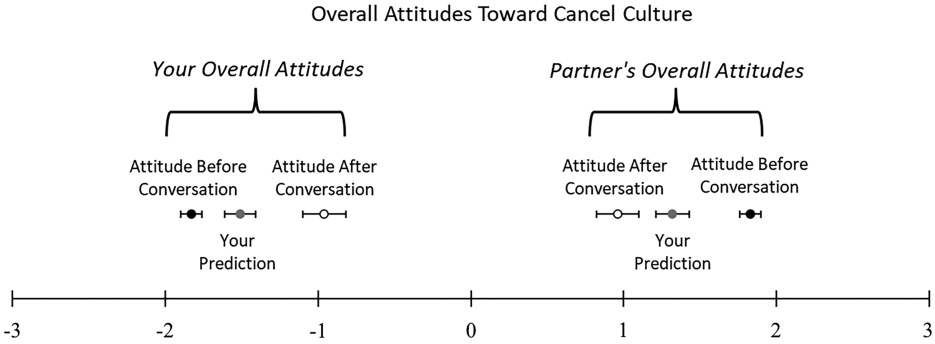
Participants not only underestimated how much their own and their partner's attitudes would depolarize, but they also underestimated how much they personally would perceive their partner's attitude to have depolarized after the conversation. Specifically, participants estimated that their partner's attitude had depolarized significantly more after the conversation than they had predicted before the conversation ($Ms_{\text{depolarization}} = 1.12$ vs. 0.51, respectively; $SDs = 1.12$ vs. 0.89, $b = 0.61$, $SE = 0.12$), $t(100.00) = 5.02$, $p < .001$, 95% CI [0.37, 0.85], providing convergent evidence that participants underestimated changes in each other's attitudes before the conversation. Participants' estimates after the conversation did not differ significantly from how much their partners' attitudes had actually depolarized ($b = -0.25$, $SE = 0.14$), $t(100.00) = -1.84$, $p = .069$, 95% CI [−0.52, 0.02], suggesting the conversations provided feedback that helped to calibrate the participants' inferences about each other's attitudes.

Predicted and Actual Attitudes Toward Each Person's Own Example

We next tested two potential explanations of why participants underestimated how much their own and others' overall attitudes would depolarize. One explanation is that participants underestimated how much they and their partner would change their minds about their respective examples of cancel culture. Such a result could occur if, for example, participants overlooked social dynamics of conversation that lead to compromise. As seen in Figure 6, we found no support for this possibility. A mixed linear model produced a significant effect of measurement type in the opposite direction as this explanation would have predicted ($b = -0.51$, $SE = 0.10$), $t(350.00) = -5.11$, $p < .001$, 95% CI [−0.70, −0.31], qualified by a significant Measurement Type × Target interaction effect ($b = -1.03$, $SE = 0.20$), $t(350.00) = -5.21$, $p < .001$, 95% CI [−1.42, −0.64]. These effects indicated that participants accurately anticipated that their attitudes toward their own examples of cancel culture would depolarize very little

Figure 5

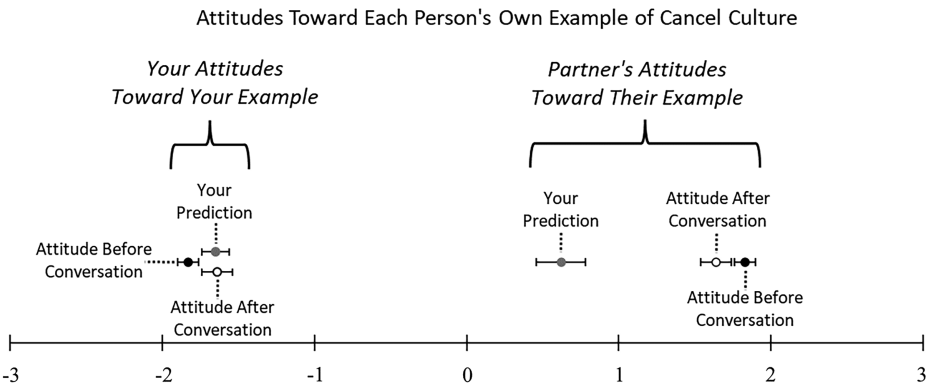
Mean Preconversation Attitudes, Predicted Attitudes, and Postconversation Attitudes for Oneself and One's Conversation Partner in Study 4



Note. We reverse coded the attitudes of participants who initially supported cancel culture, so that negative numbers represent attitudes consistent with one's own position and positive numbers represent attitudes inconsistent with one's own position for all participants. Error bars represent ±1 standard error.

Figure 6

Mean Preconversation Attitudes, Predicted Attitudes, and Postconversation Attitudes Toward Each Person's Own Example of Cancel Culture in Study 4



Note. We reverse coded the attitudes of participants who initially supported cancel culture, so that negative numbers represent attitudes consistent with one's own position and positive numbers represent attitudes inconsistent with one's own position for all participants. Dashed lines indicate which labels refer to which data points. Error bars represent ±1 standard error.

($b = 0.01$, $SE = 0.09$), $t(100.00) = 0.11$, $p = .914$, 95% CI [−0.17, 0.19], and overestimated how much their partners' attitudes toward the partners' examples of cancel culture would depolarize ($b = -1.02$, $SE = 0.17$), $t(150.00) = -5.89$, $p < .001$, 95% CI [−1.36, −0.68].[12] Thus, these measures cannot explain why participants underestimated how much their own and their partners' overall attitudes toward cancel culture would depolarize.

Predicted and Actual Attitudes Toward Each Other's Examples

The second potential explanation is that participants underestimated how much their own and others' overall attitudes would depolarize because they underestimated how much they would agree with each other's examples of cancel culture. This result would be consistent with failing to appreciate that they and their partner were focused on different examples of cancel culture, such that the conversations might have surfaced unexpected areas of agreement (see Study 3). As seen in Figure 7, this explanation was supported. A mixed linear model analogous to the ones described above produced a significant effect of measurement type in the hypothesized direction ($b = 0.76$, $SE = 0.13$), $t(350.00) = 5.69$, $p < .001$, 95% CI [0.50, 1.02], qualified by a significant Measurement Type × Target interaction effect ($b = 0.62$, $SE = 0.27$), $t(350.00) = 2.32$, $p = .021$, 95% CI [0.09, 1.15]. Participants significantly underestimated how much their own attitude toward their partner's example would lean in the direction of their partner's attitude ($b = 0.45$, $SE = 0.20$), $t(100.00) = 2.28$, $p = .025$, 95% CI [0.06, 0.84], and significantly underestimated how much their partner's attitude toward one's own example would lean in the direction of one's own attitude ($b = 1.07$, $SE = 0.19$), $t(150.00) = 5.70$, $p < .001$, 95% CI [0.70, 1.44], with somewhat larger miscalibration for the partner's attitude toward one's own example as indicated by the two-way interaction described above.[13]

These findings are consistent with the direct measures of predicted and actual agreement. Participants significantly underestimated how much they would report agreeing with their partner's reasons after the conversation ($b = 1.42$, $SE = 0.17$), $t(100.00) = 8.41$, $p < .001$, 95% CI [1.09, 1.75], and significantly underestimated how much the partner would report agreeing with one's own reasons ($b = 1.37$, $SE = 0.18$), $t(149.98) = 7.66$, $p < .001$, 95% CI [1.02, 1.72]. Thus, participants underestimated how much their own and others' overall attitudes would depolarize not because they underestimated how much they would change their minds about their respective examples of cancel culture, but rather because they underestimated how much they would agree with each other's examples of cancel culture.

Follow-Up Survey

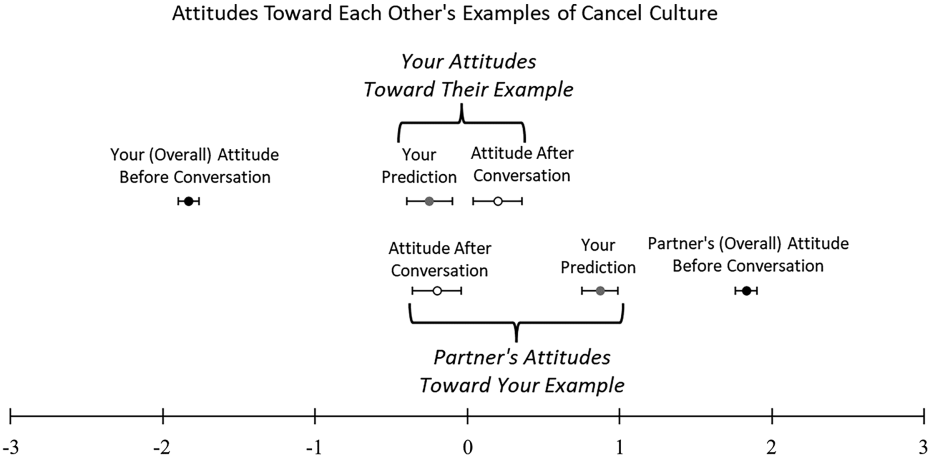
All 100 participants responded to the follow-up survey. Participants' overall attitudes in the follow-up survey were more polarized than the attitudes they reported immediately after their conversations ($b = 0.38$, $SE = 0.11$), $t(100.00) = 3.53$, $p = .001$, 95% CI [0.17, 0.59], especially among participants who initially opposed cancel culture ($b = -0.56$, $SE = 0.22$), $t(100.00) = -2.60$, $p = .011$, 95% CI [−0.99, −0.13].

[12] In analyses of the corrected data (see Footnote 11), we likewise observed a significant effect of measurement type ($b = -0.22$, $SE = 0.08$), $t(350.00) = -2.68$, $p = .008$, 95% CI [−0.38, −0.06], qualified by a significant interaction with target ($b = -0.46$, $SE = 0.16$), $t(350.00) = -2.80$, $p = .005$, 95% CI [−0.78, −0.14]. Participants accurately anticipated that their attitudes toward their own examples of cancel culture would depolarize very little, $b = 0.01$, $SE = 0.09$, $t(100.00) = 0.11$, $p = .914$, 95% CI [−0.17, 0.19], and overestimated how much their partners' attitudes toward the partners' examples of cancel culture would depolarize ($b = -0.45$, $SE = 0.13$), $t(150.00) = -3.36$, $p = .001$, 95% CI [−0.71, −0.19].

[13] In analyses of the corrected data (see Footnote 11), we observed a significant effect of measurement type ($b = 1.01$, $SE = 0.13$), $t(350.00) = 7.68$, $p < .001$, 95% CI [0.75, 1.26], and a nonsignificant interaction with target ($b = 0.05$, $SE = 0.26$), $t(350.00) = 0.19$, $p = .849$, 95% CI [−0.46, 0.56]. Participants underestimated how much their own attitude toward their partner's example would lean in the direction of their partner's attitude ($b = 0.98$, $SE = 0.18$), $t(100.00) = 5.45$, $p < .001$, 95% CI [0.62, 1.34], and underestimated how much their partner's attitude toward one's own example would lean in the direction of one's own attitude ($b = 1.03$, $SE = 0.19$), $t(150.00) = 5.32$, $p < .001$, 95% CI [0.65, 1.41].

Figure 7

Mean Predicted Attitudes and Postconversation Attitudes Toward Each Other's Examples of Cancel Culture in Study 4



Note. We reverse coded the attitudes of participants who initially supported cancel culture, so that negative numbers represent attitudes consistent with one's own position and positive numbers represent attitudes inconsistent with one's own position for all participants. Error bars represent ±1 standard error.

However, participants' overall attitudes in the follow-up survey were significantly less polarized than those they reported at baseline ($b = -0.49$, $SE = 0.10$), $t(100.00) = -5.09$, $p < .001$, 95% CI [−0.68, −0.30], with no differences in this finding between participants who initially supported and opposed cancel culture ($b = -0.30$, $SE = 0.19$), $t(100.00) = -1.56$, $p = .122$, 95% CI [−0.68, 0.08]. Importantly, the more that participants' overall attitudes depolarized in the main session, and the more they underestimated how much their overall attitudes would depolarize in the main session, the more moderate their overall attitudes remained 1 week later compared to the overall attitudes they reported at baseline, again suggesting the conversations brought about somewhat lasting attitude change (correlation between actual change in main session and sustained change at follow-up: $b = 0.41$, $SE = 0.07$, $t(100.00) = 5.47$, $p < .001$, 95% CI [0.26, 0.55]; correlation between underestimation of change in main session and sustained change at follow-up: $b = 0.28$, $SE = 0.09$, $t(100.00) = 3.25$, $p = .002$, 95% CI [0.11, 0.45]).[14]

Study 4 replicated the finding that participants underestimated how much their own and others' overall attitudes would depolarize in a spoken conversation. We found evidence that this occurred because participants underestimated how much they would agree with each other's examples of cancel culture, suggesting they did not make adequate allowance for potential differences in the examples of cancel culture they had brought to mind before the conversation.

Importantly, Study 4 did *not* support several alternative interpretations of our findings. Participants did not underestimate how much they or their partner would change their minds about their own examples of cancel culture, suggesting that our findings stem from exposure to each other's examples rather than from dynamics of conversation that dissuade participants from their attitudes toward their own examples. These results are also inconsistent with an illusion of explanatory depth (Fernbach et al., 2013), statistical regression of the participants' attitudes, or a demand characteristic, each of which predict that participants should underestimate how much they and their partner will change their minds about their respective examples of cancel culture.

Study 5: Misunderstanding Political Divides in Conversation

There are many topics on which people expect to disagree, but people do seem to agree that politics are a topic about which people's attitudes are unlikely to change (see pretest described below). Our theory, however, predicts that even people with different political views may fail to appreciate that they may be focused on different aspects of an issue, and so may depolarize more while discussing politics than they expect. Therefore, to test our hypothesis in a situation where attitude change may seem especially unlikely, Study 5 examined conversations about a polarizing political topic: the performance of U.S. President Joe Biden.

Method

Pretest

We first sought to determine whether people feel more strongly about Joe Biden's job performance, and expect less attitude change for

[14] We also analyzed the unipolar measures in which participants provided separate ratings of how much they supported and opposed cancel culture. Participants' attitudes toward their original side of the issue in the follow-up survey ($M = 1.77$, $SD = 0.68$) did not differ significantly from the attitudes they reported immediately after their conversations ($M = 1.69$, $SD = 0.73$, $b = 0.08$, $SE = 0.06$), $t(100.00) = 1.24$, $p = .216$, 95% CI [−0.05, 0.21], and were therefore significantly less polarized than those they reported at baseline ($M = 1.94$, $SD = 0.69$, $b = -0.17$, $SE = 0.07$), $t(100.00) = -2.28$, $p = .025$, 95% CI [−0.32, −0.02]. In contrast, participants' attitudes toward their partner's original side of the issue 1 week after the session ($M = 0.78$, $SD = 0.64$) fell between the attitudes they reported immediately after their conversations ($M = 0.88$, $SD = 0.67$) and those they reported at baseline ($M = 0.67$, $SD = 0.70$), and did not differ significantly from either, $|ts| < 1.81$, $ps > .075$.

this topic, compared to the other topics we have studied. Therefore, we conducted a pretest in which participants provided a series of evaluations of cats versus dogs, cancel culture, and Biden's job performance as U.S. president (see Supplemental Materials for details). Participants predicted that people with opposing attitudes toward Biden's job performance would be less likely to depolarize by the end of a 10-min conversation than people with opposing attitudes toward cats versus dogs or cancel culture (ts = 4.85 and 4.27, respectively; ps < .001). Participants rated their attitudes toward Biden's job performance to be similarly clear (ts = 1.22 and −1.30, ps = .224 and .194), somewhat more correct (ts = −4.08 and −1.57, ps = .0001 and .118), and significantly more important (ts = −21.79 and −9.48, ps < .001) than their attitudes toward the other topics. The results of this pretest suggest participants felt at least as strongly about Biden's job performance, and thought that people's attitudes should be less likely to depolarize, compared to the other topics we have studied.

Participants

We recruited participants from the U.S. participant pool on Prolific to complete the main session in exchange for $6.00 and the follow-up survey in exchange for $1.00.[15] Two hundred ten participants completed the main session without technical difficulties. We excluded 10 participants from analyses: eight because they or their partner did not state their attitudes during the 30-s video call, and two because they or their partner did not explain their reasons for their attitudes during the 10-min video call. This left a final sample size of 200 participants after data exclusions (M_{age} = 44.76, SD_{age} = 14.10; 44.00% female, 55.00% male, 1.00% other gender; 76.00% White, 10.00% Black, 2.00% Hispanic, 4.50% Asian, 7.50% other ethnicity; 54.50% liberal, 31.50% conservative, 14.00% moderate), which provided about 80% power to detect differences between predicted and actual attitude change of size b = 0.16.

Procedure

The study sessions followed a similar procedure as Study 4: Participants were matched with someone on the other side of the issue, reported predictions, had a 10-min, spoken conversation through a video call, completed survey items analogous to those they had completed before the conversation, and were contacted 1 week later for a follow-up survey.

However, we made several changes to the procedure compared to Study 4. First, participants reported their political orientation (−3 = *very liberal*, −2 = *somewhat liberal*, −1 = *slightly liberal*, 0 = *moderate*, 1 = *slightly conservative*, 2 = *somewhat conservative*, 3 = *very conservative*), and the political party they most identify with (*democratic party* vs. *republican party* vs. *independent* vs. *other*), during a prescreen survey that took place several days before the main session, rather than at the end of the main session itself. We made this change so that the participants' responses to these items could not be affected by their conversations during the main session. The participants' responses to these items, however, were not used to determine their eligibility for the main session.

Second, because participants in Study 5 discussed their attitudes toward Joe Biden's job performance, we modified the baseline measures that participants completed at the start of the main session. Participants first rated their overall attitude toward Biden's job performance (−3 = *strongly disapprove*, −2 = *somewhat disapprove*, −1 = *slightly disapprove*, 0 = *neither approve nor disapprove*, 1 = *slightly approve*, 2 = *somewhat approve*, 3 = *strongly approve*, *I don't know*). They indicated how interested they would be in voting for Joe Biden, if the 2024 presidential election were being held today (0 = *not at all interested*, 3 = *somewhat interested*, 6 = *very interested*). They then completed two feeling thermometer measures in which they indicated how unfavorable or favorable they feel toward people who approve of Joe Biden's job performance, and toward people who disapprove of Joe Biden's job performance (0 = *unfavorable*, 10 = *favorable*), sequenced so that participants first evaluated people on their own side of the issue and then evaluated people on the other side of the issue. These baseline measures were also included among participants' predictions before the conversation, in their actual evaluations after the conversation, and in the follow-up survey 1 week later.

Third, rather than ask participants to write down one example before the conversation, we asked them to write down between one and three areas of Joe Biden's job performance that were congruent with the overall attitude they had already submitted. For example, participants who strongly approved of Biden's job performance were asked to write down between one and three areas that they strongly approved of, those who strongly disapproved of Biden's job performance were asked to write down between one and three areas that they strongly disapproved of, and so on. The first area was required whereas the second and third were optional. Participants discussed these areas of Biden's job performance during their conversations.

Finally, after reporting predictions about the conversation, participants also rated how much they thought the difference between their own attitude and the other participant's attitude represents an objective disagreement or a subjective difference of opinion (−3 = *completely an objective disagreement*, 0 = *equally both*, 3 = *completely a subjective difference of opinion*), how important they consider the topic of Joe Biden's job performance to be (0 = *not important at all*, 3 = *somewhat important*, 6 = *very important*), and how closely they have followed Joe Biden's job performance during his presidency (0 = *not at all closely*, 3 = *somewhat closely*, 6 = *very closely*). Participants completed the measure of objective disagreement versus subjective difference of opinion again at the end of the postconversation survey. We did not include measures of attitude certainty due to the length of the study.

Results and Discussion

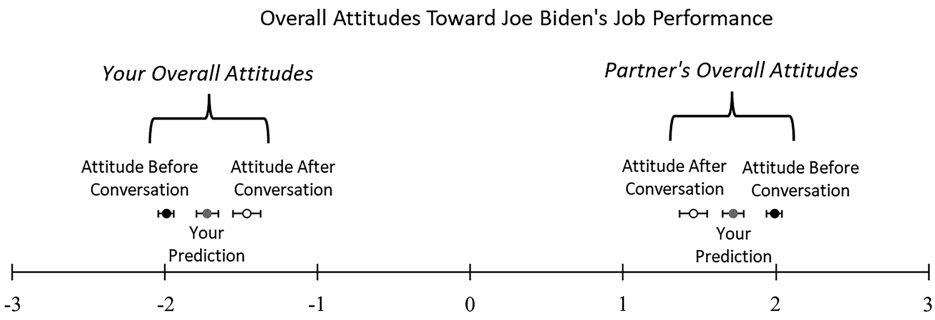
Predicted and Actual Attitude Change

Replicating the previous studies, participants underestimated how much their own and others' overall attitudes toward Joe Biden's job performance would depolarize during the 10-min conversation (see Figure 8). Whereas participants expected the conversations to narrow the divide between their initial attitudes by 14%, the conversations

[15] We conducted the study sessions between April 9 and April 26, 2023, during the third year of Joe Biden's presidency. Although Biden announced his intention to run for reelection on April 25, 2023, close to the end of data collection, it was widely assumed before this announcement that he was likely to run for reelection. National polls indicated that between 41% and 43% of Americans approved, and between 52% and 54% disapproved, of Biden's job performance at the time of these study sessions (FiveThirtyEight, 2023).

Figure 8

Mean Preconversation Attitudes, Predicted Attitudes, and Postconversation Attitudes for Oneself and One's Conversation Partner in Study 5



Note. We reverse coded the attitudes of participants who initially approved of Biden, so that negative numbers represent attitudes consistent with one's own position and positive numbers represent attitudes inconsistent with one's own position for all participants. Error bars represent ±1 standard error.

actually narrowed this divide by 27% on average, resulting in a significant effect of measurement type using the same mixed linear model as the previous studies ($b = 0.27$, $SE = 0.06$), $t(700.00) = 4.51$, $p < .001$, 95% CI [0.15, 0.38]. Thus, although participants' overall attitudes depolarized somewhat less in Study 5 than in the previous studies, participants also predicted less depolarization, and this maintained the miscalibration between predicted and actual depolarization that we have observed throughout our studies. Differences between predicted and actual depolarization did not vary for one's own attitude or the partner's attitude, as indicated by a nonsignificant Measurement Type × Target interaction ($b = 0.00$, $SE = 0.12$), $t(700.00) = 0.00$, $p = 1.000$, 95% CI [−0.23, 0.23], and did not vary for participants who approved or disapproved of Biden, as indicated by a nonsignificant Measurement Type × Side interaction ($b = 0.13$, $SE = 0.12$), $t(700.00) = 1.11$, $p = .269$, 95% CI [−0.10, 0.36].

These nonsignificant two-way interactions, however, were qualified by a significant Measurement Type × Target × Side interaction effect that we did not observe in the previous studies ($b = 0.68$, $SE = 0.24$), $t(700.00) = 2.89$, $p = .004$, 95% CI [0.22, 1.14]. This three-way interaction indicated that whereas participants who approved and disapproved of Biden were similarly likely to underestimate how much their own attitudes would depolarize ($b = -0.21$, $SE = 0.13$), $t(200.00) = -1.57$, $p = .118$, 95% CI [−0.47, 0.05], participants who approved were significantly more likely than those who disapproved to underestimate how much their partner's overall attitude would depolarize ($b = 0.47$, $SE = 0.18$), $t(400.00) = 2.66$, $p = .008$, 95% CI [0.12, 0.82].

Participants not only underestimated how much their own and their partner's attitudes would depolarize, but they also underestimated how much they personally would perceive their partner's attitude to have depolarized after the conversation. Specifically, participants estimated that their partner's attitude had depolarized significantly more after the conversation than they had predicted before the conversation ($Ms_{\text{depolarization}} = 0.58$ vs. 0.27, respectively; SDs = 0.89 vs. 0.69, $b = 0.32$, $SE = 0.07$), $t(200.00) = 4.54$, $p < .001$, 95% CI [0.18, 0.45], providing convergent evidence that participants underestimated changes in each other's attitudes before the conversation. This was especially true of participants who approved of Biden compared to those who disapproved ($b = 0.37$, $SE = 0.14$), $t(200.00) = 2.67$, $p = .008$, 95% CI [0.10, 0.64], consistent with the interaction effect described earlier. Participants' estimates of how much their partners' attitudes had depolarized after the conversation did not differ significantly from how much their partners' attitudes had actually depolarized ($b = -0.05$, $SE = 0.08$), $t(200.00) = -0.62$, $p = .537$, 95% CI [−0.21, 0.11], suggesting the conversations provided feedback that helped to calibrate the participants' inferences about each other's attitudes.

Predicted and Actual Attitudes Toward Each Person's Own Examples

As in Study 4, we tested two potential explanations of why participants underestimated how much their own and others' overall attitudes would depolarize. One explanation is that participants underestimated how much they and their partner would change their minds about their respective examples of Biden's job performance. As in Study 4, we did not find consistent support for this explanation. A mixed linear model analogous to the one described above found a nonsignificant effect of predicted versus actual depolarization ($b = 0.03$, $SE = 0.06$), $t(700.00) = 0.55$, $p = .583$, 95% CI [−0.08, 0.15], qualified by a significant interaction with self versus other ($b = -0.25$, $SE = 0.12$), $t(700.00) = -2.07$, $p = .039$, 95% CI [−0.48, −0.01]. As seen in Figure 9, participants slightly but significantly underestimated how much their attitudes toward their own examples would depolarize ($b = 0.16$, $SE = 0.07$), $t(200.00) = 2.21$, $p = .028$, 95% CI [0.02, 0.29], but nonsignificantly overestimated how much their partners' attitudes toward the partners' examples would depolarize ($b = -0.09$, $SE = 0.09$), $t(400.00) = -0.95$, $p = .341$, 95% CI [−0.28, 0.10].

Predicted and Actual Attitudes Toward Each Other's Examples

The second explanation is that participants underestimated how much their overall attitudes would depolarize because they underestimated how much they would agree about each other's examples of Biden's job performance. We found clear support for this explanation, conceptually replicating the findings of Study 4.

Figure 9

Mean Preconversation Attitudes, Predicted Attitudes, and Postconversation Attitudes Toward Each Person's Own Examples of Biden's Job Performance in Study 5

Note. We reverse coded the attitudes of participants who initially approved of Biden, so that negative numbers represent attitudes consistent with one's own position and positive numbers represent attitudes inconsistent with one's own position for all participants. Dashed lines indicate which labels refer to which data points. Error bars represent ±1 standard error.

A mixed linear model analogous to the ones described above produced only a significant effect of measurement type in the hypothesized direction ($b = 0.56$, $SE = 0.08$), $t(700.00) = 7.09$, $p < .001$, 95% CI [0.40, 0.71]. As seen in Figure 10, participants significantly underestimated how much their own attitude toward their partner's examples would lean in the direction of their partner's attitude ($b = 0.49$, $SE = 0.11$), $t(200.00) = 4.65$, $p < .001$, 95% CI [0.28, 0.70], and significantly underestimated how much their partner's attitude toward one's own examples would lean in the direction of one's own attitude ($b = 0.63$, $SE = 0.11$), $t(300.00) = 5.52$, $p < .001$, 95% CI [0.40, 0.85].

These findings are consistent with the direct measures of predicted and actual agreement. Participants significantly underestimated how much they would report agreeing with their partner's reasons after the conversation ($b = 1.06$, $SE = 0.12$), $t(200.00) = 9.03$, $p < .001$, 95% CI [0.82, 1.29], and significantly underestimated how much their partner would report agreeing with one's own reasons after the conversation ($b = 0.95$, $SE = 0.14$), $t(300.00) = 6.62$, $p < .001$, 95% CI [0.67, 1.23]. Thus, participants underestimated how much their own and others' overall attitudes would depolarize not because they underestimated how much they would change their minds about their respective examples of Biden's job performance, but rather because they underestimated how much they would agree with each other's examples of his job performance.

Consequences

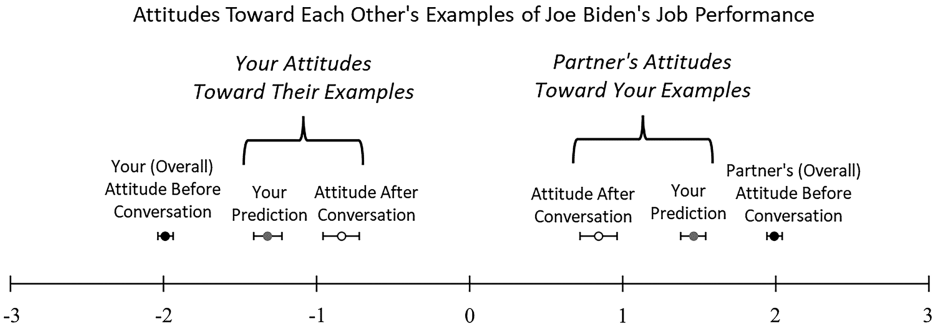
Importantly, Study 5 also finds evidence of two consequences of underestimating attitude change. Participants slightly but significantly underestimated how much their own interest in voting for Biden would depolarize during the conversation ($M_{\text{predicted-change}} = 0.02$, $M_{\text{actual-change}} = 0.17$, SDs = 0.60 and 0.83, respectively, $b = -0.15$, $SE = 0.05$), $t(200.00) = -2.92$, $p = .004$, 95% CI [−0.24, −0.05], and significantly underestimated how positive they would feel toward people on the other side of the issue after the conversation ($M_{\text{predicted-change}} = 0.47$, $M_{\text{actual-change}} = 0.77$, SDs = 1.41 and 1.72, respectively, $b = 0.30$, $SE = 0.11$), $t(200.00) = 2.78$, $p = .006$, 95% CI [0.09, 0.51]. Exploratory analyses suggested that participants underestimated changes on the voting and feeling thermometer measures because they underestimated how much their own attitudes would depolarize during the conversation, rather than because they underestimated how much their partners' attitudes would depolarize. Specifically, underestimations on the voting and feeling thermometer measures were significantly larger among participants who underestimated how much their own attitudes would depolarize than among those who did not, $|b$s$| > 0.23$, $|t$s$| > 3.14$, ps $< .003$, but were similar in magnitude among participants who underestimated how much their partners' attitudes would depolarize and among those who did not, $|b$s$| < 0.04$, $|t$s$| < 0.76$, ps $> .452$.

Follow-Up Survey

Nearly all participants (197 of 200) responded to the follow-up survey. Consistent with the previous studies, the conversations had a somewhat lasting impact on the participants' attitudes. Participants' overall attitudes toward Biden's job performance 1 week after their session were more polarized than the attitudes they reported immediately after their conversations ($b = 0.35$, $SE = 0.07$), $t(198.68) = 4.75$, $p < .001$, 95% CI [0.20, 0.49], but were significantly less polarized than those they reported at baseline ($b = -0.19$, $SE = 0.06$), $t(199.25) = -3.04$, $p = .003$, 95% CI [−0.32, −0.07]. Similarly, participants' evaluations of people on the other side of the issue 1 week after the session were not as positive as the evaluations they reported immediately after their conversations ($b = -0.48$, $SE = 0.11$), $t(197.78) = -4.29$, $p < .001$, 95% CI [−0.71, −0.26], but were significantly more positive than their original evaluations ($b = 0.28$, $SE = 0.14$), $t(197.78) = 2.10$, $p = .037$, 95% CI [0.02, 0.55]. The more that participants' overall attitudes depolarized in the main session, and the more they underestimated how much their overall attitudes would depolarize in the main session, the more moderate their overall attitudes remained

Figure 10

Mean Predicted Attitudes and Postconversation Attitudes Toward Each Other's Examples of Biden's Job Performance in Study 5



Note. We reverse coded the attitudes of participants who initially approved of Biden, so that negative numbers represent attitudes consistent with one's own position and positive numbers represent attitudes inconsistent with one's own position for all participants. Error bars represent ±1 standard error.

1 week later compared to the attitudes they reported at baseline, suggesting the conversations brought about somewhat lasting changes in these attitudes (correlation between actual change in main session and sustained change at follow-up: $|ts| > 2.51$, $ps < .014$; correlation between underestimation of change in main session and sustained change at follow-up: $|ts| > 2.21$, $ps < .028$).

In contrast, changes in the participants' interest in voting for Biden were less enduring. Participants' interest in voting for Biden was significantly more polarized 1 week after the session than it was immediately after the conversation ($b = 0.13$, $SE = 0.06$), $t(196.70) = 2.01$, $p = .046$, 95% CI [0.002, 0.26], and did not differ significantly from the interest they had reported at baseline ($b = -0.04$, $SE = 0.06$), $t(196.90) = -0.63$, $p = .533$, 95% CI [−0.16, 0.08].

In Study 5, participants with opposing attitudes toward Joe Biden's job performance as U.S. president underestimated how much their attitudes would depolarize in a spoken conversation, consistent with prior studies. This result occurred primarily because participants underestimated how much they would agree with each other's examples of Biden's job performance, suggesting that they did not anticipate the extent to which the conversation would draw their attention to different facets of Biden's presidency that they had not considered before the conversation. As a result, participants underestimated how much the conversation would depolarize their interest in voting for Biden, and underestimated how much the conversation would enhance their evaluations of people on the other side of this issue. It may be particularly noteworthy that participants underestimated how positive they would feel toward people on the other side of the issue, as this suggests that the outcomes of a short conversation can generalize to a broader group besides the conversation partner (cf. Santoro & Broockman, 2022).

Study 6: Manipulating Expectations of Attitude Change

Our theory suggests people lack interest in discussing their differences of opinion in part because they have miscalibrated beliefs that their own and others' attitudes are unlikely to change. If so, this suggests one intervention to enhance people's interest in civil discourse: Inform them that their own and others' attitudes may depolarize more than they expect. Study 6 tested this possibility by experimentally manipulating participants' expectations of attitude change, and measuring their interest in discussing Joe Biden's job performance with someone on the other side of this issue. Although participants could be hesitant to have a conversation in which they expect their own attitude to depolarize, they might also be eager to have a conversation in which they expect another person's attitude to depolarize, and so we hypothesized that our intervention would increase their overall interest in having the conversation.

Method

Participants

We recruited 461 participants from the U.S. participant pool on Prolific to complete the study in exchange for $0.80. We excluded 61 participants from analyses because they failed the attention check described below. This left a final sample size of 400 participants after data exclusions ($M_{age} = 40.57$, $SD_{age} = 13.18$; 43.75% female, 53.50% male, 2.75% other gender; 67.75% White, 9.50% Black, 5.25% Hispanic, 10.25% Asian, 7.25% other ethnicity; 54.00% liberal, 25.75% conservative, 20.25% moderate), which provided about 80% power to detect a minimum effect of size $d = 0.40$ in which the participants' interest in having the conversation differs between the underestimation-of-self-and-other condition and the control condition.

Procedure

After participants provided informed consent, they indicated how much they currently approve or disapprove of Joe Biden's job performance as U.S. president (−3 = *strongly disapprove*, −2 = *somewhat disapprove*, −1 = *slightly disapprove*, 0 = *neither approve nor disapprove*, 1 = *slightly approve*, 2 = *somewhat approve*, 3 = *strongly approve*, *I don't know*). Participants were then informed that they had been matched with another participant currently taking this survey. Unlike Study 5, however, participants were not actually matched with another person, because the procedure of Study 6 did not require social interaction. Instead, participants who approved of Joe Biden's job performance were randomly assigned to read that they had been

matched with someone who slightly disapproves, somewhat disapproves, or strongly disapproves of his job performance. Participants who disapproved of Joe Biden's job performance were randomly assigned to read that they had been matched with someone who slightly approves, somewhat approves, or strongly approves of his job performance.

Participants were asked to imagine that they were about to have a 10-min, spoken conversation with this other participant about Joe Biden's job performance. During the conversation, they and the other participant would explain their attitudes toward Joe Biden's job performance and would respond to each other's attitudes. Participants were then randomly assigned to one of four conditions, each of which accurately described different aspects of the results of our research. Participants in the control condition read:

> You might think that your own attitude and the other participant's attitude toward Joe Biden's job performance will change very little during a 10-minute conversation. Indeed, we have conducted research in which we have asked participants to predict how much their own attitude and another participant's attitude will change during a 10-minute conversation. Participants typically predict that their own attitude and another participant's attitude will change very little, and that their attitudes will be nearly as far apart after a 10-minute conversation as they had been before the conversation.

Participants in the underestimation-of-self-and-other condition read the same passage as the control condition. In addition, they then read the following passage:

> However, when participants in our research actually have a 10-minute conversation, they often find that their predictions were inaccurate. The participant's own attitude and the other participant's attitude often shift closer together during the conversation than the participant had predicted. That is, participants often find that they underestimated how much their own attitude and the other participant's attitude would change during the conversation. Therefore, although neither you nor the other participant is likely to experience a complete reversal during your conversation, your attitude toward Joe Biden's job performance and the other participant's attitude toward Joe Biden's job performance might shift closer together during a 10-minute conversation than you expect right now.

Although we hypothesized that participants in this underestimation-of-self-and-other condition would be more interested in having the conversation than those in the control condition, this could occur for either of two reasons. Participants in the underestimation-of-self-and-other condition might be more interested because they want to have a conversation that will cause *their own* attitude to depolarize, or because they want to have a conversation that will cause *the other person's* attitude to depolarize. To disentangle these explanations, we included two additional conditions in the experiment. Participants in the underestimation-of-self condition received nearly the same information as did participants in the underestimation-of-self-and-other condition quoted above, except that both passages focused solely on changes in one's own attitude without describing changes in the other participant's attitude. For example, participants in the underestimation-of-self condition read that "You might think that your own attitude toward Joe Biden's job performance will change very little," but that "your attitude toward Joe Biden's job performance might shift closer to the other participant's attitude during a 10-min conversation than you expect right now."

Likewise, participants in the underestimation-of-other condition received similar information except that the instructions focused solely on changes in the other participant's attitude without describing changes in one's own attitude. For example, participants in the underestimation-of-other condition read that "You might think that the other participant's attitude toward Joe Biden's job performance will change very little," but that "the other participant's attitude toward Joe Biden's job performance might shift closer to your attitude during a 10-minute conversation than you expect right now."

After receiving these instructions, participants in all conditions indicated how interested they would be in discussing Joe Biden's job performance with this other participant (0 = *not at all interested*, 3 = *somewhat interested*, 6 = *very interested*). After submitting their responses, participants completed a manipulation check in which they predicted what attitude they would report and what attitude the other participant would report after the conversation (−3 = *strongly disapprove*, −2 = *somewhat disapprove*, −1 = *slightly disapprove*, 0 = *neither approve nor disapprove*, 1 = *slightly approve*, 2 = *somewhat approve*, 3 = *strongly approve*).

Participants then completed an attention check in which they reported what we had told them about the results of our research. Because the underestimation-of-self-and-other manipulation included all the information from the other three manipulations combined, we tailored the response options separately in each condition to reduce confusion (see Supplemental Materials for the response options). Finally, participants reported their political orientation (−3 = *very liberal*, −2 = *somewhat liberal*, −1 = *slightly liberal*, 0 = *moderate*, 1 = *slightly conservative*, 2 = *somewhat conservative*, 3 = *very conservative*), demographic information, were debriefed about the purpose of the research, and were paid for their participation.

Results and Discussion

The manipulation was effective: Participants predicted that their own attitude would depolarize more in the underestimation-of-self and underestimation-of-self-and-other conditions than in the control condition, $|ts| > 3.91$, $ps < .001$. They likewise predicted that the other person's attitude would depolarize more in the underestimation-of-other and underestimation-of-self-and-other conditions than in the control condition, $|ts| > 3.57$, $ps < .001$.

Consistent with our hypothesis, participants were significantly more interested in having the conversation when they were informed that people underestimate how much both sides will depolarize ($M = 2.26$, $SD = 1.89$) than when they were only informed of people's predictions ($M = 1.66$, $SD = 1.87$), $t(392) = -2.26$, $p = .025$, 95% $CI_{difference} = [-1.17, -0.08]$, $d = -0.31$. This suggests that calibrating people's expectations of attitude change could enhance their interest in discussing their differences of opinion.

This result could have arisen either because participants were interested in having a conversation in which *their own* attitude would depolarize, or because they were interested in having a conversation in which *the other person's* attitude would depolarize. Consistent with the second explanation, participants who were informed that people underestimate how much others' attitudes will depolarize ($M = 2.31$, $SD = 1.99$) were significantly more interested in having the conversation than those who were only informed of people's predictions, $t(392) = -2.39$, $p = .018$, 95% $CI_{difference} = [-1.18, -0.11]$, $d = -0.34$.

In contrast, the first explanation—that participants are interested in having conversations in which their own attitudes will depolarize—received ambiguous support in our data. This is because participants' interest in having the conversation in the underestimation-of-self condition ($M = 2.04$, $SD = 1.86$) fell between the control condition ($M = 1.66$, $SD = 1.87$) and the underestimation-of-self-and-other condition ($M = 2.26$, $SD = 1.89$), and so did not differ significantly from any of the other conditions, $|ts| < 1.35$, $ps > .181$. To better understand how interested people are in having conversations in which their own attitudes and others' attitudes might depolarize, we pooled the data across the four conditions and performed exploratory regression analyses using participants' interest in having the conversation as the dependent variable, and their predictions of how much their own attitude and their partner's attitude will depolarize (derived from the manipulation check) as simultaneous independent variables. Participants' interest in having the conversation was significantly associated with how much they expected their own attitude to depolarize ($b = 0.41$, $SE = 0.11$), $t(397) = 3.58$, $p < .001$, 95% CI [0.18, 0.63], and with how much they expected their partner's attitude to depolarize ($b = 0.39$, $SE = 0.09$), $t(397) = 4.26$, $p < .001$, 95% CI [0.21, 0.57], suggesting that participants might be more interested in discussing their differences of opinion when they expect either person's attitude to depolarize than when they do not.[16]

To assess participants' preferences more directly, we additionally conducted Supplemental Study S2. In this supplemental study, participants ($N = 150$) again imagined discussing Joe Biden's job performance with someone on the other side of this issue (see Supplemental Materials). Participants indicated how interested they would be in having the conversation, separately for each of six possible outcomes that were presented within participants: if their own attitude would change very little versus would shift closer to the other person's attitude, if the other person's attitude would change very little versus would shift closer to one's own attitude, and if both people's attitudes would change very little versus would shift closer together. We counterbalanced the order in which participants evaluated changes in their own attitude and changes in the other person's attitude.

Consistent with our hypotheses and with Study 6, participants were more interested in having a conversation in which both people's attitudes would depolarize than a conversation in which neither person's attitude would depolarize, $F(1, 311.69) = 162.35$, $p < .001$, $\eta_p^2 = .53$. Importantly, this occurred primarily because participants were more interested in having a conversation in which their partner's attitude would depolarize than one in which their partner's attitude would not, $F(1, 311.69) = 191.95$, $p < .001$, $\eta_p^2 = .52$. Although participants were also significantly more interested in having a conversation in which their own attitude would depolarize than one in which their own attitude would not, $F(1, 311.69) = 40.75$, $p < .001$, $\eta_p^2 = .25$, manipulating participants' beliefs about how much their partner's attitude would depolarize affected their interest significantly more than manipulating their beliefs about how much their own attitude would depolarize, $F(1, 148) = 36.90$, $p < .001$, $\eta_p^2 = .20$.

General Discussion

People with different personal and political beliefs can learn from one another by discussing their differences of opinion. Yet people routinely avoid discussing their views with others who do not share their politics, their religion, or their personal convictions, creating "silos" in which people's views are reinforced—and often polarized—by like-minded others (Sunstein, 2002). Eight studies reveal one reason people may lack interest in discussing their differences of opinion: They have miscalibrated beliefs that their own and others' attitudes are unlikely to change. Participants with opposing attitudes toward cats and dogs (Study 1 and Supplemental Study S1), cancel culture (Studies 2 and 4), and Joe Biden's job performance as U.S. president (Study 5) underestimated how much their attitudes would depolarize in a spoken conversation. Participants attributed differences in their attitudes primarily to disagreements that a conversation seemed unlikely to resolve and overlooked differences in how they were construing an issue that a conversation could bridge rapidly (Studies 3–5).

We believe these findings hold importance for three reasons. First, changes in the participants' attitudes persisted over time. In each study, participants' attitudes remained somewhat less polarized 1 week after a 10-min conversation than they were at baseline. Second, changes in the participants' attitudes were consequential. Participants in Study 5 underestimated how much their interest in voting for the current president would depolarize and underestimated how positive they would feel toward people on the other side of the issue after their conversations. Third, miscalibrated expectations may leave people unnecessarily reluctant to reach out across personal and political divides. Participants in Study 6 and Supplemental Study S2 who learned that their own and others' attitudes might depolarize were more interested in discussing their differences of opinion than those who did not learn this information, meaning that calibrating people's expectations could remove a psychological barrier to civil discourse. Thus, our studies complement recent research suggesting that people also avoid discussing their political differences because they expect these conversations to be less positive and more hostile than they actually are (Wald et al., 2024).

Notably, our findings did not support the predictions of scholars of naive realism, who have hypothesized that people with opposing attitudes should *overestimate* how much they will persuade each other in conversations (Bland et al., 2012, pp. 270–271; Pronin, Puccio, & Ross, 2002, p. 648). Participants in our studies did not expect others' attitudes to depolarize more than their own attitudes and underestimated how much both sides would depolarize to a similar degree. That said, the predictions of naive realism might receive support in contexts where people attribute differences in their attitudes primarily to biases on the part of others, such as when people imagine interacting with abstract or hypothetical others who might seem to have weaker mental capacities than oneself (Epley & Kardas, 2021), when people explicitly try to persuade one another in a debate, a negotiation, or with a persuasive message (Moore & Cain, 2007; Swift & Moore, 2012), or when they discuss a topic with an objectively correct answer such as a math problem (see Table 2; Minson & Dorison, 2022).

Our research raises additional questions about when conversations will reduce attitude polarization more than people expect and when they will not. The spoken conversations between strangers that we examined here are similar in many respects to those people might have with strangers in a taxi, by the water cooler, or at a

[16] The results of this regression analysis hold when controlling for expected changes in the valence of the participants' attitudes (see Supplemental Materials).

Table 2
Assessment of Limitations

Dimension	Assessment
Internal validity	
Is the phenomenon diagnosed with experimental methods?	Studies with spoken conversations include within-participants (1, S1, 2, 4, and 5) and between-participants (S1) measurements of the predicted versus actual outcomes of the conversation. Study S1 experimentally manipulates whether participants report predictions before the conversation. Study 6 includes a between-participants manipulation of predicted attitude change. Study S2 includes a within-participants manipulation of predicted attitude change. Participants' initial attitudes, however, are measured rather than manipulated to maintain ecological validity.
Is the phenomenon diagnosed with longitudinal methods?	In all studies with spoken conversations, we measure participants' attitudes again 1 week after the main session and find that participants' attitudes remain somewhat less polarized 1 week after their conversations than they were at baseline.
Were the manipulations validated with manipulation checks, pretest data, or outcome data?	Participants in all studies with spoken conversations completed comprehension checks in which they selected their own attitude and their partner's attitude. We also included manipulation checks in Study 6.
What possible artifacts were ruled out?	Our data suggest that differences between predicted and actual attitude change are not explained by an illusion of explanatory depth, statistical regression of the participants' attitudes, or a demand characteristic. Analyses in the Supplemental Material indicate that our results are not affected by participants who dropped out of the studies between learning their partner's attitude and having the conversation.
Statistical validity	
Was the statistical power at least 80%?	Sensitivity power analyses indicated that all studies achieved more than 80% power except Study 6, which achieved about 59% power.
Was the reliability of the dependent measure established in this publication or elsewhere in the literature?	We measured participants' explicit attitudes using standard self-report measures.
If covariates are used, have the researchers ensured they are not affected by the experimental manipulation before including them in comparisons across experimental groups?	N/A
Were the distributional properties of the variables examined and did the variables have sufficient variability to verify effects?	Yes
Generalizability to different methods	
Were different experimental manipulations used?	Studies 1, S1, 2, 4, and 5 measured predicted and actual changes in the participants' overall attitudes. Studies 4 and 5 additionally measured predicted and actual attitudes toward the participant's own examples and toward their partner's examples. We tested our hypotheses in spoken conversations about personal preferences (Studies 1 and S1), social issues (Studies 2–4), and political issues (Studies 5–6). We manipulated predicted attitude change both between participants (Study 6) and within participants (Study S2).
Generalizability to field settings	
Was the phenomenon assessed in a field setting?	No
Are the methods artificial?	The methods were naturalistic: Participants had 10-min, spoken conversations about cats and dogs, cancel culture, and Joe Biden's job performance as U.S. president with someone whose attitude differed from their own.
Generalizability to times and populations	
Are the results generalizable to different years and historic periods?	Our methods do not allow us to investigate other years or historical periods.
Are the results generalizable across populations (e.g., different ages, cultures, or nationalities)?	Our studies with spoken conversations found similar results in the United States (Studies 1, S1, 2, and 5) and the United Kingdom (Study 4). As described in the Statement of Limitations, people from Eastern cultures exhibit more dialectical thinking and might therefore have better calibrated expectations of attitude change than the Western participants we recruited in our studies.
Theoretical limitations	
What are the main theoretical limitations?	Our studies investigated spoken conversations about personal preferences, social issues, and political issues. They did not assess whether participants would similarly underestimate how much their own and others' attitudes would depolarize in debates, negotiations, exchanges of persuasive messages, or conversations about topics that have objectively correct answers such as math problems.

Note. N/A = not applicable.

networking event, but are vastly different from the conversations people have on social media, with family members behind closed doors, or those they see on the news. These settings vary along many dimensions that we did not experimentally manipulate, including the number of conversation partners (Cooney et al., 2020), their familiarity (Davis & Rusbult, 2001), their goals (Itzchakov et al., 2020, 2024; Yeomans et al., 2022), the communication medium (Daft & Lengel, 1986; Roos et al., 2020, 2022), the structured or unstructured format of the exchange (Caluwaerts et al., 2023; Pettigrew & Tropp, 2006), the presence or absence of an audience (Bateson et al., 2006; Ernest-Jones et al., 2011), time pressure (Stuhlmacher et al., 1998), and the expectedness or unexpectedness of disagreement (e.g., Fitzsimons & Finkel, 2010; Wegner & Bargh, 1998). Future research could investigate whether negative experiences in some settings (e.g., on social media) could cause people to anticipate more negative outcomes than warranted in others (e.g., in spoken conversations), as well as how the contextual variables above might moderate differences between expected and actual attitude change.

Investigating these contexts could also shed light on complementary mechanisms. Although our studies found evidence that participants' attitudes depolarized more than expected because they underestimated how much they would agree, this depolarization might also have been enabled by background processes such as perspective taking (Todd & Galinsky, 2014), high-quality listening (Itzchakov et al., 2020, 2024), balanced processing of information (Brienza et al., 2021; Puryear & Gray, 2024; Yang et al., 2024), or concrete reasoning (vs. abstract reasoning: Trope & Liberman, 2010)—each of which may be more common in spoken conversations than in other settings. If spoken conversations are uniquely likely to reduce attitude polarization, this could help to explain why our society remains polarized despite widespread access to diverse points of view through online media.

Our research also raises broader questions about the nature of attitude polarization. Political scientists traditionally measure attitude polarization using surveys in which people report their attitudes toward relatively abstract issues, such as gun control, border security, or the president's job performance (e.g., American National Election Studies, 2021). Because these issues are multifaceted and are therefore open to multiple interpretations, these measures may unintentionally confound people's underlying attitudes with their subjective construals of an issue (Zaller, 1992; Zaller & Feldman, 1992). Our research finds that people with opposing attitudes bring to mind systematically different aspects of an issue, raising the possibility that traditional measures of attitude polarization may exaggerate the magnitude of people's underlying disagreements across the political spectrum.

Finally, our research suggests a potentially novel contribution to the literature on false polarization. Whereas existing research finds that people overestimate the *magnitude* of their disagreements across the political spectrum (Fernbach & Van Boven, 2022), our research suggests people may also overestimate the *depth* of their disagreements. Participants in our studies were accurately informed of the magnitude of their disagreements, but they nonetheless underestimated their common ground because they failed to recognize that they were focused on different aspects of these issues. If these findings generalize to other political issues, they could suggest people overestimate not only the magnitude, but also the depth, of partisan polarization.

Concluding Thought

As noted by Laplace (1814/1956) in the opening quotation, differences of opinion often stem from the "various points of view where circumstances have placed us." Much like people with different physical points of view may form very different impressions of the same landscape, people with different psychological points of view may form very different attitudes toward the same issue. Our research suggests that people with opposing attitudes often fail to appreciate that they are evaluating an issue from different psychological points of view, instead presuming their difference of opinion reflects a more fundamental disagreement than it does. Civil conversations are thus surprisingly likely to reveal common ground, to reduce attitude polarization, and to paint a more complete picture of the landscape of a contentious issue.

References

Akhtar, O., & Wheeler, S. C. (2016). Belief in the immutability of attitudes both increases and decreases advocacy. *Journal of Personality and Social Psychology*, *111*(4), 475–492. https://doi.org/10.1037/pspa0000060

Alper, S. (2020). Explaining the complex effect of construal level on moral and political attitudes. *Current Directions in Psychological Science*, *29*(2), 115–120. https://doi.org/10.1177/0963721419896362

American National Election Studies. (2021). *ANES 2020 time series study full release* (February 10, 2022 version) [Data set and documentation]. https://www.electionstudies.org

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Asch, S. E. (1952). *Social psychology*. Prentice-Hall. https://doi.org/10.1037/10025-000

Bainbridge, W. S., & Stark, R. (1981). Friendship, religion, and the occult: A network study. *Review of Religious Research*, *22*(4), 313–327. https://doi.org/10.2307/3509765

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, *2*(3), 412–414. https://doi.org/10.1098/rsbl.2006.0509

Bishop, G. D., & Myers, D. G. (1974). Informational influence in group discussion. *Organizational Behavior and Human Performance*, *12*(1), 92–104. https://doi.org/10.1016/0030-5073(74)90039-7

Bland, B., Powell, B. M., & Ross, L. (2012). Barriers to dispute resolution: Reflections on peacemaking and relationships between adversaries. In R. Goodman, D. Jinks, & A. K. Woods (Eds.), *Understanding social action, promoting human rights* (pp. 265–291). Oxford University Press.

Bohns, V. K. (2016). (Mis)understanding our influence over others: A review of the underestimation-of-compliance effect. *Current Directions in Psychological Science*, *25*(2), 119–123. https://doi.org/10.1177/0963721415628011

Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology*, *38*(3), 551–569. https://doi.org/10.1111/pops.12337

Brienza, J. P., Kung, F. Y. H., & Chao, M. M. (2021). Wise reasoning, intergroup positivity, and attitude polarization across contexts. *Nature Communications*, *12*(1), Article 3313. https://doi.org/10.1038/s41467-021-23432-1

Brodsky, A., Lee, M. J., & Leonard, B. (2022). Discovering new frontiers for dyadic and team interaction studies: Current challenges and an open-source solution—SurvConf—for increasing the quantity and richness of interactional data. *Academy of Management Discoveries*, *8*(3), 337–340. https://doi.org/10.5465/amd.2021.0257

Brundage, M., Little, A. T., & You, S. S. (2024). Selection neglect and political beliefs. *Annual Review of Political Science*, *27*(1), 63–85. https://doi.org/10.1146/annurev-polisci-041322-033325

Caluwaerts, D., Bernaerts, K., Kesberg, R., Smets, L., & Spruyt, B. (2023). Deliberation and polarization: A multi-disciplinary review. *Frontiers in Political Science*, *5*, Article 1127372. https://doi.org/10.3389/fpos.2023.1127372

Chambers, J. R., Baron, R. S., & Inman, M. L. (2006). Misperceptions in intergroup conflict: Disagreeing about what we disagree about. *Psychological Science*, *17*(1), 38–45. https://doi.org/10.1111/j.1467-9280.2005.01662.x

Chambers, J. R., & Melnyk, D. (2006). Why do I hate thee? Conflict misperceptions and intergroup mistrust. *Personality and Social Psychology Bulletin*, *32*(10), 1295–1311. https://doi.org/10.1177/0146167206289979

Cooney, G., Mastroianni, A. M., Abi-Esber, N., & Brooks, A. W. (2020). The many minds problem: Disclosure in dyadic versus group conversation. *Current Opinion in Psychology*, *31*, 22–27. https://doi.org/10.1016/j.copsyc.2019.06.032

Cowan, S. K., & Baldassarri, D. (2018). "It could turn ugly": Selective disclosure of attitudes in political discussion networks. *Social Networks*, *52*, 1–17. https://doi.org/10.1016/j.socnet.2017.04.002

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science*, *32*(5), 554–571. https://doi.org/10.1287/mnsc.32.5.554

Davis, J. L., & Rusbult, C. E. (2001). Attitude alignment in close relationships. *Journal of Personality and Social Psychology*, *81*(1), 65–84. https://doi.org/10.1037/0022-3514.81.1.65

Dorison, C. A., Minson, J. A., & Rogers, T. (2019). Selective exposure partly relies on faulty affective forecasts. *Cognition*, *188*, 98–107. https://doi.org/10.1016/j.cognition.2019.02.010

Enke, B. (2020). What you see is all there is. *The Quarterly Journal of Economics*, *135*(3), 1363–1398. https://doi.org/10.1093/qje/qjaa012

Epley, N., & Kardas, M. (2021). Understanding the minds of others: Activation, application, and accuracy of mind perception. In P. A. M. Van Lange, E. T. Higgins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 163–180). Guilford Press.

Ernest-Jones, M., Nettle, D., & Bateson, M. (2011). Effects of eye images on everyday cooperative behavior: A field experiment. *Evolution and Human Behavior*, *32*(3), 172–178. https://doi.org/10.1016/j.evolhumbehav.2010.10.006

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political extremism is supported by an illusion of understanding. *Psychological Science*, *24*(6), 939–946. https://doi.org/10.1177/0956797612464058

Fernbach, P. M., & Van Boven, L. (2022). False polarization: Cognitive mechanisms and potential solutions. *Current Opinion in Psychology*, *43*, 1–6. https://doi.org/10.1016/j.copsyc.2021.06.005

Fiorina, M. P., & Abrams, S. J. (2008). Political polarization in the American public. *Annual Review of Political Science*, *11*(1), 563–588. https://doi.org/10.1146/annurev.polisci.11.053106.153836

Fishkin, J., Siu, A., Diamond, L., & Bradburn, N. (2021). Is deliberation an antidote to extreme partisan polarization? Reflections on "America in one room". *American Political Science Review*, *115*(4), 1464–1481. https://doi.org/10.1017/S0003055421000642

Fitzsimons, G. M., & Finkel, E. J. (2010). Interpersonal influences on self-regulation. *Current Directions in Psychological Science*, *19*(2), 101–105. https://doi.org/10.1177/0963721410364499

FiveThirtyEight. (2023, July 4) *How popular is Joe Biden?* https://projects.fivethirtyeight.com/biden-approval-rating/

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38. https://doi.org/10.1037/0033-2909.117.1.21

Gilovich, T. (1990). Differential construal and the false consensus effect. *Journal of Personality and Social Psychology*, *59*(4), 623–634. https://doi.org/10.1037/0022-3514.59.4.623

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Griffin, D. W., Dunning, D., & Ross, L. (1990). The role of construal processes in overconfident predictions about the self and others. *Journal of Personality and Social Psychology*, *59*(6), 1128–1139. https://doi.org/10.1037/0022-3514.59.6.1128

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, *65*(1), 399–423. https://doi.org/10.1146/annurev-psych-010213-115045

Hutchens, M. J., Hmielowski, J. D., & Beam, M. A. (2019). Reinforcing spirals of political discussion and affective polarization. *Communication Monographs*, *86*(3), 357–376. https://doi.org/10.1080/03637751.2019.1575255

IBM Corp. (2023). *IBM SPSS statistics for Windows* (Version 29.0.2.0) [Computer software]. https://www.ibm.com/support/pages/ibm-spss-statistics-2902-documentation

Itzchakov, G., Weinstein, N., Leary, M., Saluk, D., & Amar, M. (2024). Listening to understand: The role of high-quality listening on speakers' attitude depolarization during disagreements. *Journal of Personality and Social Psychology*, *126*(2), 213–239. https://doi.org/10.1037/pspa0000366

Itzchakov, G., Weinstein, N., Legate, N., & Amar, M. (2020). Can high quality listening predict lower speakers' prejudiced attitudes? *Journal of Experimental Social Psychology*, *91*, Article 104022. https://doi.org/10.1016/j.jesp.2020.104022

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, *76*(3), 405–431. https://doi.org/10.1093/poq/nfs038

Jocker, T., van der Brug, W., & Rekker, R. (2024). Growing up in a polarized party system: Ideological divergence and partisan sorting across generations. *Political Behavior*, *46*(4), 2263–2286. https://doi.org/10.1007/s11109-024-09917-x

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*(1), 1–24. https://doi.org/10.1016/0022-1031(67)90034-0

Kardas, M., Nordgren, L., & Rucker, D. (2025). *Unnecessarily divided: Civil conversations reduce attitude polarization more than people expect*. https://osf.io/tgnjk/?view_only=81cedf8e423042e5ac2fb98d68e37bb0

Kennedy, K. A., & Pronin, E. (2008). When disagreement gets ugly: Perceptions of bias and the escalation of conflict. *Personality and Social Psychology Bulletin*, *34*(6), 833–848. https://doi.org/10.1177/0146167208315158

Laplace, P. S. (1956). Concerning probability. In J. R. Newman (Ed.), *The world of mathematics* (Vol. 2, pp. 1325–1333). Simon and Schuster. (Original work published 1814)

Larrick, R. P., Mannes, A. E., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In J. I. Krueger (Ed.), *Social judgment and decision making* (pp. 227–242). Psychology Press.

Levendusky, M. (2009). *The partisan sort: How liberals became Democrats and conservatives became Republicans*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226473673.001.0001

Levy, G., & Razin, R. (2019). Echo chambers and their effects on economic and political outcomes. *Annual Review of Economics*, *11*(1), 303–328. https://doi.org/10.1146/annurev-economics-080218-030343

Liao, T. F., & Stevens, G. (1994). Spouses, homogamy, and social networks. *Social Forces*, *73*(2), 693–707. https://doi.org/10.2307/2579826

Lord, C. G., & Lepper, M. R. (1999). Attitude representation theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 31, pp. 265–343). Academic Press. https://doi.org/10.1016/S0065-2601(08)60275-0

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Mill, J. S. (1859). *On liberty*. Broadview Press.

Minson, J. A., & Dorison, C. A. (2022). Toward a psychology of attitude conflict. *Current Opinion in Psychology*, *43*, 182–188. https://doi.org/10.1016/j.copsyc.2021.07.002

Molnar, A. (2019). SMARTRIQS: A simple method allowing real-time respondent interaction in Qualtrics surveys. *Journal of Behavioral and Experimental Finance*, *22*, 161–169. https://doi.org/10.1016/j.jbef.2019.03.005

Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods*, *22*(1), 6–27. https://doi.org/10.1037/met0000086

Moore, D. A., & Cain, D. M. (2007). Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes*, *103*(2), 197–213. https://doi.org/10.1016/j.obhdp.2006.09.002

Mueller, T. S. (2021). Blame, then shame? Psychological predictors in cancel culture behavior. *The Social Science Journal*. Advance online publication. https://doi.org/10.1080/03623319.2021.1949552

Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, *83*(4), 602–627. https://doi.org/10.1037/0033-2909.83.4.602

Oosterhoff, B., Poppler, A., & Palmer, C. A. (2022). Early adolescents demonstrate peer-network homophily in political attitudes and values. *Psychological Science*, *33*(6), 874–888. https://doi.org/10.1177/09567976211063912

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist*, *54*(9), 741–754. https://doi.org/10.1037/0003-066X.54.9.741

Petrocelli, J. V., Tormala, Z. L., & Rucker, D. D. (2007). Unpacking attitude certainty: Attitude clarity and attitude correctness. *Journal of Personality and Social Psychology*, *92*(1), 30–41. https://doi.org/10.1037/0022-3514.92.1.30

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369–381. https://doi.org/10.1177/0146167202286008

Pronin, E., Puccio, C., & Ross, L. (2002). Understanding misunderstanding: Social psychological perspectives. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 636–665). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.038

Puryear, C., & Gray, K. (2024). Using "balanced pragmatism" in political discussions increases cross-partisan respect. *Journal of Experimental Psychology: General*, *153*(5), 1189–1212. https://doi.org/10.1037/xge0001554

Rattan, A., & Georgeac, O. A. M. (2017). Understanding intergroup relations through the lens of implicit theories (mindsets) of malleability. *Social and Personality Psychology Compass*, *11*(4), Article e12305. https://doi.org/10.1111/spc3.12305

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: "Naive realism" in intergroup perception and conflict. *Journal of Personality and Social Psychology*, *68*(3), 404–417. https://doi.org/10.1037/0022-3514.68.3.404

Roos, C. A., Koudenburg, N., & Postmes, T. (2022). Dealing with disagreement: The depolarizing effects of everyday diplomatic skills face-to-face and online. *New Media & Society*, *24*(9), 2153–2176. https://doi.org/10.1177/1461444821993042

Roos, C. A., Postmes, T., & Koudenburg, N. (2020). The microdynamics of social regulation: Comparing the navigation of disagreements in text-based online and face-to-face discussions. *Group Processes & Intergroup Relations*, *23*(6), 902–917. https://doi.org/10.1177/1368430220935989

Ross, L. (2013). Perspectives on disagreement and dispute resolution: Lessons from the lab and the real world. In E. Shafir (Ed.), *The behavioral foundations of public policy* (pp. 108–125). Princeton University Press. https://doi.org/10.2307/j.ctv550cbm.12

Santoro, E., & Broockman, D. E. (2022). The promise and pitfalls of cross-partisan conversations for reducing affective polarization: Evidence from randomized experiments. *Science Advances*, *8*(25), Article eabn5515. https://doi.org/10.1126/sciadv.abn5515

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, *88*(6), 895–917. https://doi.org/10.1037/0022-3514.88.6.895

Stuhlmacher, A. F., Gillespie, T. L., & Champagne, M. V. (1998). The impact of time pressure in negotiation: A meta-analysis. *International Journal of Conflict Management*, *9*(2), 97–116. https://doi.org/10.1108/eb022805

Sunstein, C. (2002). The law of group polarization. *Journal of Political Philosophy*, *10*(2), 175–195. https://doi.org/10.1111/1467-9760.00148

Swift, S. A., & Moore, D. A. (2012). Bluffing, agonism, and the role of overconfidence in negotiation. In R. Croson & G. E. Bolton (Eds.), *The Oxford handbook of economic conflict resolution* (pp. 266–278). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199730858.013.0019

Teeny, J. D., & Petty, R. E. (2022). Attributions of emotion and reduced attitude openness prevent people from engaging others with opposing views. *Journal of Experimental Social Psychology*, *102*, Article 104373. https://doi.org/10.1016/j.jesp.2022.104373

Todd, A. R., & Galinsky, A. D. (2014). Perspective-taking as a strategy for improving intergroup relations: Evidence, mechanisms, and qualifications. *Social and Personality Psychology Compass*, *8*(7), 374–387. https://doi.org/10.1111/spc3.12116

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Wald, K. A., Kardas, M., & Epley, N. (2024). Misplaced divides? Discussing political disagreement with strangers can be unexpectedly positive. *Psychological Science*, *35*(5), 471–488. https://doi.org/10.1177/09567976241230005

Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 446–496). McGraw-Hill.

Westfall, J., Van Boven, L., Chambers, J. R., & Judd, C. M. (2015). Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science*, *10*(2), 145–158. https://doi.org/10.1177/1745691615569849

Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 37–65). Lawrence Erlbaum Associates.

Yang, Z., Kung, F. Y. H., Brienza, J. P., & Chao, M. M. (2024). Bridging social divides: The role of wise reasoning in improving intergroup relations. *Translational Issues in Psychological Science*, *10*(1), 69–81. https://doi.org/10.1037/tps0000389

Yeomans, M., Schweitzer, M. E., & Brooks, A. W. (2022). The conversational circumplex: Identifying, prioritizing, and pursuing informational and relational motives in conversation. *Current Opinion in Psychology*, *44*, 293–302. https://doi.org/10.1016/j.copsyc.2021.10.001

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge University Press. https://doi.org/10.1017/CBO9780511818691

Zaller, J. R., & Feldman, S. (1992). A simple theory of the survey response: Answering questions versus revealing preferences. *American Journal of Political Science*, *36*(3), 579–616. https://doi.org/10.2307/2111583

Zhao, X., & Epley, N. (2022). Surprisingly happy to have helped: Underestimating prosociality creates a misplaced barrier to asking for help. *Psychological Science*, *33*(10), 1708–1731. https://doi.org/10.1177/09567976221097615

Received July 31, 2024
Revision received June 24, 2025
Accepted August 19, 2025 ■