In Press, Psychological Science

Misplaced Divides?: Discussing Political Disagreement with Strangers Can Be Unexpectedly Positive

Kristina Wald¹, Michael Kardas², & Nicholas Epley³

University of Pennsylvania¹ University of Oklahoma² University of Chicago³

Corresponding author: Nicholas Epley 5807 South Woodlawn Avenue Booth School of Business University of Chicago Chicago, IL 60637 Phone: 773-834-1266 Email: epley@chicagobooth.edu

Abstract

Differences of opinion between people are common in everyday life, but discussing those differences openly in conversation may be unnecessarily rare. We report 3 experiments (N = 1,264) demonstrating that people's interest in discussing important but potentially divisive topics is guided by their expectations about how positively the conversation will unfold, leaving them more interested in having a conversation with someone who agrees versus disagrees with them. People's expectations about their conversations, however, were systematically miscalibrated, such that people underestimated how positive these conversations would be—especially in cases of disagreement. Miscalibrated expectations stemmed from underestimating the degree of common ground that would emerge in conversation, and from failing to appreciate the power of social forces in conversation that create social connection. Misunderstanding the outcomes of conversation could lead people to avoid discussing disagreements more often, creating a misplaced barrier to learning, social connection, free inquiry, and free expression.

Statement of Relevance

People are commonly advised to avoid discussing potentially divisive topics in conversation, based on expectations that these conversations will be hostile and unpleasant. Avoiding these topics, however, creates a divide between people that keeps them from more accurately understanding others' perspectives, from identifying common ground and consensus, and from connecting with others in meaningful and authentic ways. At a societal level, such avoidance could enhance polarization and harden political divides. Our research suggests that people's beliefs about discussing disagreements, however, may be systematically miscalibrated, such that they underestimate how positively others will respond when discussing disagreements in dyadic conversation, and hence may be avoiding conversations that they would be happy to have had. More calibrated expectations about the outcomes of conversation would yield wiser decisions about when to discuss potentially divisive topics and when to avoid them, potentially leaving people both better connected and better informed.

Misplaced Divides?: Discussing Political Disagreement with Strangers Can Be Unexpectedly Positive

"...I am convinced that men hate each other because they fear each other. They fear each other because they don't know each other, and they don't know each other because they don't communicate with each other, and they don't communicate with each other because they are separated from each other. And God grant that something will happen to open channels of communication..." Martin Luther King, 1962

People are commonly advised to avoid discussing potentially divisive topics such as politics and religion in conversation, presumably to avoid the negative reactions that are expected to follow from discussing disagreements about deeply held values (Skitka & Bauman, 2008; Skitka & Morgan, 2014). Increasing political polarization may further heighten these fears (Finkel et al., 2020), as more extreme disagreement would presumably be expected to yield even more negative reactions. Explicit advice to avoid discussing disagreements may be unnecessary, however, as people seem to intuitively avoid engaging with others who disagree with their political views (Chen & Rohla, 2018; Gerber et al., 2012; Motyl et al., 2014; Skitka et al., 2005; Sun & Slepian, 2020), and seek out others with like-minded views (Buss, 1984; Byrne, 1961; Sprecher, 2014, 2019).

Consistent with King's conviction, however, we suggest that people's expectations about the consequences of discussing potential areas of disagreement can be systematically miscalibrated such that they overestimate how negatively these conversations would actually go. As a result, people may be overly reluctant to talk about potentially divisive topics in everyday life. Misunderstandings about the outcomes of discussing political disagreements may therefore leave people more separated from those with differing opinions than appropriately calibrated expectations would lead them to be.

We suggest that miscalibrated expectations about discussing disagreement could stem from three mechanisms. First, because expectations about outgroups are likely based on stereotypes that exaggerate central differences between groups (compared to perceptions of ingroups; Ames, 2004; Eyal & Epley, 2017), people may underestimate how much common ground they share with an outgroup member compared to an ingroup member, and hence underestimate how positive an interaction with an outgroup member might be (Fernbach & Van Boven, 2022; Mallett et al., 2008). Because conversations tend to reveal common ground to enable coordinated communication (a process called "grounding," Clark & Brennan, 1991), conversation may be especially likely to reveal unexpected areas of agreement. If people underestimate the amount of common ground present in cases of disagreement, and if conversation tends to reveal common ground, then the conversation may be a significantly more positive experience than expected. Second, dyadic conversation is a cooperative exchange that can strengthen social bonds by sharing attention, disclosing personal information, and showing responsiveness (Collins & Miller, 1994; Reis & Shaver, 1988; Wohltjen & Wheatley, 2021). However, people tend to underestimate how positive conversations (in general) will be, both in terms of positive experience and learning, suggesting that people do not fully anticipate the power of these social forces in dyadic conversations (Atir et al., 2022; Epley & Schroeder, 2014; Sandstrom & Boothby, 2021; see Epley et al., 2022 for a review). If so, then people's expectations about the outcomes of social interaction may be based more on the type of person they are interacting with (e.g., whether they agree or disagree) than on the type of interaction they are having (e.g., whether they are engaging in a dyadic conversation versus a non-interactive exchange; Kruger et al., 2005). If people fail to appreciate how social forces present in conversation can draw people together, then they may be especially likely to underestimate the positive outcomes of having an actual conversation with someone who holds opposing views.

Finally, people learn the outcomes of conversations they actually have, but do not learn the outcomes of conversations they avoid. Negative expectations that encourage avoidance are therefore likely to be based on less direct experience than positive expectations that encourage engagement (Denrell, 2005; Fazio et al., 2004; Hogarth et al., 2015). If people avoid discussing disagreements because they think the conversation will be unpleasant, then they will miss having the very conversations that could calibrate their expectations. Further, if this avoidance causes people to discuss their disagreements with strangers only rarely, then they might forget just how positive their past experiences have been, or they might not bring these past experiences to mind when anticipating the outcome of a future conversation. In contrast, if people choose to discuss important topics with people they already agree with—because they think they will be more pleasant—then people would also be more calibrated about how these conversations about the experience of discussing disagreement are likely to be especially miscalibrated compared to their expectations about discussing agreement (Epley et al., 2022).

Several existing findings support our hypotheses, but do not involve conversations about the potentially divisive topics of conversation that people often choose to avoid, such as politics and religion. For instance, people in one series of experiments expected that asking a stranger potentially sensitive questions, such as "how much is your salary?," would be a more negative experience than it actually was (Hart et al., 2021). In another series, participants underestimated how positive it would be to have a conversation about relatively deep and intimate topics, and did so more than for relatively shallow topics (Kardas et al., 2022). Most closely related to our hypotheses, political partisans expected that being exposed to an opposing political viewpoint— such as by reading an explanation of why someone voted for an opposing candidate, or by listening to someone from an opposing political party explain their views—would be a more negative experience than it actually was, and these miscalibrated expectations guided their interest in being exposed to opposing viewpoints (Dorison et al., 2019). We advance this literature by examining people's expectations and experiences in dyadic conversations about

potentially divisive topics, a context that may be both uniquely powerful for creating social connection in daily life but also especially likely to be avoided due to concerns about hostility or aggression. We also test how the context of an interaction moderates people's expectations and experiences to understand when and why people may systematically misunderstand the outcomes of social interaction.

We describe three experiments testing our hypotheses that people's expectations about the nature of a conversation guide their interest in having or avoiding it (Experiment 1), and that expectations about discussing disagreement are systematically miscalibrated such that people underestimate how positively such conversations will go with strangers (Experiment 2), compared to discussing areas of agreement. We test how underestimating common ground (Experiments 2-3) and the power of social forces in conversation (Experiment 3) could explain why people underestimate the positive outcomes of discussing disagreements. Better calibrated expectations about the outcomes of conversation could lead to wiser decisions about opening channels of communication with others.

Open Practices Statement

All studies were preregistered. All preregistrations, experimental materials, data, and analysis code are available on OSF:

https://osf.io/j6us9/?view_only=30f209930f104e8a99fd0a296033c9ce.

Experiment 1: From Expectations to Interest

Method

Participants

We recruited participants online through Prolific in exchange for \$1.00. We included an attention check question at the very beginning of the survey in order to weed out any low-effort participants and/or bots before they could proceed. We recruited a total of 471 participants in order to achieve our final targeted sample of 450 participants who passed the attention and comprehension check questions (52.67% female; 76.89% White; $M_{age} = 36.15$, $SD_{age} = 12.48$). *Procedure*

In order to test our first prediction that people's expectations about the nature of a conversation guide their interest in having one, and that people would generally be more interested in talking with someone who agreed rather than disagreed with them, we asked participants to imagine that they were about to engage in a conversation on a political topic with a stranger who either agreed or disagreed with them on an important issue at the time. Specifically, we first asked participants for their opinions on nine potentially divisive topics related to politics and/or religion, including abortion, climate change, and belief in God (see Appendix for exact statements). For each statement, participants answered the following question: "How much do you personally agree or disagree with this statement?" (-5 = strongly disagree, 5 = strongly agree).

We then randomly selected one of the nine topics, excluding any that participants reported neither agreeing nor disagreeing with (i.e., reported a 0 on the -5 to 5 scale), and randomly assigned participants to either the agreement or disagreement condition. Participants in

the agreement condition imagined discussing the selected topic for 10 minutes with someone who agreed with their opinion on the topic, whereas participants in the disagreement condition imagined discussing the selected topic with someone who disagreed with their opinion on the topic. On the next page, we asked participants to report the other person's opinion on the topic as an attention check. We allowed participants two attempts to answer this question correctly. Only two participants failed both attempts and were therefore excluded from analyses.

Conversation Expectations. Participants reported how positive or negative they expected the conversation would be in response to the following items, on scales ranging from 0 (*nothing* or *not at all*) to 10 (*a lot* or *very*): "How much do you think you would enjoy discussing this topic with this person?"; "How awkward do you think it would be to discuss this topic with this person?"; "How much do you think you would learn from this person by discussing this topic?"; "How hostile do you think this person would be toward your opinion on this topic?"; "How much do you think you would like this person?"; "How much do you think this person?"; "How much do you think this person would be toward your opinion on this topic?"; "How much do you think you would like this person?"; "How much do you think this person would be toward your opinion on this topic?"; "How much do you think you would like this person?"; "How much do you think this person?" As a manipulation check, we then asked participants to report how similar they thought their partner's opinion was to their own on a scale ranging from 0 (*very different*) to 10 (*very similar*).

Approach/Avoidance. To measure participants' interest in discussing their assigned topic during this conversation (versus avoiding the topic and discussing anything else), we asked participants to indicate their preference for avoiding or engaging on an 11-pt. scale ranging from -5 (*try very hard to AVOID discussing [topic]*) to 5 (*try very hard to DISCUSS [topic]*), with 0 labeled "neither try to avoid nor try to discuss."

Finally, we asked participants to report their gender, age, race/ethnicity, and political orientation. We also asked participants to write one sentence about what they did in the experiment (in order for us to filter out any non-human or low-effort responses). We did not exclude any participants on this basis.

Results

Manipulation check. As intended, participants in the agreement condition expected that their discussion partner's opinion would be significantly more similar to their own (M = 7.34, SD = 2.02) than did participants in the disagreement condition (M = 1.21, SD = 1.68), F(1, 432) = 1201.45, p < .001, $\eta_p^2 = .74$. This main effect was qualified by a significant conversation × topic interaction, F(8, 432) = 2.65, p = .008, $\eta_p^2 = .05$, indicating that the magnitude of this difference varied by topic. We describe these results fully in the Supplemental Online Materials.

Expectations. As preregistered, we computed a composite ($\alpha = .80$) of participants' expected enjoyment, awkwardness (reverse-scored), learning, and hostility (reverse-scored) in the conversation to create a single measure of conversation expectations.

A 2 (conversation type: agree vs. disagree) × 9 (topic) ANOVA on participants' conversation expectations yielded the predicted main effect of conversation type, F(1, 432) = 436.11, p < .001, $\eta_p^2 = .50$, indicating that participants expected more positive experiences when talking to someone they agreed with (M = 6.89, SD = 1.71) compared to someone they disagreed with (M = 3.37, SD = 1.87). This main effect was qualified by a significant conversation × topic





Figure 1. Participants' expectations about how positive a conversation would be with someone who agreed versus disagreed with them for each of 9 topics (Experiment 1). Error bars represent ± 1 standard error.

We observed similar main effects of conversation type in participants' expectations of how much they would like their partner, F(1, 432) = 265.02, p < .001, $\eta_p^2 = .38$, how much participants thought their partner would like them, F(1, 432) = 384.37, p < .001, $\eta_p^2 = .47$, and on their expected sense of connection, F(1, 432) = 331.48, p < .001, $\eta_p^2 = .43$. We again observed significant interactions between conversation type and topic for each of these measures, indicating the effect of conversation type varied somewhat by conversation topic. Because of the consistency across these measures, we report them fully in the Supplemental Online Materials.

Approach/Avoidance. Overall, participants did not report being very interested in having a conversation with a stranger about any of the discussion topics. However, participants were

more interested in *avoiding* the conversation when they disagreed with the other person (M = -1.53, SD = 2.85) than when they agreed (M = 0.28, SD = 2.28), F(1, 432) = 56.02, p < .001, $\eta_p^2 = .11$. We observed a marginally significant interaction with conversation topic, F(8, 432) = 1.94, p = .053, $\eta_p^2 = .03$, yielding the biggest effect of agreement versus disagreement on separation of church and state, t(432) = 3.53, p < .001, d = 1.39, and the smallest for belief in God, t(432) = .005, p = .996, d = 0.00. As predicted, participants' stronger reluctance to talk with someone they disagreed versus agreed with was significantly mediated by the composite measure of participants' conversation expectations, b = -2.43, 95% CI = [-2.94, -1.95], p < .001.

Although people's interest in having a conversation with a stranger is guided by many factors, these results suggest that people may be especially *un*interested in talking with a stranger who disagrees with them on politically divisive topics at least in part because they expect that the experience will be relatively negative. Experiment 2 tests the extent to which these expectations are calibrated in live conversations between strangers who either agree or disagree with each other on at least one of the nine potentially divisive topics used in Experiment 1. Misunderstanding the outcomes of conversation could lead to misplaced preferences about whom to talk with versus avoid, potentially leading people to avoid conversations they would feel good—and better informed—to have had.

Experiment 2: Expectations versus Experiences

Method

Participants

Participants completed this experiment in a university-operated laboratory in the downtown area of a large Midwestern city in exchange for \$6.00. We targeted a total of 300 participants (150 pairs). However, due to limitations on the number of participants available at our location, we ended data collection prior to reaching this goal after we could no longer feasibly obtain additional participants. We excluded 21 participants from our analyses: 5 due to incomplete data, 10 because their opinions did not align as needed on any of the topics, and 6 because of procedural errors (e.g., participants selecting the wrong opinion for the discussion topic and therefore anticipating a discussion involving disagreement when they actually agreed, or vice versa). This yielded a final sample of 198 participants recruited through a diverse community-based participant pool (34.85% female; 23.23% White; $M_{age} = 41.56$, $SD_{age} = 14.45$; 27.78% politically conservative). Although we originally planned to end the study sessions of any participants whose attitudes did not align as needed for their condition, and did so for 10 participants as described above, we later modified this procedure by reassigning such pairs to the other condition. Our final sample therefore includes 12 pairs who participated in a different condition than what they were randomly assigned to because their opinions did not align as needed for their condition (e.g., they were assigned to the disagreement condition but did not disagree on any of the conversation topics). Results are not meaningfully changed whether we include these pairs in our analyses or not. Because this violation of random assignment did not stem from self-selection by the participants themselves, we have included these 12 pairs in order to analyze data from as many participants as possible.

Procedure

To test the extent to which people's expectations about discussing agreement versus disagreement were calibrated, we asked participants to actually discuss a political or religious issue with another participant who agreed or disagreed with them, measuring their expectations before the conversation and their actual experiences afterward. We included a third uncertainty condition in which participants were not told in advance whether they agreed or disagreed with their conversation partner in order to test how baseline expectations-when people were unaware of their conversation partner's position—compared to the agree and disagree conditions (when the partner's position was known). We predicted that participants would underestimate how positive their conversations would be to a larger extent when they disagreed with their partner than when they agreed, and that the magnitude of miscalibration would be somewhere in between for participants in the uncertain condition. Because participants in the uncertain condition were randomly assigned to discuss a topic about which they agreed or disagreed, with equal frequency—yielding higher rates of disagreement in this uncertain condition than people are likely to encounter in everyday life-this uncertain condition provides a conservative test of whether people underestimate the positive experience of discussing political topics when they are uncertain of another person's attitude in daily life.

Once in the lab, an experimenter led each participant to their own room to complete a survey privately and confidentially, without meeting their partner. Both participants in each pair completed an online survey assessing the extent to which they agreed or disagreed with a series of statements about the same nine topics related to politics and/or religion as in Experiment 1. Participants indicated the extent to which they agreed or disagreed with each statement on a scale ranging from -5 (*strongly disagree*) to 5 (*strongly agree*). We also asked several additional measures about participants' attitudes, which we report in the Supplement as they are not central to our hypotheses. Finally, we asked participants to indicate any topics that they would categorically refuse to discuss in order to make sure that we were sampling from topics that participants would actually discuss. We did this for ethical reasons, given that some of these topics might involve traumatic experiences that we did not feel the participants should be forced to talk about in an experiment. A total of 94 (out of 198) participants.

We randomly assigned participants to one of three conditions (agreement, disagreement, or uncertainty) and then selected one of the statements for the participants to discuss in their conversation. In the agreement condition, we selected a statement for which the participants' attitudes fell on the same side of the scale, such that both participants supported the statement or opposed the statement. When participants agreed on multiple statements, we selected the statement with the smallest difference in attitude ratings. In the disagreement condition, we selected a statement for which the participants' attitudes fell on opposite sides of the scale, such that one participant supported the statement and the other opposed it. When participants disagreed on multiple statements, we selected the statements, we selected the statement with the largest difference in attitude ratings. In the uncertainty condition, we selected either a statement on which participants agreed

or a statement on which they disagreed, such that half the pairs in the uncertainty condition were randomly assigned to a statement they agreed on and the other half to a statement they disagreed on.

After participants completed the survey indicating their positions on the nine statements, the experimenter informed each participant that they would be having a conversation with another participant, and informed them of the statement they would be discussing. In the agreement and disagreement conditions, the experimenter also told participants whether the other participant agreed or disagreed with them, including the other participant's exact numeric rating on the statement. In the uncertainty condition, the experimenter did not provide any information about the other participant's opinion. While receiving this information, we also showed participants the following discussion prompts that they could follow when having their conversation: "1. What is your position on this issue? Why do you think you feel this way? 2. How important is this issue to you? Is there any aspect of the issue that is especially important?"

After learning about the conversation they were about to have, participants reported how positive or negative they expected the conversation would be, and how similar they thought their partner's position was to their own, on the same items used in Experiment 1. The experimenter then led one participant into the same room as the other, provided them with a reminder of the discussion prompts, and instructed them to have a discussion for the next 10 minutes. The experimenter then left the room and came back after the 10 minutes had passed. A computer camera in the room video-recorded these discussions (which participants were informed of, and consented to, during the informed consent process).

After the discussion, the experimenter brought one participant back to their original room so that each participant could complete the final set of questions on the computer privately. Participants answered the same survey items used to measure their expectations before the conversation, except that they were phrased in terms of their actual experiences (e.g., "How much did you enjoy discussing this topic with your partner?").

Finally, participants provided demographic information (including their political orientation), indicated how often they read or watch the news, indicated how much they care about American politics (both reported in the Supplement), and were then debriefed and paid for their time.

Conversation Coding. After conducting the experiment, we asked two research assistants to evaluate the nature of each conversation along several dimensions. We did not pre-register this coding procedure because we did not recognize its value until after conducting the experiment, meaning that they should be considered exploratory and subject to a more strongly powered replication. The primary purpose of these coders was to examine the extent to which participants actually discussed the videos as instructed. We obtained recordings of 90 out of the 99 unique conversation sessions, with the remaining 9 missing due to experimenter or technical errors (e.g., forgetting to properly save the recordings).

Both research assistants were blind to participants' experimental condition. We asked these coders to evaluate how much time the participants spent on topic in the conversation, and

how much time they spent in conversation in total, in minutes and seconds using a stopwatch. After watching the conversation, we asked coders to rate how much agreement was expressed in the conversation, how much disagreement was expressed in the conversation, what condition the research assistants thought participants were in ("agree" or "disagree"), how much participants seemed to enjoy the conversation, and how hostile the conversation appeared to be. Coders reported their evaluations of expressed agreement and disagreement on scales ranging from 0 (none at all) to 5 (complete agreement/disagreement), and their evaluations of enjoyment and hostility on scales ranging from 0 (not at all) to 10 (a lot/very).

Results

Because of interdependence within our experimental design, both within-participants due to repeated measures and within-dyads, all analyses involving conversation experiences reported below (and in Experiment 3) use ANOVAs on mixed linear models with repeated observations nested within participants and participants nested within pairs, unless otherwise noted.

At the suggestion of a reviewer, we conducted a post-hoc power analysis which indicated that our sample size, using this nested design and the resulting data we report below, had 99% power to detect our primary 2 (phase: expectation vs. experience) \times 3 (conversation: agree vs. disagree vs. uncertain) interaction on our main composite measure. The R code for this power analysis is included on the OSF page cited earlier.

Manipulation check. As intended, participants expected that their discussion partner's opinion would be significantly more similar to their own in the agreement condition (M = 7.83, SD = 2.09) than in the disagreement condition (M = 3.17, SD = 2.94), t(195) = 10.52, p < .001. The uncertainty condition (M = 6.30, SD = 2.17) fell in between, with participants expecting greater similarity than in the disagreement condition, t(195) = -7.07, p < .001, but less similarity than in the agreement condition, t(195) = -7.07, p < .001, but less similarity than in the agreement condition, t(195) = 3.63, p = .001.

Conversation measures. We computed a composite measure of participants' conversation expectations/experiences in the same way as in Experiment 1 ($\alpha = .60$ for expectations and $\alpha = .64$ for experiences), and analyzed the rest of our dependent measures individually. Although our pre-registration indicated that we would only conduct analyses on a composite of these items if they yielded $\alpha > .70$, and the alpha did not meet this threshold, we nevertheless decided to report the results below using this composite for ease of communication and consistency across experiments. The pattern of results remains unchanged when analyzing each item separately. We include our pre-registered analyses on individual items in the Supplemental Materials.

A 2 (phase: expectation vs. experience) × 3 (conversation: agree vs. disagree vs. uncertain) mixed-model ANOVA¹ on our composite measure yielded significant main effects of phase, F(1, 195) = 144.21, p < .001, $\eta_p^2 = .43$, and condition, F(2, 195) = 6.01, p = .003, $\eta_p = .11$, qualified by a significant phase × condition interaction, F(2, 195) = 10.44, p < .001, $\eta_p^2 = .11$, qualified by a significant phase × condition interaction, F(2, 195) = 10.44, p < .001, $\eta_p^2 = .11$, qualified by a significant phase × condition interaction, F(2, 195) = 10.44, p < .001, $\eta_p^2 = .001$, $\eta_p^2 = .0$

¹ Following a reviewer's suggestion, we also analyzed our data in a 2 (phase: expectation vs. experience) x 2 (conversation: agree vs. disagree) x 2 (partner's attitude information: informed vs. not informed) mixed-model ANOVA. This analysis yielded results similar to those we report in the main text. We report these results in the Supplemental Materials.

.10. As predicted, participants underestimated how positive their conversations would be in all conditions, but this tendency was largest in the disagreement condition, t(195) = -9.69, p < .001, d = -1.80, and smallest in the agreement condition, t(195) = -3.85, p < .001, d = -0.65, with the degree of miscalibration in the uncertain condition falling in between, t(195) = -7.01, p < .001, d = -1.19. As shown in Figure 2, participants' expectations about how positive the conversation would be varied significantly across conditions, F(2, 195) = 13.29, p < .001, $\eta_p^2 = .22$, but their reported experiences did not, F(2, 195) = 0.77, p = .465, $\eta_p^2 = .02$. Very few participants reported having negative experiences (below a 5 on our 0-10 scale) across all conditions.



Figure 2. How positive participants expected, versus actually experienced, their conversations to be when they were talking with someone who agreed with them, who disagreed with them, or whose position was uncertain (Experiment 2). Diamonds show condition means. Error bars represent ± 1 standard error.

As shown in Figure 3, we observed similar patterns on our measures of liking, perceived partner's liking, and connectedness. The phase × condition interaction was nonsignificant for liking, F(2, 195) = 2.40, p = .093, $\eta_p^2 = .02$, and was significant for perceived partner's liking and connectedness, Fs(2, 195) = 5.43 & 11.04, respectively, ps < .005, $\eta_p^2 s = .05 \& .10$.



Figure 3. Mean expected and experienced liking, perceived partner's liking, connectedness, and similarity of opinion when talking with someone participants agreed with, disagreed with, and whose position was uncertain in Experiment 2. Error bars represent ± 1 standard error.

Finally, we also tested one of our proposed mechanisms—underestimating common ground—by testing whether participants would be especially likely to underestimate how similar they would perceive the other participant's opinion to be to theirs when they disagreed (and the least likely when they agreed). A 2 (phase: expectation vs. experience) × 3 (conversation: agree vs. disagree vs. uncertain) mixed-model ANOVA on perceived similarity of opinion yielded main effects of phase, F(1,195) = 90.61, p < .001, $\eta_p^2 = .32$, and condition, F(2, 195) = 29.15, p < .001, $\eta_p^2 = .38$, qualified by a significant interaction, F(2, 195) = 16.98, p < .001, $\eta_p^2 = .15$. As shown in Figure 3, participants underestimated similarity of opinion the most in the disagreement condition, t(195) = -9.10, p < .001, d = -1.69, and did so less in the uncertainty condition, t(195) = -5.70, p < .001, d = -0.96. Participants did not significantly underestimate similarity of opinion in the agreement condition, t(195) = -1.35, p = .177, d = -0.23. Although participants in the disagreement condition recognized that they had less similar opinions after the conversation than did participants in the agreement condition, this gap was significantly smaller than participants anticipated before the conversation.

Alternate Design and Analysis. Following a reviewer's request, we also analyzed our data in a 2 (phase: expectation vs. experience) x 2 (conversation: agree vs. disagree) x 2 (partner's attitude information: informed vs. not informed) mixed-model ANOVA. We did not conceptualize our experiment in this design, and hence these analyses were not pre-registered. Although our pre-registered sample sizes were not based on this analysis, a posthoc power analysis indicated that we have 89% power to detect an effect on this 3-way interaction on our main composite measure. As expected, we observed what would be the predicted significant 3way interaction on our composite measure, as well as on our measures of partner's liking, connectedness, and similarity (p's < .014). The interaction on our measure of liking showed the same pattern but is statistically nonsignificant (p = .069). Overall, the 3-way interaction is indicating that the degree of miscalibration between expectations and experiences is largest in the disagree/informed condition, and smallest in the agree/informed condition, with the miscalibration for the uninformed conditions in the middle—consistent with our pre-registered analyses. We report complete details of this alternate analysis the Supplemental Materials.

Coding Conversations. Because the two coders' ratings were significantly correlated with each other on all items (rs > .45, ps < .001), we averaged them together to create a composite measure for all items. All mean ratings are reported in Table 1.

Our primary interest was the extent to which conversations might differ when people disagreed versus agreed on the topic of discussion (with conversations in the uncertain condition coded as agreement or disagreement based on whether they actually agreed or disagreed on the topic of discussion). It is possible, for instance, that people did not find discussing disagreement to be as unpleasant as expected because they did not actually spend time discussing their disagreement in the conversation, choosing to avoid the potentially divisive topic instead of engaging with it.

Results suggest this was not the case. Conversations involving disagreement versus agreement did not differ significantly in terms of the time spent on topic (Ms = 6.72 vs. 5.94, respectively, t(88) = -1.53, p = .129, d = -0.32), or the total time in conversation (Ms = 10.26 vs. 10.07, t(88) = -1.32, p = .191, d = -0.28). Given that experimenters were instructed to let participants talk for 10 minutes, this latter measure simply indicates that participants typically talked for their entire allotted time in conversation.

Conversations did, however, vary significantly in the degree of disagreement versus agreement expressed in the conversation. As expected, conversation partners expressed more disagreement when they actually disagreed on the conversation topic than when they agreed (Ms = 1.49 vs. 0.33, respectively, t(88) = -5.98, p < .001, d = -1.27), and expressed less agreement when they disagreed versus agreed (Ms = 3.22 vs. 3.85, t(88) = 3.42, p < .001, d = 0.72). Consistent with participants' reported experiences, these differences in the nature of the conversation did not seem to affect participants' enjoyment of the conversation as coders did not perceive significant differences in the extent to which participants seemed to be enjoying conversations involving disagreement versus agreement (Ms = 6.65 vs. 6.69, respectively, t(88) = 0.14, p = .886, d = 0.03). Coders also perceived very little hostility in the conversations overall

(means less than 1 on scales ranging from 0-10), but did observe nonsignificantly more hostility expressed in conversations involving disagreement compared to agreement (Ms = 0.85 vs. 0.35, respectively, t(88) = -1.80, p = .075, d = -0.38). The coders correctly predicted whether the participants agreed or disagreed with each other, on average, 76% of the time.

	Experiment 2		Experiment 3				
	Agree	Disagree	Agree		Disagree		
			Dialogue	Monologue	Dialogue	Monologue	
Time On Topic (Min)	5.94_{a}	6.72 _a	6.36 _b	3.71 _a	9.34 _c	4.15 _{a,b}	
Total Time Discussing (Min)	10.07_{a}	10.26 _a	7.31c	3.79 _a	10.03_{b}	4.30 _a	
Expressed Agreement	3.85 _a	3.22b	4.77 _b	4.50_{b}	3.24c	0.82 _a	
Expressed Disagreement	0.33 _a	1.49_{b}	0.23 _a	0.25 _a	1.76 _b	4.26 _c	
Enjoyment	6.69 _a	6.65 _a	8.45 _b	6.23 _a	8.60_b	5.95 _a	
Hostility	0.35 _a	0.85 _a	0.03 _a	0.00_{a}	0.07_{a}	0.16 _a	

Table 1. Mean coder ratings of conversation sessions in Experiments 2 and 3. Within each experiment, means that differ significantly (p < .05) are indicated by different subscripts.

Choice, Advice, and Expectations. As we described in Experiment 1, we believe the miscalibrated expectations documented in Experiment 2 matter because they may guide both the choices people make in their own lives and the advice they might give to others about whom to talk to and whom to avoid in conversation. To test this, we conducted a follow-up experiment (N= 400) using a 2 (decision: personal choice vs. advice) x 2 (goal: enjoy conversation vs. no goal) between-participants experimental design (see Supplemental Materials for full details). In this online experiment, we first described the procedure for Experiment 2, and then asked participants either to indicate whether they would choose to be in the agreement or disagreement condition themselves (personal choice condition), or to indicate whether they would advise a specific family member or friend to be in the agreement or disagreement condition (advice condition). We also either gave them the explicit goal to enjoy the conversation or did not mention any goal at all. Overall, a significant majority of participants (69.25%) either chose for themselves, or advised a friend, to talk with someone they agreed with (rather than someone they disagreed with) on a political topic, $\chi^2 = 59.29$, p < .001, with nonsignificant effects of target, goal, and the interaction. Participants also reported their expectations for a conversation with someone they agreed with and someone they disagreed with on one of the conversation topics (we did not specify which topic), on the same measures used by participants in Experiment 2. We replicated the same patterns we observed in Experiment 2 (with nonsignificant interactions by target and goal conditions on all measures). In addition, those participants who chose or advised agreement expected a larger gap in experiences between conversations of agreement versus disagreement (on the composite measure) compared to those who chose/advised disagreement, F(1, 398) = 41.71, p < .001, $\eta_p^2 = .09$. Consistent with their expectations that a

conversation with someone they disagreed with would be relatively unpleasant, these participants not only chose for themselves, but also advised others, to avoid those conversations.

Taken together, these findings replicate existing research documenting that people tend to underestimate how positive they will feel talking with a stranger (Dunn et al., 2007; Epley & Schroeder, 2014; Kardas et al., 2022; Sandstrom & Boothby, 2021), but suggest that this gap is especially large when having a conversation about areas of disagreement. Although participants expected a significantly less positive experience talking with someone they disagreed with on an important issue compared to someone they agreed with, their actual experiences did not differ significantly between these conversations. These miscalibrated expectations could lead people to both avoid, and to advise avoiding, conversations that might be not only surprisingly positive, but objectively positive.

In the supplemental materials, we report another experiment (Experiment S1) that provides additional evidence of the robustness of these results in which participants had a conversation with someone they knew they agreed or disagreed with on the same important topics used in Experiment x. Results indicated that participants again underestimated how positive they would feel talking with a stranger, but that this gap was especially large when discussing an area of disagreement. This pattern again emerged because of the differences in people's expectations of discussing the topic with someone they agreed versus disagreed with was significantly larger than the difference in their actual experiences of discussing the topic.

We suggest that miscalibrated expectations about discussing more potentially divisive disagreements stem from misunderstanding the power of dyadic conversation to focus on common ground and strengthen social connection. We test that mechanism directly in Experiment 3 by manipulating not only the type of person in the interaction—someone they agreed or disagreed on a political topic with—but also the type of interaction they would be having with another person-whether they would be having a dyadic conversation or not. Specifically, participants randomly assigned to the dialogue condition had a conversation with another person, whereas participants randomly assigned to the monologue condition described their beliefs and position on the topic of discussion and then watched their partner do the same. Because the dialogue condition includes an actual conversation involving responsiveness and cooperation, whereas the monologue condition does not, we expected that participants would have a more positive experience when discussing disagreement in dialogue than in monologue. We predicted that participants' expectations of the interaction, however, would be driven primarily by their level of (dis)agreement with the person they were talking with and would be relatively insensitive to the type of interaction they would be having. This would lead people to underestimate how much they would enjoy interacting with another person the most when they were having an actual conversation with a person they disagree with.

Experiment 3: Dialogue vs. Monologue

Method Participants

Due to challenges recruiting participants with diverse political opinions (i.e., those who would disagree on the political topics in our experiment), we recruited participants from two sources to have a video conversation online: from an online laboratory of people recruited from the United States maintained by a university (n = 156, including one participant who indicated "other" when asked where they were recruited from), and from Prolific (https://www.prolific.co/) (n = 84) in order to obtain our pre-registered sample size of 240 participants (before any exclusions). Conversation sessions included participants from one or both sources. Participants completed the experiment in exchange for a \$6 Amazon gift card (lab participants) or \$6 (Prolific participants). We excluded any results from 43 participants who started the experiment but did not finish it (e.g., due to partner no-shows, technical issues, or otherwise leaving the experiment early). Of those who finished the experiment, as preregistered, we excluded 12 participants due to procedural errors (e.g., the participant missing the key manipulation information due to experimenter error, being mistakenly asked to discuss a topic that did not match their assigned condition, or being mistakenly assigned to different topics within a pair), and 12 participants because their prediction of their partner's attitude before the conversation indicated that they had misunderstood the key experimental manipulation. This yielded a final sample of 216 participants (56.94% female, 61.57% White, $M_{age} = 34.44$, $SD_{age} = 13.46$; 37.96% politically conservative or in the middle). Within this sample, seven pairs completed the experiment in a different condition than the one they had been randomly assigned to because their opinions did not align as needed on any of the topics. The results are not meaningfully affected by whether these seven pairs are included or excluded, and so we retain them in all analyses to report as much data as possible. In addition, we could not identify the partners of three participants in the data file due to errors recording their ID numbers; again, results are not meaningfully affected when excluding these three participants, so we include them in all analyses below.

Procedure

In order to increase the odds of recruiting participants for each session who were likely to agree or disagree on the possible topics of conversation as our procedure required, all participants first completed a pre-screening questionnaire online to make them eligible to participate in the experiment. This pre-screening questionnaire asked participants to report their political orientation (*generally liberal, generally conservative, equally liberal and conservative,* or *I don't know*) and required them to successfully use the video recording software we used in the main experiment. Any participants who selected "I don't know" for their political orientation were told that they were ineligible for the main experiment. (Note that we only asked participants about their general political orientation in order to identify people who might be likely to agree or disagree on certain political topics, and did not have any hypotheses about how political

orientation, per se, might moderate either people's expectations or experiences. Testing how political orientation might moderate the results we observed is an interesting topic for future research but is beyond the scope of this paper.)

We directed all other participants to a page where they could provide their general availability to participate in the main experiment. Unbeknownst to them, we randomly assigned participants at that point to either the agreement or the disagreement conditions. We then used both their condition assignment and reported political orientation to match participants in timeslots, based on our belief that participants' general political leanings could be used as a proxy for the likelihood of agreeing or disagreeing on at least one of the possible conversation topics as our experimental procedure required. In the disagreement conditions, we paired one liberal and one conservative participant together. In the agreement conditions, we paired two liberal or two conservative participants together. Due to our expectations that our sampling of people online would identify people who were more liberal than the average American, we included those who selected "equally liberal and conservative" among the conservative participants because we expected they would be likely to disagree with liberal participants on at least one of these topics. Conducting our analyses below removing these politically in the middle participants, however, does not meaningfully alter the significance levels of our analyses (see Supplement). Once we paired participants, research assistants emailed each participant to confirm their timeslot.

We conducted the main experimental procedure virtually over the Zoom videoconferencing platform (<u>https://zoom.us/</u>). We scheduled two participants for each timeslot. The experimenter separated participants at the beginning of the experiment by putting each into a separate "breakout room" in order to give verbal instructions to each participant individually, and to provide each participant with a link to the online experimental survey.

Political Statement Opinions. As in Experiment 2, each participant reported their position on a range of potentially divisive political and religious topics. We included several timely topics in addition to others used in Experiment 2 (e.g., wearing masks to prevent the spread of COVID-19, renaming Confederate military bases), for a total of 12 topics (see Appendix). For each topic, participants reported how much they agreed or disagreed with the statement. Once finished with the attitude survey, the experimenter identified their discussion topic based on their reported opinions and the condition they had been randomly assigned following the same procedure used in Experiment 2. The experimenter then informed each participant of the chosen topic, pasted it into the chat window, and asked them to select it from the list of topics shown at that point on their survey in order to advance to the next page of the survey.

Participants randomly assigned to the dialogue condition then read that they would be having a conversation with the other participant in their session on the selected topic. Participants randomly assigned to the monologue condition, in contrast, read that they would record a video of themselves sharing their opinion on the selected topic, and that they would then watch their partner's video. We reminded participants of the topic they would discuss and informed them whether the other participant agreed or disagreed with them on the topic (without telling them their partner's exact numeric rating, unlike in Experiment 2). We also showed participants the same prompts as in Experiment 2 to help direct their conversation or monologue.

In order to measure participants' expectations about the interaction, we asked them to answer the same items used to measure expectations in Experiment 2 except that the items referenced the "interaction" rather than the "conversation" in order to appropriately describe both the dialogue and monologue conditions. We also added the question, "To what extent do you think your partner has false or inaccurate beliefs about this topic?" (0 = not at all, 10 = very much), along with three exploratory items measuring humanization of the other participant, which we report in the Supplemental Materials. Finally, we asked participants to guess the extent to which their partner had agreed or disagreed with the statement to be discussed on a scale ranging from -5 (*strongly disagree*) to 5 (*strongly agree*).

Once both participants had completed these measures, the experimenter started the interaction. In the dialogue conditions, the experimenter moved participants into the same breakout room on Zoom, asked them to discuss the topic for as long as they wanted and to signal to the experimenter when finished by sending a message in the Zoom chat. The experimenter then moved participants back into separate breakout rooms once their conversation was finished. In the monologue conditions, the experimenter explained to each participant separately how to record themselves describing their opinion on the topic (using a separate Qualtrics survey with the video recording software embedded), and then sent each participant a link to the other participant's recording in the Zoom chat once both had created their recordings. Each participant then watched the other participant's recording.

After finishing their interaction, participants reported their experiences on the same measures used to capture their expectations before the interaction, but phrased in the past tense to measure their actual experience of the interaction. Participants then re-rated their own opinion on the issue, and also predicted their partner's opinion in response to the following question: "How much do you think <u>your partner</u> personally agrees or disagrees with this statement (after having had the interaction)?" (-5 = *strongly disagree*, 5 = *strongly agree*). On the next survey page, participants then answered an exploratory question asking how much participants now thought their partner had agreed or disagreed with the statement before having had the interaction (see Supplemental Materials).

Finally, participants provided demographic information, were debriefed, and submitted a form to receive their payment in the form of an Amazon gift card (lab participants) or were paid on Prolific.

Conversation Coding. We obtained videos for 100 out of 118 unique conversation sessions (including some sessions for which we had excluded one member of the pair), with the remaining 18 missing due to technical or experimenter errors. We again asked two research assistants to evaluate the nature of each interaction following the same procedure as in Experiment 2. These analyses were again not pre-registered as we did not recognize their value until after conducting the experiment, and they should again be considered exploratory and subject to a more strongly powered replication. The primary purpose of these coders was to

examine the extent to which participants actually discussed the videos as instructed. We coded the videos from Experiment 3 at the same time as the videos from Experiment 2 using two different research assistants, both of whom were blind to participants' attitude condition (although were obviously aware of their interaction condition).

Results

As in Experiment 2, all analyses involving conversation expectations and experiences reported below use ANOVAs on mixed linear models with repeated observations nested within participants and participants nested within pairs, unless otherwise noted.

At the suggestion of a reviewer, we conducted a post-hoc power analysis which indicated that this sample size, using this nested design and the resulting data we report below, had 79% power to detect our primary 2 (phase: expectation vs. experience) \times 2 (attitude: agree vs. disagree) \times 2 (interaction: dialogue vs. monologue) interaction on our main composite measure. The R code for this power analysis is included on the OSF page cited earlier.

Manipulation check. As intended, participants expected that their partner's opinion was significantly more similar to their own before the conversation in the agreement conditions (M = 7.94, SD = 1.57) than in the disagreement conditions (M = 2.13, SD = 1.97), t(212) = 22.93, p < .001.

Interaction Expectations and Experiences. For this experiment, our pre-registration specified analyses of individual items. However, because the individual items show the same pattern as the composite measure of interaction expectations and experiences that we calculated in Experiment 2 (with only our awkwardness measure yielding a non-significant interaction term, and our hostility measure yielding a marginally significant interaction), we report the composite measure of participants' interaction expectations and experiences ($\alpha = 0.70$ for expectations; $\alpha = 0.55$ for experiences) in the main text to ease presentation and report the pre-registered item-level analyses in the Supplemental Materials.

As predicted, a 2 (phase: expectation vs. experience) × 2 (attitude: agree vs. disagree) × 2 (interaction: dialogue vs. monologue) mixed-model ANOVA on our composite measure yielded main effects of phase, F(1, 212) = 157.41, p < .001, $\eta_p^2 = .43$, and attitude, F(1, 212) = 44.23, p < .001, $\eta_p^2 = .17$, and a nonsignificant effect of interaction type, F(1, 212) = 2.86, p = .092, $\eta_p^2 = .01$, qualified by the critical three-way interaction, F(1, 212) = 7.64, p = .006, $\eta_p^2 = .03$. As shown in Figure 4, participants' expectations were affected by the type of person they were interacting with, such that they expected a more positive interaction when they agreed with their partner than when they disagreed with their partner, F(1, 212) = 67.89, p < .001, $\eta_p^2 = .24$. Their expectations were not significantly affected, however, by the type of interaction they would be having, F(1, 212) = 1.11, p = .293, $\eta_p^2 < .01$. Participants' actual experiences, however, *were* significantly affected by the type of interaction they were having, especially when discussing disagreement, $F_{\text{interaction}}(1, 212) = 7.60$, p = .007, $\eta_p^2 = .06$. Although participants generally underestimated how positive their experience in the interaction would be, including when simply learning about someone's opinion they disagreed with (replicating Dorison et al., 2019), this was especially true when having a *conversation* with someone participants disagreed with.



Figure 4. Mean expected and experienced positive conversational experiences by attitude condition and interaction condition in Experiment 3. Diamonds represent condition means. Error bars represent ± 1 standard error.

Notably, these results also make it clear that not all interactions between those who disagree are equally positive experiences. As shown in Figure 4, participants' experience of listening to someone they disagreed with in the monologue condition was not as positive as actually having a conversation with someone they disagreed with, t(212) = 4.44, p < .001, d = 1.18. However, as we also observed in Experiment 2, having a conversation with someone participants disagreed with did not differ significantly from the experience of having a conversation in this context with someone participants agreed with, t(212) = -0.20, p = .841, d = -0.05. Again, as shown in Figure 4, very few participants had negative experiences (below a 5 on our 0-10 scale) across all conditions.

We replicated the same pattern of significant three-way interactions on the additional measures of connectedness, F(1, 212) = 7.91, p = .005, $\eta_p^2 = .04$, liking, F(1, 212) = 8.63, p = .004, $\eta_p^2 = .04$, and perceptions of others' liking, F(1, 212) = 7.25, p = .008, $\eta_p^2 = .03$ (see Table 2). These findings provide additional support for our prediction that people underestimate the positivity of disagreement because they expect the quality of their interactions to be driven by the type of person they are talking to, without appreciating how the social forces present in a conversation can create social connection and a sense of common ground.

	Agree				Disagree				
	Dialogue		Monologue		Dialogue		Monologue		
	Expect	Experience	Expect	Experience	Expect	Experience	Expect	Experience	
Liking	6.54 _b	8.52 _c	6.73 _b	8.17 _c	5.12 _a	8.04 _c	5.29 _a	6.07 _{a,b}	
Partner's Liking	6.27 _{b,c}	7.75 _d	6.22 _{b,c}	6.15 _b	4.45 _a	7.24 _{c,d}	4.29 _a	4.24 _a	
Connected	5.90 _b	7.43 _c	6.05 _b	6.81 _{b,c}	3.63 _a	6.65 _{b,c}	3.58 _a	3.96 _a	
Similarity	7.73c	9.27 _d	8.17 _{c,d}	8.95 _d	2.18a	5.69 _b	2.07a	2.80a	
False Beliefs	1.52 _{a,b}	0.56 _a	1.36 _{a,b}	0.97 _{a,b}	4.24c	2.16b	4.31c	4.38c	

Table 2. Means of additional measures in Experiment 3. Within each row, means that differ significantly (p < .05) are indicated by different subscripts.

Further support for these hypotheses comes from participants' perceived similarity of opinion with their partner, where we observed significant main effects of phase, F(1, 212) =112.80, p < .001, $\eta_p^2 = .35$, attitude, F(1, 212) = 466.95, p < .001, $\eta_p^2 = .82$, and interaction condition, F(1, 212) = 7.59, p = .007, $\eta_p^2 = 0.07$, again qualified by the predicted three-way interaction, F(1, 212) = 10.65, p = .001, $\eta_p^2 = .05$. Participants again expected that their perceived similarity in attitudes would be driven by whether the person agreed versus disagreed with them, F(1, 212) = 526.63, p < .001, $\eta_p^2 = .83$, when their actual experience of similarity was affected by how they interacted with their partner, F(1,212) = 17.50, p < .001, $\eta_p^2 = .14$. This means that participants were especially likely to underestimate how similar they would find their partner's attitudes to be when they had a conversation with someone they disagreed with. We observed a similar pattern of results in participants' perceptions of the extent to which their partner held false beliefs, including a significant three-way interaction, F(1, 212) = 4.23, p =.041, $\eta_p^2 = .02$. Because conversation requires cooperation and coordination, it is likely to focus on areas of common ground between people that might otherwise be missed in the absence of a conversation. This, again, was a consequence of conversation that participants did not anticipate when discussing disagreement.

Perceived and Actual Attitudes. Participants' beliefs about their partner's attitudes somewhat matched their perceptions of common ground. To conduct these analyses, we reverse scored the responses of those who opposed a given position so that positive difference scores before versus after the conversation reflected more moderate perceived attitudes. A 2 (attitude: agree vs. disagree) × 2 (interaction: dialogue vs. monologue) × 2 (time: perceived attitude before vs. after conversation) ANOVA on participants' beliefs about their partner's attitudes yielded a significant main effect of time, F(1, 212) = 8.47, p = .004, $\eta_p^2 = .04$, qualified by a significant

attitude × time interaction, F(1, 212) = 46.90, p < .001, $\eta_p^2 = .18$. Participants in the disagree condition believed their partners held more moderate attitudes after the interaction than before the interaction (Ms = 2.18 vs. 3.55, respectively), t(212) = 6.49, p < .001, d = 0.95, whereas participants in the agree condition showed the opposite pattern (Ms = 4.52 vs. 3.98, respectively), t(212) = -2.99, p = .003, d = -0.38. A nonsignificant 3-way interaction indicates that perceived differences in attitude before versus after the interaction did not differ significantly between the monologue and dialogue conditions, F(1, 212) = 2.44, p = .120, $\eta_p^2 = .01$.

Participants' actual attitudes before versus after the interaction showed a similar pattern, F(1, 212) = 25.16, p < .001, $\eta_p^2 = .11$, with participants reporting significantly more moderate attitudes after the interaction when they disagreed with their partner (Ms = 2.64 vs. 4.13, respectively), t(212) = 7.58, p < .001, d = 1.11, but not reporting significantly more moderate attitudes after the interaction when they agreed with their partner (Ms = 4.66 vs. 4.84, respectively), t(212) = 1.03, p = .302, d = 0.13. A nonsignificant three-way interaction indicates that these patterns did not differ between the dialogue and monologue conditions, F(1, 212) = 0.83, p = .363, $\eta_p^2 < .01$. These results suggest that participants' attitudes may have changed over the course of the interaction, although these results have to be interpreted with some caution. Because we selected topics that participants reported disagreeing on the most, and hence were likely to report having more extreme attitudes about before the interaction, any noise in the measurement of participants' attitudes before the interaction would yield more moderate attitudes after the interaction due to regression to the mean (Chen & Risen, 2010).

Although participants' beliefs about their partner's attitudes showed a pattern similar to the participants' actual reported attitudes, additional exploratory analyses (conducted on complete pairs only) indicated that participants thought their partners had more moderate attitudes after talking about a topic they disagreed on than their partners actually did (Ms = 1.75vs. 2.84, SDs = 2.70 vs. 2.57), t(192) = -3.97, p < .001, d = -0.81, but we did not observe significant miscalibration in the other conditions (monologue-disagreement: Ms = 2.47 vs. 2.63, SDs = 2.45 vs. 3.01; dialogue-agreement: Ms = 4.63 vs. 4.72, SDs = 0.99 vs. 0.88; monologue-agreement: Ms = 4.57 vs. 4.78, SDs = 0.86 vs. 0.50), all ts < 1.15. This pattern is confirmed in a 2 (attitude: agree vs. disagree) \times 2 (interaction: dialogue vs. monologue) \times 2 (measurement type: perceived post-conversation attitudes, actual post-conversation attitudes) ANOVA that yielded a marginally significant three-way interaction, F(1, 192) = 3.28, p = .072, $\eta_{p}^{2} = .02$. Again, these calibration results in the disagreement condition are somewhat difficult to interpret because we selected, without participants' knowledge, topics in the disagreement condition that participants were the most extremely different on, which may explain why participants thought their partner had more moderate attitudes than they actually did even at baseline in the disagreement conditions (Ms = 3.59 vs. 4.16), t(376.17) = -7.42, p < .001, d = -0.75.

Conversation Coding. As in Experiment 2, the two coders' evaluations were significantly correlated on all measures (rs > .43, ps < .001), and so we averaged them together to create a

composite measure for all items. We could not calculate the correlation for hostility because one coder indicated no hostility expressed in any interaction.

It is again possible that participants had a more positive experience discussing disagreement than expected because they simply avoided talking about the topic, or maybe talked less, than participants in other conditions. As shown in Table 1, this was not the case. A 2 (attitude: agree vs. disagree) x 2 (interaction: monologue vs. dialogue) ANOVA on the estimated time spent on topic yielded main effects of attitude, F(1, 96) = 6.93, p = .010, $\eta_p^2 =$.07, and interaction condition, F(1, 96) = 36.36, p < .001, $\eta_p^2 = .27$, qualified by a marginally significant interaction, F(1, 96) = 3.81, p = .054, $\eta_p^2 = 0.04$. Participants actually spent the most time on topic when they were having a conversation with someone they disagreed with (M =9.34 minutes), and spent less time on topic when having a conversation with someone they agreed with (M = 6.36 minutes), t(96) = -3.29, p = .001. Participants generally spent less time talking on topic in the monologue condition, but did not talk significantly less when they disagreed with their partner (M = 4.15 minutes) than when they agreed (M = 3.71), t(96) = -0.47, p = .636 (means represent the sum of both participants' monologues). The total time spent talking largely mirrored the time spent on topic, yielding significant main effects of attitude, F(1,96) = 5.85, p = .017, $\eta_p^2 = .06$, and interaction condition, F(1, 96) = 47.88, p < .001, $\eta_p^2 = .33$, with a nonsignificant interaction, F(1, 96) = 2.74, p = .101, $\eta_p^2 = .03$. Discussing disagreement was not avoided, but rather seemed to be the most engaging condition in our experiment.

The interactions did vary significantly across conditions in the degree of disagreement and agreement expressed in the interaction. Consistent with conversation creating opportunities to express common ground, separate 2 (attitude: agree vs. disagree) x 2 (interaction: monologue vs. dialogue) ANOVAs on expressed disagreement and agreement yielded significant interactions, Fs (1, 96) = 56.15 & 37.76, respectively, ps < .001, $\eta_p^2 = .37 \& 0.28$. As shown in Table 1, when participants disagreed, they expressed more disagreement and less agreement in the monologue condition than in the dialogue condition, ts (96) = -9.73 & 8.91, ps < .001. Unable to respond directly in conversation to another's statement to identify points of agreement and disagreement, or to learn from each other in the process of the conversation, participants in the monologue condition simply stated their views without identifying areas of overlap that seemed to emerge in the conversations.

Coders' ratings of enjoyment in the same 2×2 ANOVA described above yielded only a significant main effect for interaction condition, F(1, 96) = 111.74, p < .001, $\eta_p^2 = .54$, indicating that coders believed the participants were enjoying their interaction significantly more in the dialogue condition than in the monologue condition, t(96) = 10.57, p < .001. Also consistent with participants' own reported experiences, the coders did not think the participants were enjoying their conversation significantly more when they agreed versus disagreed, t(96) = 0.31, p = .760. Because these interactions contained so little hostility, with one coder indicating no hostility expressed in any interaction, we report the average ratings in Table 1 but do not analyze them here.

General Discussion

People quite reasonably avoid conversations that they expect will be unpleasant, but our experiments suggest that some reasons for avoiding conversations may be somewhat unreasonable. Participants in Experiment 1 expected that conversations about potentially divisive topics would be less positive when they disagreed than when they agreed and were therefore more interested in avoiding conversations with someone they disagreed with. Experiments 2 and 3, however, suggest that these expectations may be systematically miscalibrated. In both experiments, participants expected more positive experiences talking with someone they agreed with than someone they disagreed with about a potentially divisive topic. In reality, participants had similarly positive experiences in both cases, meaning that participants were the most miscalibrated about the conversation they would also be most inclined to avoid. Mistakenly fearing a negative interaction may create misplaced partisan divides, not only keeping people from connecting with each other but also keeping people from learning about each other and from each other. Mistakenly avoiding these interactions may indeed, as Martin Luther King suggested, create more fear in social life than is warranted.

Of course, not all interactions across partisan lines are equally pleasant, as the context in which interactions occur has a profound effect on its outcomes. Two groups with opposing views who meet on the street to shout at each other will end up with a very different interaction than two people with opposing views who meet at a coffee shop to talk with each other. Even these same two people who meet at a coffee shop to talk are likely to have a very different interaction than if they were to meet online to type at each other. And yet, the power of these situational forces to guide social interaction is easy to overlook either when explaining social interaction (Gilbert & Malone, 1995) or when anticipating it (Kruger et al., 2005; Kumar & Epley, 2021; Schroeder et al., 2017). Participants in Experiment 3 expected that the outcomes of their interaction would depend on the type of person they were interacting with—whether they agreed or disagreed with the person-but did not expect that the outcomes would depend on the type of interaction they were having—whether they were engaging in dialogue or monologue. Our experiments do not suggest that people misunderstand how positively they will experience all types of interactions across partisan divides to a similar degree. Instead, they suggest that people may be uniquely likely to misunderstand the outcomes of dyadic conversation across partisan divides.

We believe our findings are of both theoretical and practical importance. Theoretically, our experiments support some existing research (Dunn et al., 2007; Mallett et al., 2008) suggesting that the well-documented preference for similarity in choosing whom to interact with (Montoya et al., 2008) may be guided by miscalibrated expectations about the outcomes of interacting with those who are similar versus different from us. Most existing research on homophily has primarily examined people's preferences, but more research is needed to examine how people's preferences for homophily actually align with their experiences (e.g., Dunn et al., 2007; Mallett et al., 2008). Our experiments also contribute to emerging research suggesting that people may underestimate how positively others respond to sociality more broadly, potentially

leading people to be less sociable than might be optimal for their own and others' wellbeing (see Epley et al., 2022 for a review). Our research documents an important moderator of undersociality: the extent to which another person disagrees on an important issue. Participants in Experiments 2 and 3 underestimated how positive an interaction would be to a greater degree when they disagreed with their conversation partner than when they agreed. Finally, our research enriches our understanding of interpersonal and intergroup conflict. Considerable existing research documents the power of intergroup contact to diminish partisan animosity, diminish intergroup prejudice, and create friendships (Broockman & Kalla, 2016; Bruneau et al., 2021; Pettigrew & Tropp, 2006; White et al., 2021). Despite these consistently positive outcomes, our experiments help to explain why people may nevertheless avoid intergroup contact: because they mistakenly expect their experience will be relatively negative.

Practically, our findings suggest that the popular advice—and common intuition—to censor oneself by avoiding discussions of politics and religion may be too extreme, at least in conversations with strangers. In addition to creating connection, calibrating beliefs, and learning about opposing viewpoints, being more open and honest in dyadic conversation may also increase feelings of authenticity and decrease regret (McDonald et al., 2020). Modern life now affords a wide array of opportunities for interacting across partisan divides, increasingly involving text-based communication channels or social media that lack the social forces that create the surprisingly positive outcomes we observed in dyadic conversations. A person wanting to connect positively across partisan divides would be wise not to simply open any channel of communication to someone on the other side, but rather to rely on the age-old technology of speaking and listening. Recognizing the extent to which partisan fears are miscalibrated could be a critical first step to opening the channels of communication that enable partisans to connect with each other more wisely, to know each other better, and to mistakenly fear each other less.

References

- Ames, D. R. (2004). Inside the Mind Reader's Tool Kit: Projection and Stereotyping in Mental State Inference. *Journal of Personality and Social Psychology*, 87(3), 340–353. https://doi.org/10.1037/0022-3514.87.3.340
- Atir, S., Wald, K. A., & Epley, N. (2022). Talking with strangers is surprisingly informative. Proceedings of the National Academy of Sciences, 119(34), e2206992119. https://doi.org/10.1073/pnas.2206992119
- Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on doorto-door canvassing. *Science*, 352(6282), 220–224. https://doi.org/10.1126/science.aad9713
- Bruneau, E., Hameiri, B., Moore-Berg, S. L., & Kteily, N. (2021). Intergroup Contact Reduces Dehumanization and Meta-Dehumanization: Cross-Sectional, Longitudinal, and Quasi-Experimental Evidence From 16 Samples in Five Countries. *Personality and Social Psychology Bulletin*, 47(6), 906–920. https://doi.org/10.1177/0146167220949004
- Buss, D. M. (1984). Marital assortment for personality dispositions: Assessment with three different data sources. *Behavior Genetics*, 14(2), 111–123. https://doi.org/10.1007/BF01076408
- Byrne, D. (1961). Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology*, *62*(3), 713–715. https://doi.org/10.1037/h0044721
- Chen, M. K., & Risen, J. (2010). How choice affects and reflects preferences: Revisiting the free-choice paradigm. *Journal of Personality and Social Psychology*, 99(4), 573–594. https://doi.org/10.1037/a0020217
- Chen, M. K., & Rohla, R. (2018). The effect of partisanship and political advertising on close family ties. *Science*. https://doi.org/10.1126/science.aaq1433
- Clark, H. H., & Brennan, S. E. (1991). *Grounding in Communication* (L. B. Resnick, J. M. Levine, & S. D. Teasley, Eds.; pp. 222–233). American Psychological Association.
- Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, 116(3), 457–475. https://doi.org/10.1037/0033-2909.116.3.457
- Denrell, J. (2005). Why Most People Disapprove of Me: Experience Sampling in Impression Formation. *Psychological Review*, 112(4), 951. https://doi.org/10.1037/0033-295X.112.4.951
- 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
- Dr. Martin Luther King's visit to Cornell College—Cornell College. (1962). Cornell College News Center. https://news.cornellcollege.edu/dr-martin-luther-kings-visit-to-cornellcollege/
- Dunn, E. W., Biesanz, J. C., Human, L. J., & Finn, S. (2007). Misunderstanding the affective consequences of everyday social interactions: The hidden benefits of putting one's best face forward. *Journal of Personality and Social Psychology*, 92(6), 990–1005.

- Epley, N., Kardas, M., Zhao, X., Atir, S., & Schroeder, J. (2022). Undersociality: Miscalibrated social cognition can inhibit social connection. *Trends in Cognitive Sciences*, 26(5), 406– 418. https://doi.org/10.1016/j.tics.2022.02.007
- Epley, N., & Schroeder, J. (2014). Mistakenly seeking solitude. *Journal of Experimental Psychology*, *143*(5), 1980–1999.
- Eyal, T., & Epley, N. (2017). Exaggerating Accessible Differences: When Gender Stereotypes Overestimate Actual Group Differences. *Personality and Social Psychology Bulletin*, 43(9), 1323–1336. https://doi.org/10.1177/0146167217713190
- Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology*, 87(3), 293. https://doi.org/10.1037/0022-3514.87.3.293
- 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
- Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, 370(6516), 533–536. https://doi.org/10.1126/science.abe1715
- Gerber, A. S., Huber, G. A., Doherty, D., & Dowling, C. M. (2012). Disagreement and the Avoidance of Political Discussion: Aggregate Relationships and Differences across Personality Traits. *American Journal of Political Science*, 56(4), 849–874. https://doi.org/10.1111/j.1540-5907.2011.00571.x
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21. https://doi.org/10.1037/0033-2909.117.1.21
- Hart, E., VanEpps, E. M., & Schweitzer, M. E. (2021). The (better than expected) consequences of asking sensitive questions. *Organizational Behavior and Human Decision Processes*, 162, 136–154. https://doi.org/10.1016/j.obhdp.2020.10.014
- Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The Two Settings of Kind and Wicked Learning Environments. *Current Directions in Psychological Science*, 24(5), 379–385. https://doi.org/10.1177/0963721415591878
- Kardas, M., Kumar, A., & Epley, N. (2022). Overly shallow?: Miscalibrated expectations create a barrier to deeper conversation. *Journal of Personality and Social Psychology*, 122(3), 367–398. https://doi.org/10.1037/pspa0000281
- Kruger, J., Epley, N., Parker, J., & Ng, Z. W. (2005). Egocentrism over e-mail: Can we communicate as well as we think? *Journal of Personality and Social Psychology*, 89(6), 925–936.
- Kumar, A., & Epley, N. (2021). It's surprisingly nice to hear you: Misunderstanding the impact of communication media can lead to suboptimal choices of how to connect with others. *Journal of Experimental Psychology: General*, 150(3), 595–607.

- Mallett, R. K., Wilson, T. D., & Gilbert, D. T. (2008). Expect the unexpected: Failure to anticipate similarities leads to an intergroup forecasting error. *Journal of Personality and Social Psychology*, 94(2), 265–277. https://doi.org/10.1037/0022-3514.94.2.94.2.265
- McDonald, R. I., Salerno, J. M., Greenaway, K. H., & Slepian, M. L. (2020). Motivated secrecy: Politics, relationships, and regrets. *Motivation Science*, 6(1), 61–78. https://doi.org/10.1037/mot0000139
- Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, 25(6), 889–922. https://doi.org/10.1177/0265407508096700
- Motyl, M., Iyer, R., Oishi, S., Trawalter, S., & Nosek, B. A. (2014). How ideological migration geographically segregates groups. *Journal of Experimental Social Psychology*, *51*, 1–14. https://doi.org/10.1016/j.jesp.2013.10.010
- 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
- Reis, H. T., & Shaver, P. (1988). Intimacy as an Interpersonal Process. In S. W. Duck (Ed.), Handbook of Personal Relationships: Theory, Relationships and Interventions (pp. 367– 389).
- Sandstrom, G. M., & Boothby, E. J. (2021). Why do people avoid talking to strangers? A mini meta-analysis of predicted fears and actual experiences talking to a stranger. *Self and Identity*, *20*(1), 47–71. https://doi.org/10.1080/15298868.2020.1816568
- Schroeder, J., Kardas, M., & Epley, N. (2017). The Humanizing Voice: Speech Reveals, and Text Conceals, a More Thoughtful Mind in the Midst of Disagreement. *Psychological Science*, 28(12), 1745–1762.
- Skitka, L. J., & Bauman, C. W. (2008). Moral Conviction and Political Engagement. *Political Psychology*, *29*(1), 29–54. https://doi.org/10.1111/j.1467-9221.2007.00611.x
- 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
- Skitka, L. J., & Morgan, G. S. (2014). The Social and Political Implications of Moral Conviction. *Political Psychology*, 35(S1), 95–110. https://doi.org/10.1111/pops.12166
- Sprecher, S. (2014). Effects of Actual (Manipulated) and Perceived Similarity on Liking in Getacquainted Interactions: The Role of Communication. *Communication Monographs*, *81*(1), 4–27. https://doi.org/10.1080/03637751.2013.839884
- Sprecher, S. (2019). Does (Dis)Similarity Information about a New Acquaintance Lead to Liking or Repulsion? An Experimental Test of a Classic Social Psychology Issue. Social Psychology Quarterly, 82(3), 303–318. https://doi.org/10.1177/0190272519855954
- Sun, K. Q., & Slepian, M. L. (2020). The conversations we seek to avoid. Organizational Behavior and Human Decision Processes, 160, 87–105. https://doi.org/10.1016/j.obhdp.2020.03.002

White, S., Schroeder, J., & Risen, J. L. (2021). When "enemies" become close: Relationship formation among Palestinians and Jewish Israelis at a youth camp. *Journal of Personality* and Social Psychology, 121(1), 76–94. https://doi.org/10.1037/pspi0000331

Wohltjen, S., & Wheatley, T. (2021). Eye contact marks the rise and fall of shared attention in conversation. *Proceedings of the National Academy of Sciences*, 118(37), e2106645118. https://doi.org/10.1073/pnas.2106645118

Appendix: Topics Used in the Experiments

Experiments 1-2

I support reproductive rights including protecting legal abortions.

I support the legalization of same-sex marriage.

I support passing stricter gun-control legislation.

I support the Black Lives Matter movement for racial equality.

I support separating families at the US-Mexico border as a deterrent for illegal immigration.

I support enforcing regulations to combat climate change.

I believe in the existence of one or more Gods.

I believe there should be no restrictions placed on religious freedom.

I support strict separation between church and state.

Experiment 3

I support reproductive rights including protecting legal abortions.

I support the legalization of same-sex marriage.

I support passing stricter gun-control legislation.

I support the Black Lives Matter movement for racial equality.

I support defunding police departments.

I support the renaming of all military bases that are currently named after Confederate soldiers.

I believe organizations (e.g., colleges, business, government offices) should require members of their community to get vaccinated against COVID-19.

I support passing more restrictive mail-in voting laws for U.S. elections.

I support enforcing regulations to combat climate change.

I believe in the existence of one or more Gods.

I believe there should be no restrictions placed on religious freedom.

I support strict separation between church and state.

Supplemental Materials for "Misplaced Divides?: Discussing Political Disagreement with Strangers Can Be Unexpectedly Positive"

Topic	(Man	Similarity ipulation Ch	eck)	Liking			Partner's Liking		
	Agree	Disagree	d	Agree	Disagree	d	Agree	Disagree	d
Church and State	7.89 _a	0.62_{b}	3.97	7.43_a	2.5 _b	2.53	7.36_a	2.38 _b	2.69
Black Lives Matter	7.19 _a	0.88_{b}	3.45	6.68 _a	2.58b	2.10	6.65 _a	1.54 _b	2.77
Climate Change	7.33 _a	1.10_{b}	3.40	6.50 _a	3.76b	1.41	6.44 _a	3.76b	1.45
Abortion	7.50 _a	0.81_{b}	3.65	6.36 _a	3.28b	1.58	6.29 _a	2.47 _b	2.07
Same-Sex Marriage	8.04 _a	1.19 _b	3.74	6.38 _a	3.06 _b	1.70	5.92 _a	2.22 _b	2.00
Family Separation	7.65 _a	1.48 _b	3.37	6.61 _a	3.41b	1.65	6.57 _a	2.67 _b	2.11
Gun Control	7.12 _a	1.56 _b	3.03	5.81 _a	3.38 _b	1.25	6.35 _a	3.12 _b	1.75
Religious Freedom	6.45 _a	1.58 _b	2.66	5.55 _a	3.08_{b}	1.27	5.23 _a	3.08_{b}	1.16
Believe in God	6.73 _a	1.88_{b}	2.64	6.14 _a	5.08 _a	0.54	6.05_{a}	4.38 _b	0.90

Experiment 1

The variation by topic for our additional measures is shown in Table S1.

Topic	Connectedness			Approach/Avoidance			
	Agree	Disagree	d	Agree	Disagree	d	
Church and State	6.89 _a	1.77 _b	2.49	1.11 _a	-2.42 _b	1.39	
Black Lives Matter	5.87 _a	1.46 _b	2.15	-0.03 _a	-2.12 _b	0.83	
Climate Change	6.11 _a	2.57 _b	1.72	1.78_{a}	-0.67 _b	0.96	
Abortion	5.61 _a	1.88 _b	1.82	-0.54 _a	-1.66 _a	0.44	
Same-Sex Marriage	5.62 _a	2.12b	1.70	-0.12a	-2.16b	0.80	
Family Separation	6.13 _a	2.15 _b	1.94	1.09 _a	-1.30 _b	0.94	
Gun Control	5.77 _a	2.44 _b	1.62	-0.12a	-1.38a	0.50	
Religious Freedom	4.91 _a	1.92 _b	1.46	0.05 _a	-1.46 _b	0.59	
Believe in God	5.59 _a	3.92 _b	0.81	-0.23 _a	-0.23 _a	0.00	

Table S1. Means and Cohen's *d*'s for additional measures in Experiment 1. Within each measure, means that differ significantly (p < .05) in a row are indicated by different subscripts.

Experiment 2

Attitude perceptions measures. We also asked participants the following questions about their opinions on each of the topics (all 0 = not at all, 10 = very or a lot): "How certain are you

about your opinion on this statement?"; "How much is your opinion on this statement connected to your core moral beliefs or convictions?"; "How much do you care about your opinion on this statement?" Averaging across all topics and all participants, the means and standard deviations of each of these measures is presented in Table S2 below.

	М	SD
Certainty	8.65	1.29
Morality	8.28	1.69
Care	8.29	1.66

Table S2. Means and standard deviations of attitude perception measures in Experiment 2.

Demographic questions: Engagement with the news and with American politics. On average, participants reported reading the news pretty often (M = 4.00, SD = 1.02, on a scale from 1 = never to 5 = a great deal), and reported caring about American politics a fair amount (M = 3.90, SD = 1.12, on a scale from 1 = not at all to 5 = very much).

Analyses on separate items in interaction experiences composite. On our measure of enjoyment, we observed the following: a main effect of phase, F(1, 195) = 84.55, p < .001, $\eta_p^2 = .30$, no main effect of condition, F(2, 195) = 2.00, p = .141, $\eta_p^2 = .04$, and a significant Phase × Condition interaction, F(2, 195) = 3.89, p = .022, $\eta_p^2 = .04$. On our measure of awkwardness, we observed a main effect of phase, F(1, 195) = 33.47, p < .001, $\eta_p^2 = .15$, a marginal effect of condition, F(2, 195) = 2.75, p = .069, $\eta_p^2 = .05$, and a significant Phase × Condition interaction, F(2, 195) = 4.79, p = .009, $\eta_p^2 = .05$. On our measure of learning, we observed a main effect of phase, F(1, 195) = 33.93, p < .001, $\eta_p^2 = .15$, a marginal effect of condition, F(2, 195) = 2.37, p= .096, $\eta_p^2 = .02$, and a significant Phase × Condition interaction, F(2, 195) = 5.08, p = .007, η_p^2 = .05. On our measure of hostility, we observed a main effect of phase, F(1, 195) = 98.23, p <.001, $\eta_p^2 = .33$, a significant effect of condition, F(2, 195) = 7.72, p < .001, $\eta_p^2 = .14$, and a significant Phase × Condition interaction, F(2, 195) = 7.72, p < .001, $\eta_p^2 = .14$, and a

Analyzing with a different condition design. We pre-registered this experiment as a 2 (phase: expectation vs. experience) × 3 (conversation: agree vs. disagree vs. uncertain) design, as reported in the main paper, but it is also possible to conceptualize this experiment as a 2 (phase: expectation vs. experience) × 2 (conversation: agree vs. disagree) × 2 (partner's attitude information: informed vs. not informed) design. This analysis yielded significant 3-way interactions on our composite measure, F(1,194) = 10.28, p = .002, $\eta_p^2 = .05$, and our measures of partner's liking, F(1,194) = 6.09, p = .014, $\eta_p^2 = .03$, connectedness, F(1,194) = 11.00, p = .001, $\eta_p^2 = .05$, and similarity, F(1,194) = 43.71, p < .001, $\eta_p^2 = .18$. The 3-way interaction on our measure of liking showed the same pattern but was statistically nonsignificant, F(1,194) = 3.34, p = .069, $\eta_p^2 = .02$. Figure S1 visualizes the means on our composite measure when analyzing this way.

Overall, these 3-way interactions indicate that the degree of miscalibration between expectations and experiences is largest in the disagree/informed condition, and smallest in the agree/informed condition, with the miscalibration for the uninformed conditions in the middle— consistent with our analyses reported in the main paper.



Figure S1. Results on the composite measure in Experiment 2 when analyzing as a 2 (phase: expectation vs. experience) x 2 (conversation: agree vs. disagree) x 2 (partner's attitude information: informed vs. not informed) experimental design.

Data exclusions. As described in the main paper, 21 participants were excluded from our main analyses in Experiment 2. Here, we provide additional information on these exclusions. 7 excluded participants were in the agreement condition, 6 were in the disagreement condition, and 8 were in the uncertainty condition (thus, there were not clear differences by condition). These numbers include some participants who did not complete the full study, for reasons explained below. Among these participants, we excluded 15 before they reached the topic assignment part of the study. For 10 of these 15 participants, their opinions did not align as needed by their randomly assigned condition on any of the topics (e.g., they did not disagree on a topic if they were in the disagree condition), and we initially stopped running the experiment when this happened rather than running them in a non-random condition (as we did in later study sessions). The remaining 5 of these 15 participants simply stopped taking the study before getting to that point in the experiment for reasons that were not recorded by the experimenters. Of the additional 6 exclusions who completed the full study, 4 were excluded for misreporting their opinions on the topic and thus being assigned to discuss a topic that did not match their condition. The remaining 2 were in the same pair and the research assistant suggested discarding the pair in the researcher log due to a potential procedural mistake, and because that pair's

opinions did not align as needed on any of the topics anyway. Thus, we cannot provide further analysis of pre-interaction attitudes for the exclusions in this study.

Experiment S1

Experiment 3 in the main text is an exact replication of another experiment we conducted that contained a programming error in the survey that rendered it useless for testing our primary 3-way interaction. We only discovered this error shortly after completing the experiment. Specifically, this error prevented participants in the monologue-disagreement condition from seeing the key experimental manipulation, meaning that they did not know that their partner disagreed with them on the issue they were discussing before their interaction. We therefore fixed the programming error and conducted the experiment again (Experiment 3 in the main text). However, because this error only affected our ability to test our key predictions involving the monologue conditions, the dialogue conditions were unaffected. At the associate editor's request, we report results from the dialogue conditions here as a replication test.

Participants. Participants were recruited the same way as in Experiment 3, from both the University of Chicago online laboratory and from Prolific. As in Experiment 3, participants completed the experiment in exchange for a \$6 Amazon gift card (lab participants) or \$6 (Prolific participants). We excluded a total of 21 participants in the dialogue conditions for procedural issues (e.g., seeing the wrong condition instructions due to research assistant errors or participants mistakenly selecting the wrong option in their survey, discussing a topic for which their opinions did not align as needed based on their condition, etc.), yielding a final sample of 119 participants in the dialogue conditions (61.34% female, 48.74% non-White, $M_{age} = 28.61$, $SD_{age} = 11.32$; 36.13% politically conservative or in the middle).

Procedure. Our procedure was identical to that of Experiment 3, except for two minor differences. First, we did not include the comprehension check of the other person's attitude that we used in Experiment 3 (although we did still ask participants to guess their partner's attitude before the interaction). Second, the discussion topics were slightly different from those in Experiment 3, as we modified them to be most timely for the time period in which we conducted each experiment. In Experiment S1, we included the following two topics that were not used in Experiment 3: "I believe everyone should wear a mask in public until there is a vaccine to prevent COVID-19" and "I support separating families at the US-Mexico border as a deterrent for illegal immigration." We did not include the following two topics that were used in Experiment 3: "I believe organizations (e.g., colleges, businesses, government offices) should require members of their community to get vaccinated against COVID-19," and "I support passing more restrictive mail-in voting laws for U.S. elections." The rest of the topics were the same as in Experiment 3.

Results

Manipulation Check. As intended, participants expected that their partner's opinion was significantly more similar to their own before the conversation in the agreement condition (M = 7.72, SD = 1.64) than in the disagreement condition (M = 2.98, SD = 2.13), t(117) = 13.69, p < .001.

Interaction Expectations and Experiences. A 2 (phase: expectation vs. experience) × 2 (attitude: agree vs. disagree) mixed model ANOVA on the composite measure (analyzing only dialogue condition participants) yielded main effects of phase, F(1,117) = 157.76, p < .001, $\eta_p^2 =$.57, and attitude, F(1,117) = 24.95, p < .001, $\eta_p^2 = .18$, qualified by the critical 2-way interaction, F(1,117) = 20.80, p < .001, $\eta_p^2 = .15$. As shown in Figure S2, these results replicate the pattern in Experiment 2 and suggest that participants underestimated the positivity of discussing topics of disagreement to a greater degree than discussing those of agreement. We replicated this pattern of a significant 2-way interaction on our measures of liking, F(1,117) =7.01, p = .009, $\eta_p^2 = .06$, perceived partner's liking, F(1,117) = 13.88, p < .001, $\eta_p^2 = .11$, and connectedness, F(1,117) = 4.34, p = .039, $\eta_p^2 = .04$. As in Experiments 2 and 3, this underestimation of positive experiences seemed to stem at least in part from greater underestimation of common ground in the disagreement condition than in the agreement condition: We again observed the predicted 2-way interaction on our measures of similarity, $F(1,117) = 14.09, p < .001, \eta_p^2 = .11$, and one's partner's false beliefs, F(1,117) = 4.77, p = .031, $\eta_{\rm p}^2 = .04$. Overall, the results of the dialogue condition in this study show additional robustness to the results in Experiment 2 in the main paper.



Figure S2. Mean expected and experienced positive conversational experiences by attitude condition in Experiment S1 (dialogue conditions only). Error bars represent ± 1 standard error.

Experiment 3

Humanization measures. To measure humanization of the other participant, we asked participants the following questions (preregistered as exploratory) about the other participant before the interaction (all 0 = not at all [trait], 10 = very [trait]): "How intelligent do you think you will find your partner to be?"; "How thoughtful do you think you will find your partner to be?"; "How rational do you think you will find your partner to be?" After the interaction, we asked the same questions, phrased in the past tense.

When analyzing a composite of these measures, there was a main effect of phase, F(1, 212) = 85.73, p < .001, $\eta_p^2 = .29$, a main effect of agreement condition, F(1, 212) = 51.51, p < .001, $\eta_p^2 = .20$, a nonsignificant main effect of interaction condition, F(1, 212) = 1.55, p = .214, $\eta_p^2 < .01$, and a nonsignificant Phase × Agreement Condition × Interaction Condition interaction, F(1, 212) = 1.39, p = .241, $\eta_p^2 < .01$.

Predicted partner rating of agreement before interaction (asked after interaction). After the interaction, participants rated their own attitude on the topic again and predicted their partner's attitude after the interaction (reported in the main paper). Following these questions, we also asked: "How much do you think your partner personally agreed or disagreed with this statement (before having had the interaction)?" (-5 = strongly disagree, 5 = strongly agree). We did not observe a significant interaction between agreement condition and interaction condition on this item (reverse-coded so that participants who had a partner who opposed the issue were switched to the positive side of the scale), F(1, 212) = 0.01, p = .914, $\eta_p^2 < .01$.

Analyses on separate items in interaction experiences composite. On our measure of enjoyment, we observed significant main effects of phase, F(1, 212) = 109.50, p < .001, $\eta_p^2 =$.34, attitude, F(1, 212) = 41.45, p < .001, $\eta_p^2 = .16$, and interaction type, F(1, 212) = 13.41, p < .001.001, $\eta_p^2 = .06$, qualified by a significant 3-way interaction, F(1, 212) = 4.92, p = .028, $\eta_p^2 = .02$. On our measure of awkwardness, we observed a significant main effect of phase, F(1, 212) =137.16, p < .001, $\eta_p^2 = .39$, a marginal effect of attitude, F(1, 212) = 3.21, p = .075, $\eta_p^2 = .01$, a nonsignificant effect of interaction type, F(1, 212) = 0.12, p = .725, $\eta_p^2 < .01$, and a nonsignificant 3-way interaction, F(1, 212) = 1.14, p = .287, $\eta_p^2 < .01$. On our measure of learning, we observed a nonsignificant effect of phase, F(1, 212) = 2.12, p = .147, $\eta_p^2 < .01$, a marginally significant effect of attitude, F(1, 212) = 3.06, p = .083, $\eta_p^2 = .03$, and nonsignificant effect of interaction type, F(1, 212) = 0.23, p = .633, $\eta_p^2 < .01$, qualified by a significant 3-way interaction, F(1, 212) = 5.62, p = .019, $\eta_p^2 = .03$. On our measure of hostility, we observed a significant effect of phase, F(1, 212) = 145.86, p < .001, $\eta_p^2 = .41$, a significant effect of attitude, $F(1, 212) = 119.16, p < .001, \eta_p^2 = .51, a nonsignificant effect of interaction type, F(1, 212) =$ 0.05, p = .824, $\eta_p^2 < .01$, and a marginally significant 3-way interaction, F(1, 212) = 2.92, p = .01.089, $\eta_p^2 = .01$.

Analyses without political independents. As noted in the main paper, we paired participants up based on their reported political ideologies in the prescreen solely to maximize the chance that they would agree or disagree as needed on any of the political topics in the experiment. To do this, we included participants who indicated being "equally liberal and conservative" in the same group as conservative participants. To test the robustness of our results, we also conducted our analyses including only participants who indicated being either liberal or conservative. Note that we did not track participants' responses from the prescreen survey to the main experiment (because, again, the prescreen was only for being able to match up participants and schedule timeslots), but we did ask participants to report their political orientation again at the end of the experiment. This question was different than the one used on the prescreen survey, asking participants to report their political orientation on a 7-point scale instead of using only 3 response options, but we relied on it for this analysis because it was the only proxy available that could be matched to participants' responses in the main experiment. Twenty-four participants provided a response in the middle of this 7-point scale (i.e., "neither liberal nor conservative"). As shown below in Figure S2, excluding all sessions including these participants still yields the predicted significant 3-way interaction on our main composite measure, F(168) = 9.84, p = .002, $\eta_p^2 = .06$.



Figure S3. Results from Experiment 2 excluding all sessions with any participant who indicated being "neither liberal nor conservative."

Data exclusions. As described in the main paper, 24 participants were excluded from our main analyses in Experiment 2. Here, we provide additional information on these exclusions. 4 participants were excluded because the research assistant made a mistake while administering the survey that prevented the participants from viewing the manipulations, 6 because the research assistant mistakenly assigned them to discuss a topic that was opposite to their condition (e.g., a topic they actually disagreed on but were told they agreed on, or vice versa), 2 because each participant within the pair was mistakenly assigned to a different topic, and 12 because the participant's prediction of their partner's attitude before the conversation indicated that they had misunderstood the key experimental manipulation (i.e., they failed a comprehension check). Of these exclusions, we cannot analyze condition information for the first 4, but of the other 8 who were excluded due to procedural errors, 4 were in the disagree/monologue condition and 4 were

in the agree/dialogue condition. Of the 12 who were excluded for failing the comprehension check about the other person's attitude, 1 was in the agree/monologue condition, 4 were in the disagree/dialogue condition, and 7 were in the disagree/monologue condition.

Regarding their pre-interaction attitudes, the 4 who were excluded who did not see the agreement/interaction condition manipulations had an average (absolute value) opinion rating of 4.25 (on a -5 to 5 scale) on the selected topic; the 8 who were excluded due to topic selection errors had average opinion ratings of 4.25 in the agreement conditions and 5.00 in the disagreement conditions (though these individuals were improperly assigned topics within the pair anyway); and the 12 who were excluded due to failing the comprehension check had a rating of 5.00 for the one participant in the agreement condition and an average rating of 4.36 for those in the disagreement conditions. Overall, these means are not markedly different from the overall (absolute value) means of 4.84 in the agreement conditions and 4.13 in the disagreement conditions among our main sample, suggesting no major differences in attitude extremity.

An additional 43 participants started Experiment 3 but did not finish the study. Although the experimenters did not record notes about why these participants did not continue the session, it is likely that they did not continue because their study partner never showed up to the study session, they encountered technical issues, or they otherwise chose to leave the study. Of these participants, 7 were in the agree/dialogue condition, 18 in agree/monologue, 7 in disagree/dialogue, and 11 in disagree/monologue. Although there were more incomplete responses in the monologue conditions than the dialogue conditions, we do not believe this stemmed from differences in how positive the participants expected their interactions to be, because the pattern of incomplete responses is the opposite of what might be predicted based on people's expectations (i.e., people would presumably most want to leave the study in the disagree/dialogue condition, based on their pre-interaction expectations).

Experiment S2: Choosing vs. Advising

As described in the main text, we conducted another experiment in which we described our procedure in Experiment 2 to a new group of participants, and then asked them to indicate whether they would personally choose to be in the agree or disagree condition, or to indicate what condition they would advise their friend/family member to be in. We also varied whether we explicitly instructed participants to try to have a positive experience or said nothing about their goal in the experiment.

Participants. We recruited participants from a virtual laboratory participant pool run by the University of Chicago who completed the experiment in exchange for \$1.60. We included an attention check question at the very beginning of the survey in order to exclude any low-effort participants (or bots) before they could proceed. We recruited a total of 569 participants in order to achieve our final targeted sample of 400 participants who passed the attention check and completed the full survey (72.50% female, 3.00% other gender; 48.50% White; $M_{age} = 28.83$, $SD_{age} = 11.57$).

Procedure. Participants read that we had previously conducted an experiment in which we asked a group of participants to discuss a topic related to politics or religion with another

person for 10 minutes. We then showed participants the 9 topics used in Experiment 2. We explained that we had assigned each participant to discuss one of the topics with either someone who agreed or disagreed with their own opinion on the topic.

We randomly assigned all participants to one of four conditions in a 2 (decision: personal choice vs. advice) x 2 (goal: enjoy conversation vs. no goal) between-participants design. In the advice conditions, we asked participants to write down the initials of a friend or family member on the next page. Participants in the personal choice conditions did not see this page. All participants then clicked to a page with one of our main dependent measures. In the choice conditions, we asked participants to imagine that they were about to participate in the experiment themselves. In the no goal condition, we then asked participants the following: "Would you choose to discuss one of the 9 political/religious topics with someone who agreed with your opinion on that topic, or with someone who disagreed with your opinion on that topic?" (multiple choice: I would choose to discuss with someone who AGREED with my opinion vs. I would choose to discuss with someone who DISAGREED with my opinion). In the enjoy conversation condition, we added the following to beginning of the question, keeping the rest of the text the same: "If your goal was to have a positive experience in the conversation, would you..." Finally, in the advice conditions, we asked participants to imagine that the friend/family member they had just thought of was about to participate in the experiment, and phrased the question similarly to the choice condition, except that we asked "Would you advise [initials] to discuss..." In the enjoy conversation condition, we similarly modified the text at the beginning of the question to indicate that their goal was for the friend/family member to have a positive experience in the conversation. In all conditions, we provided a reminder of the 9 political/religious topics at the bottom of the page.

On the next page, we asked participants to imagine that they had been asked to discuss one of the 9 topics (without specifying exactly which) either with someone who agreed with their opinion on that topic, or with someone who disagreed with their opinion on that topic. We then asked participants to answer the same conversation expectation questions used in Experiment 2, minus the similarity question, either providing their own expectations in the personal choice condition or reporting their expectations for their friend/family member in the advice condition. We asked participants to answer these expectation questions for both the agree and disagree conditions, with the order of the pages with the two sets of questions randomized. Within the text of each question, we also included a reminder of the person's attitude for that question (e.g., "How much do you think you would enjoy discussing the topic with this person (who agrees with you)?") to ensure that participants did not forget which type of target we were asking them about. On all pages, we also included reminders of the 9 topics.

Finally, participants reported demographic information (gender, age, race, political orientation) and wrote one sentence describing what they did in the study in order to allow us to exclude any obviously low-effort or bot-like responses. We did not exclude any participants on this basis. Participants then read a full debriefing and were redirected to a payment form to receive their payment.

Results

Choice/advice measure. Overall, 69.25% of participants across all conditions selected (either as a choice for themselves or as advice for a friend/family member) the agreement condition over the disagreement condition, $\chi^2 = 59.29$, p < .001.

As pre-registered, we conducted a logistic regression with the choice/advice measure as the outcome, and target condition, goal condition, and their interaction as predictors. This yielded nonsignificant main effects of target condition, b = -0.22, z = -0.76, p = .445, and goal condition, b = 0.349, z = 1.14, p = .256, and also a nonsignificant interaction, b = 0.674, z = 1.51, p = .132.

Conversation expectations. As pre-registered, we created the same composite of expectations ($\alpha = 0.69$) used in all experiments in the main paper, and analyzed the rest of the measures individually. For each measure, we conducted a 2 (decision: personal choice vs. advice) x 2 (agreement: agree vs. disagree) ANOVA, with repeated measures for the agreement variable. As predicted, we replicated the results in the experiments in the main paper, such that there was a main effect of agreement on participants' expectations of the conversation on all measures, F(1, 398)'s > 492.21, ps < .001. Overall, participants had more positive expectations for the conversations with someone who agreed versus disagreed with them. We also observed significant main effects of decision condition on the composite measure, F(1, 398) = 4.78, p = .029, $\eta_p^2 = .01$, partner's liking, F(1, 398) = 8.64, p = .003, $\eta_p^2 = .02$, and connectedness, F(1, 398) = 6.12, p = .014, $\eta_p^2 = .02$, such that participants thought they would have a better experience themselves than their friend/family member, but thought that their friend/family member would be liked more by the other person than they would be. All interactions were nonsignificant, F(1, 398)'s < 2.94, ps > .087.

We conducted an additional analysis that was not pre-registered to test whether participants who made different choices or provided different advice also expected different outcomes in conversations involving agreement versus disagreement. To conduct this analysis, we conducted an ANOVA with choice/advice (i.e., participants' own selection of agree vs. disagree, between-participants), conversation type (agree vs. disagree, within-participants), and their interaction as factors on our main composite measure. As shown in Figure S4, participants who chose or advised agreement expected a larger gap in experiences between conversations of agreement versus disagreement, F(1, 398) = $41.71, p < .001, \eta_p^2 = .09.$



Figure S4. Expectations of discussing agreement and disagreement for those who chose/advised agreement versus those who chose/advised disagreement in Supplemental Experiment 2.