Talking across the Aisle*

Luca Braghieri Peter Schwardmann Egon Tripodi

June 2, 2026

Abstract

We conduct an experiment that engages U.S. Democrats and Republicans in video conversations about policy-relevant facts. We study self-selection into conversations and their effect on information aggregation and affective polarization. Participants prefer co-partisan conversations, believing cross-partisan conversations to be less informative and less pleasant. There is more to learn from counter-partisans, but participants find it harder to extract knowledge from them. Our rich audiovisual data reveal that co- and cross-partisan conversations are strikingly similar in content and tone. Yet, knowledge extraction is impeded by participants' persistent lack of trust in the knowledge of counter-partisans. In contrast, cross-partisan interactions prove more enjoyable than anticipated and significantly reduce affective polarization, an effect that persists in an obfuscated follow-up survey three months later. More emotionally engaged conversations produce larger reductions in affective polarization. Policies encouraging cross-partisan interactions may be more successful at reducing affective polarization than at promoting information aggregation.

Keywords: cross-partisan interactions, partisan sorting, echo chambers, information diffusion, affective polarization, misperceptions.

JEL codes: C93, D83, D9.

---

* Braghieri: Bocconi University, CEPR, CESifo, and IGIER; luca.braghieri@unibocconi.it. Schwardmann: Carnegie Mellon University, CEPR, CESifo; pschwardmann@gmail.com. Tripodi: Hertie School, CESifo; egontrpd@gmail.com. We are grateful to Nejla Asimovic, Davide Cantoni, John Conlon, Stefano DellaVigna, Guillaume Fréchette, Ingar Haaland, Jeanne Hagenbach, Eran Halperin, Macartan Humphreys, Alex Imas, David Laibson, Ro'ee Levy, Matt Lowe, Johanna Mollerstrom, Lucy Page, Ricardo Perez-Truglia, Devin Pope, Chris Roth, Andreas Stegmann, Klaus Schmidt, Mateusz Stalinski, Michael Thaler, Mattie Toma, Sevgi Yuksel and Florian Zimmermann, as well as seminar participants at the ESA North America Meeting, Polarize Conference in Bergen, BEDI conference in Pittsburgh, CESifo Behavioral conference, University of Bonn, Frankfurt School and Goethe University Frankfurt, University of Warwick, Ohio State University, LMU Munich, and University of Zurich for useful comments and suggestions. Yves Le Yaouanq played an integral part in the genesis of the project. The authors gratefully acknowledge funding from the "Democracy in the 21st Century" research area of CIVICA. Tripodi acknowledges funding by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119). Ethics approval was obtained at the Hertie School (approval # 20221101-07). The experiments were pre-registered in the AsPredicted registry (#155100 and #173633). We thank Nadine Daum, Dorothea Günther, Eleonora Guseletova, Susanna Hasinnen, Tehya Lim, Riya Kejriwal, Viktoria Kruse and Myung Won (Misha) Seong for excellent research assistance and Marlene Bargou for programming the experimental interface. All errors are our own. Replication files and experimental instructions available at https://osf.io/r56b8/.

1. Introduction

Over the past two decades, scholars, public intellectuals, and policymakers have raised concerns about the negative effects of political echo chambers, online and offline environments in which individuals primarily interact and share information with co-partisans rather than counter-partisans (Bishop and Cushing, 2008; Sunstein, 2001).[1]

These concerns are generally articulated along two distinct dimensions (Bishop and Cushing, 2008; Sunstein, 2001, 2017). The first dimension relates to information: echo chambers might impair the aggregation of information that is differentially distributed across party lines, thus lowering the quality of political decision-making. The second dimension relates to social cohesion: the siloing of people into homogeneous groups reduces opportunities for cross-partisan contact that might foster empathy with and understanding of individuals on the other side of the political aisle. Despite considerable public and academic interest in echo chambers and the significant hope policymakers place in initiatives that foster cross-partisan contact, there remain significant gaps in our understanding of whether and why people self-select into politically homogeneous interactions and of the consequences of these interactions for information sharing and social cohesion.

In this paper, we make progress on these questions by engaging U.S. Democrats and Republicans in naturalistic, face-to-face video conversations about policy-relevant facts, either with co-partisans or counter-partisans. Despite the unscripted nature of the conversations, we retain substantial experimental control, enabling us to examine four key dimensions: (i) self-selection into co- vs. cross-partisan interactions, measured by participants' willingness to pay for conversations; (ii) expected and actual learning from the conversations; (iii) the affective consequences of the conversations; and (iv) the drivers of these effects.

The experiment proceeds as follows. After providing their first names or aliases, political leanings, and feelings about their own and the other party, participants take an initial quiz. The quiz consists of 14 factual multiple-choice questions covering topics that, accord-

[1]The discussion about online echo chambers was sparked by the diffusion of the internet and the rise of social media (Sunstein, 2001; Gentzkow and Shapiro, 2011); the discussion about offline echo chambers was primarily fueled by the increased degree of political homophily in patterns of geographic sorting (Bishop and Cushing, 2008; Brown et al., 2024).

ing to an exploratory survey run prior to the experiment, both Democrats and Republicans deem contentious and relevant to policy (e.g., immigration, policing, healthcare), alongside questions covering basic knowledge of U.S. politics (e.g., naming the Speaker of the House). After completing the quiz, participants are informed of the opportunity to have a conversation with another participant in the study. Depending on the treatment assignment, the conversation partner is either a co-partisan or a counter-partisan. Participants are informed of the political affiliation of their conversation partner and of the fact that, after the conversation, they will be given a chance to revise their answers to the quiz.[2] Participants then state, in an incentive-compatible elicitation: i) their willingness to pay for having the conversation, and ii) their expected improvement in the number of correct answers to the quiz as a result of having the conversation. The willingness-to-pay decision is implemented for 5 percent of the sample. Our analyses focus on the remaining 95 percent of the sample, which, independently of their willingness to pay, is released into an eight-minute video conversation with a fellow participant. After the conversation, participants are given the opportunity to revise their answers to the quiz. Participants are then asked to predict their expected improvement one more time before answering a series of questions about their experience of the conversation, their feelings about their own and the other party, and their demographics.[3]

Our first finding is that participants have a higher willingness to pay for interacting with co-partisans than for interacting with counter-partisans. This relative preference for co-partisan conversations implies that partisans, if given the choice, would self-select into echo chambers. We also elicit open-ended responses about the considerations driving participants' stated preferences (Haaland et al., 2025). Participants assigned to cross-partisan conversations are less likely to mention a desire to improve their quiz performance and more likely to mention concerns about the conversation being unpleasant. Both of these factors are highly predictive of the elicited willingness to pay and are among the most frequently stated open-ended considerations. These findings suggest that both instrumental motives (i.e., a desire to learn) and hedonic motives (i.e., the expected discomfort from an

[2]Whether the initial or revised answers to the quiz are payoff-relevant is resolved via a virtual coin flip and communicated to participants at the end of the experiment.

[3]We recruited participants online, through Prolific and CloudResearch, enabling us to achieve a scale and a level of demographic and geographic diversity that would be hard to achieve in an offline setting.

unpleasant social interaction) drive participants' relative preference for co- versus cross-partisan interactions.

Our second set of novel results examines learning and the instrumental value of co- and cross-partisan conversations. We study both actual and expected learning from those conversations by measuring participants' knowledge before and after the interaction and comparing how much they expect to improve to an objective ground truth. We find that participants expect co-partisan conversations to lead to significantly larger improvements in the revised quiz than cross-partisan conversations. These expectations are qualitatively correct, as co-partisan conversations actually do lead to larger improvements in the revised quiz, albeit only at marginal statistical significance ($p = 0.064$).

Next, we decompose the actual improvement in the quiz into two components: potential improvement and difficulties in knowledge extraction. We show that potential improvement, defined as the number of questions that a participant does not have the correct answer to and her conversation partner does, is greater in cross-partisan conversations. This is a consequence of knowledge being distributed across party lines. At the same time, the rate of knowledge extraction, defined as actual improvement conditional on the potential to improve from the conversation, is significantly lower in cross-partisan interactions.

To better understand why participants extract less knowledge from counter-partisans, we open the black box of the conversations themselves. We combine research-assistant codings of the videos, high-dimensional semantic embeddings of the transcripts, and participants' own reports about their partners and the interaction. Across these measures, co- and cross-partisan conversations look strikingly similar in observable content and tone: participants share information, justify their answers, and navigate disagreement in comparable ways. Perhaps surprisingly, cross-partisan conversations are neither more emotional nor more confrontational than co-partisan ones (Beknazar-Yuzbashev et al., 2025). The difference arises instead in participants' interpretation of the information they receive. Specifically, participants are less trusting of counter-partisans' knowledge, and this mistrust persists through conversations (Thaler, 2024), and predicts failures of knowledge extraction.

Our third set of results focuses on the hedonic aspects of the conversations. Although participants initially expressed more concern about how much they would enjoy cross-partisan conversations compared to co-partisan ones, we find that, on average, they rate

both types of interactions as equally enjoyable ex-post. Crucially, cross-partisan conversations lead to a significant reduction in affective polarization, an effect that persists in an obfuscated follow-up survey we administer more than three months after the end of the experiment. Unlike in the case of knowledge extraction, where distrust of counter-partisans' knowledge inhibits learning even though conversations unfold similarly, affective polarization is responsive to the positive surprise of cross-partisan conversations being as pleasant as co-partisan ones. Using our rich audiovisual data, we further find that depolarization is concentrated in conversations with greater emotional engagement, marked by personal disclosure, expression of feelings, and validation of one's partner.

Our findings help contextualize the widespread concerns about echo chambers as well as some of the policies proposed to address them. On the social cohesion front, we find that fact-based cross-partisan political conversations reduce affective polarization in the medium term, as conversations in other settings, often not exclusively focused on politics, have been shown to do in the short-term (Santoro and Broockman, 2022; Blattner and Koenen, 2023; Fang et al., 2025; Hobolt et al., 2024). Here, we identify emotional engagement as an important mediator. On the instrumental front, we paint a more pessimistic picture. Our results show that, even in settings where knowledge is distributed across the aisle, it might be harder for individuals to harness knowledge when talking to counter-partisans. Thus, the simple policy of encouraging cross-partisan interactions might increase social cohesion but fail to significantly improve information aggregation and people's propensity to sort into echo chambers for instrumental reasons going forward. Our results furthermore suggest that pessimistic beliefs about the potential to learn from counter-partisans are both a driver of self-selection away from cross-partisan conversations and a barrier to actually learning from them. This suggests that policies that successfully reduce biases about the other side's informedness may lead to meaningful increases in information sharing.

This paper contributes to various strands of the literature. Motivated by the concern that echo chambers harm information aggregation and exacerbate affective polarization (Sunstein, 2001, 2009, 2017), the first of these strands establishes the existence of echo chambers, both online and offline (Braghieri et al., 2024; Brown et al., 2024; Flaxman et al., 2016; Gentzkow and Shapiro, 2011; González-Bailón et al., 2023; Guess et al., 2018; Guess, 2021; Nelson and Webster, 2017). Our experiment documents the kind of self-selection that can

explain the emergence of echo chambers and proceeds to isolate hedonic and instrumental motives as plausible drivers of such self-selection.

Our paper's most novel contribution is the investigation of the expected and actual instrumental value of cross-partisan conversations, which connects to a large literature on social learning encompassing both theoretical work (DeGroot, 1974; Banerjee, 1992; Bikhchandani et al., 1992; Morris, 2001; Jackson and Yariv, 2007; Golub and Jackson, 2010; Golub and Sadler, 2016) and empirical analyses (Banerjee et al., 2013, 2019; Barrera et al., 2020; Braghieri, 2024; Chandrasekhar et al., 2022; Conlon et al., 2021; Guriev et al., 2023; Henry et al., 2022; Fehr et al., 2024; Graeber et al., 2024). Departing from much of this literature, our experiment explores bi-directional learning through conversations — a complex and yet fundamentally human mode of engagement and information exchange. Additionally, our study uniquely examines social learning through the lens of partisan identity, making ours the first investigation of: i) the beliefs that drive partisans to favor co-partisans over counter-partisan conversations, ii) the capacity of partisans to learn from counter-partisans in political conversations, and iii) the mechanisms by which, in these conversations, partisans glean less knowledge from counter-partisans than co-partisans.[4]

Our paper also contributes to a fast-growing literature on the potential of intergroup contact to reduce affective polarization and prejudice more generally (Allport, 1954; Bazzi et al., 2019; Boisjoly et al., 2006; Corno et al., 2022; Dahl et al., 2021; Dustmann et al., 2019; Enos, 2014; Fang et al., 2025; Blattner and Koenen, 2023; Lowe, 2021; Mousa, 2020; Paluck et al., 2019; Pettigrew and Tropp, 2006a; Rao, 2019; Rossiter, 2023; Rossiter and Carlson, 2024; Santoro and Broockman, 2022; Schindler and Westcott, 2021; Scacco and Warren, 2018).[5] Our

[4] A related literature, primarily employing information provision experiments, examines how individuals select and update from information sources with differing partisan labels or affiliations (Acemoglu et al., 2024; Bauer et al., 2023; Belot and Briscese, 2022; Chopra et al., 2024; Garcia-Hombrados et al., 2024; Kashner and Stalinski, 2024; Jo, 2017; Robbett et al., 2023; Burnitt et al., 2024). We build on this literature by focusing on face-to-face conversations, which: i) represent an extremely common way in which people learn in the real world, and ii) possess unique qualities that distinguish them from information provision experiments. For example, information exchange in conversations can break down when interactions become heated and confrontational. Similarly, the dynamic nature of conversations enables trust to evolve as the dialogue progresses.

[5] Alongside social segregation or lack of contact, theoretical and empirical contributions to political economics point to voter overconfidence (Ortoleva and Snowberg, 2015), competition in news provision (Perego and Yuksel, 2022), and the evolution of learning technologies (Yuksel, 2022) as important sources of increasing ideological disagreement.

paper is unique in studying learning and depolarization in the same setting and in showing that differences between the effects of co- and cross-partisan conversations depend more on participants' expectations and interpretations of the conversations than on differences in the features of the conversations themselves. As detailed in our meta-analysis in Appendix E, our study also stands out in terms of sample size, topic of discussion, and the ability to observe persistent effects on affective polarization. Beyond this, our rich audiovisual data and conversation transcripts allow us to move past the question of whether cross-partisan contact depolarizes and ask how it does so, identifying emotional engagement as the key conversational ingredient. This speaks directly to mechanisms emphasized in the contact hypothesis literature (Allport, 1954), which has long argued that meaningful, empathetic interaction is what makes contact effective, but has rarely been able to test this claim with direct measures of what happens inside the interaction itself.

The rest of the paper is organized as follows. Section 2 describes the experimental design and sample. Section 3 presents results, focusing in turn on treatment differences in self-selection into co- and cross-partisan conversations (Section 3.1), the informational value of conversations (Section 3.2), and the hedonic consequences of conversations (Section 3.3). Section 4 concludes.

2. Experimental Design and Sample

2.1. Structure of the Experiment

The overarching aim of the experiment is to facilitate naturalistic fact-based face-to-face conversations about politics between Democrats and Republicans, while imposing enough structure to measure i) the hedonic and instrumental motives that drive partisan sorting into political conversations, ii) the quality of information sharing in these conversations, and iii) the consequences of cross-partisan political conversations for social cohesion. The structure of the experiment is summarized in Figure 1. See the replication files for the complete set of experimental instructions.

After stating their first names or aliases, political leanings, and the warmth they feel toward their own and the other party, participants are given ten minutes to complete an Initial Quiz consisting of fourteen factual multiple-choice questions about politics. The quiz

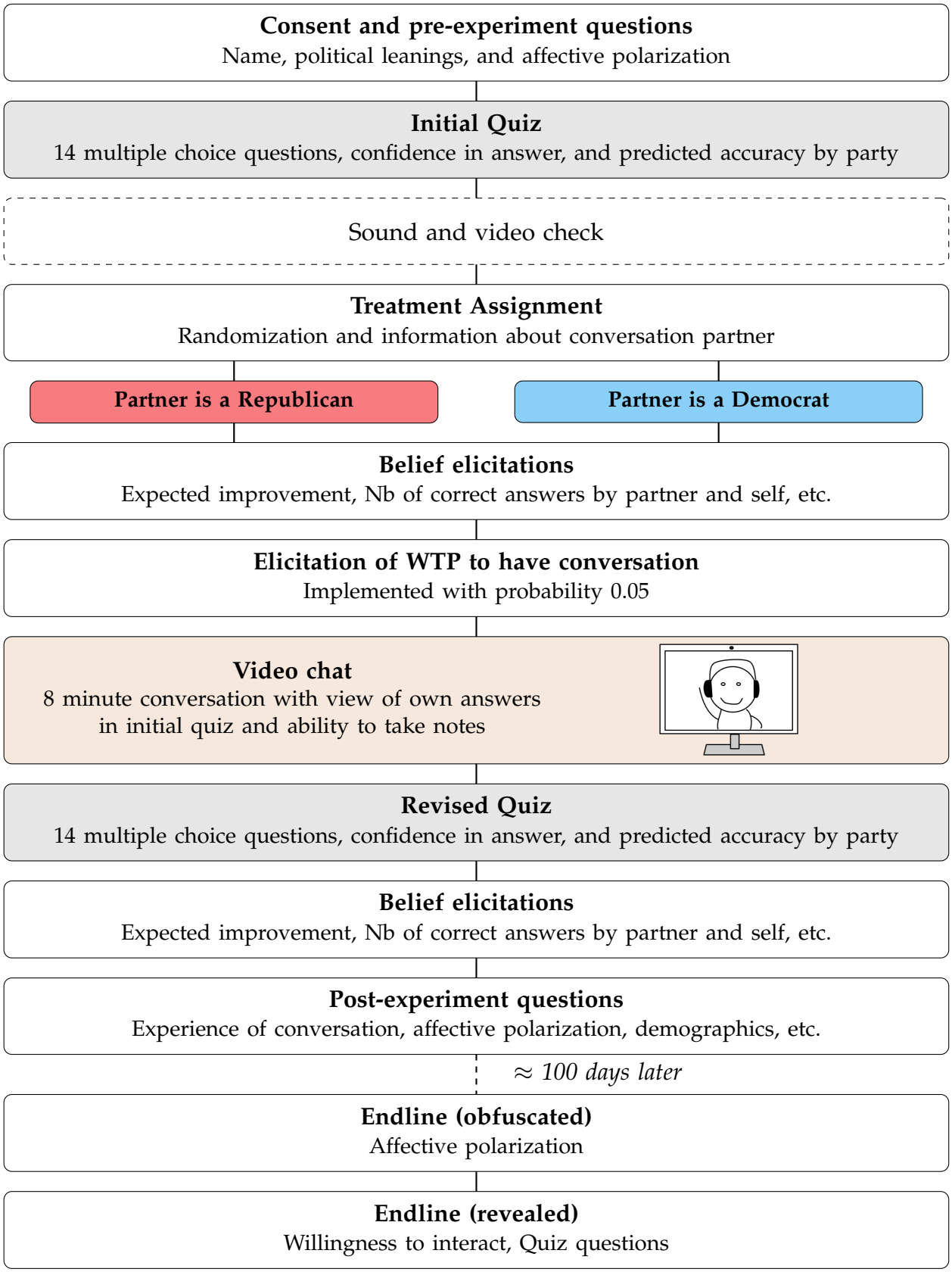

Figure 1: Design Overview

is reported in its entirety in Appendix Table A.1. A majority of quiz questions focus on topics that, according to an exploratory survey run prior to the experiment, both Democrats and Republicans deem important and contentious (e.g., immigration, policing, healthcare). The quiz also contains some basic knowledge questions about U.S. politics (e.g., naming the Speaker of the House). In order to incentivize the quiz, we provide a piece rate of 0.6 dollars per correct answer. For each quiz question, we also elicit participants' unincentivized confidence in their own answers and their beliefs about the fraction of Democrats and Republicans from a pilot version of the study who answered the question correctly.

After completing the quiz, participants are told that they will have the opportunity to engage in an eight-minute conversation about the quiz with another participant in the study and that, afterward, they will be given a chance to revise their answers to the quiz in light of the information obtained in the conversation. Participants are also informed that, at the end of the experiment, one of the two quizzes will be selected at random for payment. Depending on treatment assignment, the conversation partner is either a Democrat or a Republican, thus generating random variation in whether the conversation is between co-partisans or counter-partisans.[6] We introduce participants' prospective conversation partners by their first names or aliases and political affiliations. Thus, participants are informed of the political leanings of their conversation partner before entering the conversation, capturing real-world situations where political affiliations are either known or readily inferred.[7]

Next, we measure participants' incentivized expectation of how many more questions they will answer correctly in the Revised Quiz compared to the Initial Quiz. We then inform participants that, with a 5 percent probability, they will be assigned to the role of "deciders". Unlike other participants, "deciders" are given a choice as to whether or not they want to engage in the conversation with their partner. Introducing the role of "decider" allows us to elicit, from each participant, a willingness to pay to have (or avoid having) the eight-minute conversation.

[6]We note that 20 percent of participants drop out after treatment assignment for reasons primarily due to technical issues with the video call. Reassuringly for internal validity, Appendix Table A.2 shows no differential attrition across treatments or partisan affiliation composition.

[7]Although the political affiliation of conversation partners is often not explicitly stated, research on mediated group discussions suggests that when politically charged topics arise, most participants quickly infer each other's political affiliations (Hobolt et al., 2024).

All participants who are not randomly assigned to the role of deciders and who pass an audio and video check are released into an eight-minute unstructured video conversation with their partners. On one side of the screen, participants see a box showing a live video of their conversation partner as in a standard video call. On the other side of the screen, participants see their own answers to the Initial Quiz and a text box that they can use to take notes. We tell participants that they can use the video chat to discuss their quiz answers with their partner, but, other than that, the conversations are completely unstructured.

Immediately after the conversation, participants are given the opportunity to revise their answers to the quiz. While taking the Revised Quiz, participants are shown both their answers to the Initial Quiz and the notes that they took during the conversation with their partner.

After participants complete the Revised Quiz, we collect several additional measures. First, we again ask participants to report their expected improvement on the quiz as a result of having had the conversation. Second, we elicit five commonly used measures of affective polarization that we describe in more detail in the next section. Third, we elicit participants' beliefs about the extent to which they found their conversation partner knowledgeable. Fourth, we ask participants about their experience in the conversation. Finally, we elicit demographic characteristics.

Approximately 100 days after the intervention, we re-contacted our study participants for an "obfuscated" follow-up survey (Haaland and Roth, 2020). This survey had two main purposes: first, to measure the persistence of our affective polarization results and, second, to probe the sensitivity of those results to experimenter demand effects. The follow-up survey was "obfuscated" in the sense that we designed the survey in such a way as to make it seem unrelated to the main part of the experiment described above. Specifically, we modified the recruitment template, we changed the account from which we invited participants to take part in the study, we formatted the survey differently, and, at least for the first part of the survey, we made no reference to the main experiment. Only after participants answered a battery of questions about affective polarization were they informed that the survey was connected to our main study. We then administered our quiz one last time.[8] The full set of

[8]A unique aspect of our follow-up survey is that the degree of obfuscation decreases as respondents progress in the survey. As a result, our primary outcome of interest — a common measure of affective po-

instructions for the follow-up survey can be found in the replication files.

Discussion of design trade-offs. Our experimental design strives to strike a balance between ecological validity and experimental control. Specifically, we sought to make the conversations as naturalistic as possible to mirror the most common way in which humans typically exchange information. Allowing for naturalistic conversations is particularly important in our setting, because we wanted to capture the possibility of breakdowns in information transmission due to conversations becoming heated, confrontational, or otherwise uncooperative.

As a complement to this naturalistic approach, we introduced sufficient experimental control to accurately measure information transmission and learning. For this purpose, we incorporated a structured, albeit somewhat artificial, quiz to capture participants' best guesses about factual statements whose accuracy we could objectively verify and incentivize. To illustrate the value of this design feature, we note that some less structured experiments document a greater convergence of partisan opinions after cross-partisan conversations (Hobolt et al., 2024; Fang et al., 2025). It is tempting to interpret such convergence in opinions as a marker of learning. However, convergence in opinion and learning are distinct phenomena, and a key advantage of our design's ability to measure learning is that we can make that distinction. In particular, our experiment, in line with the existing literature, documents markers of converging opinions: participants' responses to the quiz questions become less stereotypical of their party's typical answers after cross-partisan conversations (see Section D), and participants in those conversations increase the extent to which they value the ideas of the opposing party (see Section 3.3.2). However, these markers of convergence in opinions do not imply learning. In fact, the experimental results allow us to rule out even small positive effects of cross-partisan conversations on actual learning (see Section 3.2).

Another feature of our experimental design warrants further discussion. By incentivizing quiz performance, our experiment artificially increases the instrumental value of the conversations and focuses them on the topics covered by the quiz. While this design choice limits our ability to assess the "natural" magnitude of the instrumental value of co- and

larization — is elicited under full obfuscation, whereas other outcomes are elicited under less obfuscation.

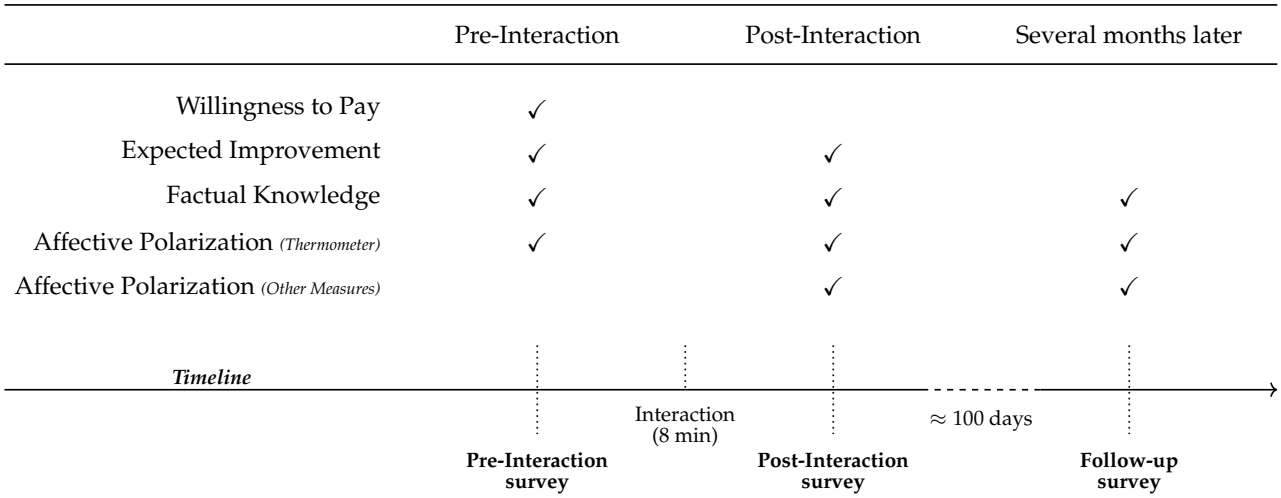

Figure 2: Timeline of Outcomes

cross-partisan political conversations in real-world settings such as elections, it enables us to explore both perceived and actual differences in the instrumental value of co- and cross-partisan interactions, as well as how instrumental considerations influence self-selection into co- vs. cross-partisan interactions. We focus on the difference in the instrumental value of co- and cross-partisan conversations, because the relative value of instrumental and hedonic motives for engaging in partisan interactions varies naturally across contexts, whereas the differences we document are likely to remain relevant in any settings where the instrumental value of politically charged conversations is high.

2.2. Main outcomes

We measure several of our main outcomes at different points in time. Figure 2 summarizes when each outcome is measured.

Willingness to pay. We use the incentive-compatible Becker-DeGroot-Marschak mechanism to measure willingness to pay to have (or avoid having) the conversation (Becker et al., 1964). Specifically, our willingness-to-pay elicitation proceeds in three steps. First, we ask participants whether they would be willing to give up some of their baseline payment in order to have the conversation with their partner, whether they would need to receive an additional bonus on top of their baseline payment in order to have the conversation with their partner, or whether they are indifferent between having and not having the conver-

sation (which involves them keeping their baseline payment and receiving no additional bonus). Second, we elicit the minimum amount of money — on a scale from zero to twice the baseline payment — that would make participants indifferent between having the conversation and not having it. If participants choose a number below their baseline payment, it means they would be willing to give up some of their baseline payment in order to have the conversation; conversely, if participants choose a number above their baseline payment, it means they would need to be paid a bonus on top of their baseline payment in order to have the conversation. Third, we inform participants via a pop-up banner of any inconsistency between their categorical response in step one and their stated willingness to pay in step two. Specifically, whenever an inconsistency is detected, the pop-up banner explains the inconsistency and discourages participants from continuing to the next page. The pop-up banner disappears as soon as the inconsistency is resolved. For ease of interpretation, we subtract participants' willingness to pay from the baseline payment. This way, a positive willingness to pay indicates that participants are happy to pay some money in order to have a conversation with their partner, a negative willingness to pay indicates that participants need to be paid some money in order to have a conversation with their partner, and a willingness to pay of zero indicates that participants are indifferent between having and not having the conversation.

Expected improvement. Participants are asked to report their expectation of the additional number of correct answers in the Revised Quiz, which occurs after the conversation, compared to the Initial Quiz. Responses are numerical and participants are allowed to include one decimal point. To incentivize this question, we use a binarized scoring rule (Hossain and Okui, 2013) that is incentive-compatible irrespective of a participant's degree of risk aversion. The rule works as follows: the closer a participant's answer is to the realized state of the world (her actual improvement), the higher the probability that she wins a fixed bonus payment.[9] Of course, incentivizing the expected improvement question might affect participants' behaviors in the Revised Quiz. Specifically, one might worry that participants tailor their answers in the Revised Quiz to match their stated expectations. In order to check whether incentives affect participants' answers to the expected improvement question, we

[9]Danz et al. (2022) show that simplifying the instructions of the binarized scoring rule improves the accuracy of belief elicitation. Based on this insight, we provide a non-quantitative explanation of the incentives in the survey instructions and we include a link to the quantitative details for further reference.

incentivize the question only for a randomly drawn half of participants. As shown in Appendix Table A.3, we find no evidence that incentives affect participants' answers to the expectation question or their behavior in the Revised Quiz.

Factual knowledge. We measure factual knowledge as the score on our fourteen-question political quiz. To make sure that correct answers are strongly correlated with actual knowledge of a topic, each multiple-choice question has five possible answers, only one of which is correct. The political quiz is incentivized, but we note that there are pros and cons to incentivizing it. The obvious advantage of incentivizing the knowledge quiz is that participants are more likely to care about choosing the correct answer. The main disadvantage is that participants might attempt to cheat by, for instance, looking up the answers to the questions on the Internet.

We implemented several strategies to minimize or measure cheating behavior. First, we informed participants from the start that the study aimed to assess only what they knew or learned during the experiment, and that any outside research would be penalized. Second, we designed several of the quiz questions to be difficult to answer through online searches and we limit the time available to answer the quiz. Third, we introduced a surprise three-question bonus quiz near the end of the study, containing questions that could be easily answered through online searches. The bonus quiz had the same incentive structure as the other quizzes. By examining the percentage of participants who answered all three questions in the bonus quiz correctly, we can arguably gain an estimate of the number of participants who might have cheated by searching for answers online. Less than 2 percent of participants answered all three questions correctly, alleviating concerns about cheating.[10] Lastly, unless cheating varies significantly by treatment status, it is unlikely to affect most of our findings.[11] The list of questions for both the main and surprise quizzes can be found in Appendix Table A.1.

Affective polarization. We include five standard measures of affective polarization (Levy, 2021). First, as our main measure, we employ a *feeling thermometer* by eliciting respondents' feelings towards each party's affiliates on a scale from 0 (extremely negative) to 100 (ex-

[10]The main results remain qualitatively identical when these participants are excluded from the analysis.

[11]To test for differential cheating across treatments, we compare the time taken to reach the conversation after the treatment revelation page and find no significant differences.

tremely positive), and we construct a measure of affective polarization based on the difference between the two. Second, we measure the *difficulty in understanding the perspectives of others*: respondents rate the perceived difficulty of understanding each party's point of view on a 5-point scale from 1 (not at all difficult) to 5 (extremely difficult). Third, we elicit the perceived *importance of considering each party's perspective* on a 5-point scale from 1 (not important at all) to 5 (extremely important). Fourth, we ask respondents to report their *perceived number of good ideas* for each party using a 4-point scale, with options ranging from 0 (almost no good ideas) to 3 (a lot of good ideas). For each of the first four measures, we compute the difference between the values elicited for one's own and the other party, re-orienting variables in such a way that larger values indicate higher levels of polarization. Fifth, we elicit the emotional reactions to hypothetical *marriages of one's own children to an out-party member*, on a 3-point scale from 0 (not upset at all) to 2 (very upset). This response is used directly, without computing a gap between attitudes towards one's own and other party members. We then derive a composite index of affective polarization by: i) standardizing each outcome, ii) orienting each outcome in such a way that higher numbers always indicate a higher degree of affective polarization, and iii) taking an equally weighted average of the standardized and re-oriented variables.

2.3. Procedures

Recruitment. We recruited study participants from both Prolific and CloudResearch Connect. The algorithm that randomly matches participants with a Democrat or with a Republican required that a sufficient number of participants take the experiment at the same time. We induced this required thickness by rolling out various recruitment surveys on a daily basis that invited participants for the main experimental session later in the day. We also used the recruitment survey to screen out Independents. Overall, we conducted 24 sessions of the experiment between December 13^{th}, 2023 and March 25^{th}, 2024.

On May 4^{th}, 2024, on average 98 days after participants took the first survey, we targeted those participants again and invited them to the obfuscated follow-up study. As discussed, we used different Prolific and CloudResearch accounts to recruit participants, as well as a different consent form.

Software. We ran the recruitment and follow-up surveys using Qualtrics. The main

Table 1: Sample Demographics

	(1) Overall	(2) Democrats	(3) Republicans
Age	42.00	40.24	44.44
	(0.43)	(0.54)	(0.69)
Female	0.47	0.51	0.42
	(0.02)	(0.02)	(0.02)
White	0.77	0.71	0.85
	(0.01)	(0.02)	(0.02)
Black	0.15	0.19	0.09
	(0.01)	(0.02)	(0.01)
Asian	0.11	0.13	0.07
	(0.01)	(0.01)	(0.01)
Latino Identity	0.07	0.08	0.06
	(0.01)	(0.01)	(0.01)
Graduated College	0.22	0.25	0.17
	(0.01)	(0.02)	(0.02)
Household Income over 50k	0.70	0.67	0.75
	(0.01)	(0.02)	(0.02)
Urban Residence	0.54	0.60	0.45
	(0.02)	(0.02)	(0.02)
Voted for Trump	0.34	0.02	0.79
	(0.02)	(0.01)	(0.02)
Voted for Biden	0.55	0.87	0.09
	(0.02)	(0.01)	(0.01)
Observations	993	577	416

Notes: This table presents summary statistics for our main sample. Column (1) shows the overall sample; columns (2) and (3) split it along party lines. Standard errors in parentheses.

experiment was programmed using oTree (Chen et al., 2016), with a custom integration to the Daily API that allows us to create pair-specific video-call rooms and store the recordings of these calls.

Preregistration. The main hypotheses, experimental design, and sample size criteria for both the main experiment and the follow-up survey were pre-registered on `AsPredicted.org` (with pre-registration IDs #155100 and #173633, respectively). Pre-registration can be found in the replication files.

2.4. Sample and Reweighting

As shown in Table A.2, attrition was modest and not differential by treatment.[12]

Table 1 presents summary statistics for our main sample. Consistent with known difficulties in recruiting Republicans for online experiments, our sample features around 28 percent fewer Republicans than Democrats (Kashner and Stalinski, 2024). Compared to

[12]Following Lin et al. (2016), we also estimate a model in which our main indicator for attrition is regressed on the treatment indicator, covariates (age, gender, partisan affiliation, and baseline polarization) and the interaction of treatment indicators with covariates. An F-test of joint significance of these interaction terms allows us to test for treatment-by-covariate differences in attrition, for which we find no significant evidence ($p = 0.185$).

Table 2: Balance

	(1) Co	(2) Cross	(3) p-value (2)-(3)	(4) Co (weight)	(5) Cross (weight)	(6) p-value (5)-(6)
Age	41.491	42.539	0.226	42.110	42.539	0.631
	(0.593)	(0.630)		(0.632)	(0.630)	
Female	0.477	0.461	0.599	0.474	0.461	0.685
	(0.022)	(0.023)		(0.023)	(0.023)	
White	0.754	0.783	0.285	0.780	0.783	0.915
	(0.019)	(0.019)		(0.019)	(0.019)	
Black	0.167	0.124	0.054	0.149	0.124	0.250
	(0.017)	(0.015)		(0.016)	(0.015)	
Asian	0.098	0.114	0.431	0.090	0.114	0.217
	(0.013)	(0.014)		(0.013)	(0.014)	
Latino Identity	0.077	0.066	0.521	0.075	0.066	0.605
	(0.012)	(0.011)		(0.012)	(0.011)	
Graduated College	0.220	0.211	0.722	0.206	0.211	0.848
	(0.018)	(0.019)		(0.018)	(0.019)	
Household Income over 50k	0.695	0.715	0.503	0.704	0.715	0.713
	(0.020)	(0.021)		(0.021)	(0.021)	
Urban Residence	0.560	0.519	0.192	0.537	0.519	0.576
	(0.022)	(0.023)		(0.023)	(0.023)	
Republican	0.344	0.498	<0.001	0.500	0.498	0.949
	(0.021)	(0.023)		(0.023)	(0.023)	
Voted for Trump	0.287	0.401	<0.001	0.410	0.401	0.772
	(0.020)	(0.022)		(0.024)	(0.022)	
Voted for Biden	0.597	0.492	0.001	0.474	0.492	0.587
	(0.022)	(0.023)		(0.023)	(0.023)	
Affective Polarization (baseline)	41.552	39.897	0.361	38.989	39.897	0.626
	(1.252)	(1.310)		(1.321)	(1.310)	
Confidence in Initial Quiz	64.129	63.847	0.772	64.649	63.847	0.428
	(0.665)	(0.710)		(0.720)	(0.710)	
Score in Initial Quiz	6.487	6.568	0.668	6.446	6.568	0.527
	(0.129)	(0.138)		(0.135)	(0.138)	
Observations	509	484	993	509	484	993

Notes: This table presents a balance test for participants in our main sample. In columns (4)–(6), Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. Robust standard errors in parentheses.

Democrats, Republicans in our sample are slightly older, more likely to be males, more likely to be white, less likely to have a college degree, and more likely to have a household income above 50 thousand dollars.

Table 2 presents a balance table that compares the characteristics of individuals in the co-partisan and the cross-partisan interaction groups.[13] As shown in the table, the co-partisan group features fewer self-identified Republicans, fewer self-declared Trump voters, and more self-declared Biden voters. This imbalance is mechanical and stems from the interaction of two forces: i) we recruited fewer self-identified Republicans than Democrats as discussed above, and ii) the cross-partisan interaction condition necessarily features an equal number of Republicans and Democrats. As a result of these two forces, there are relatively fewer Republicans than Democrats in the co-partisan group, thus mechanically generating the imbalance shown in the balance table. We address this imbalance by reweighting

[13] Appendix Table A.4 provides balance tests at different stages of the experiment.

observations so as to mimic having an equal number of Democrats and Republicans both in the co-partisan and the cross-partisan interaction groups. We do this for all of our analyses below. Reassuringly, this procedure brings balance also to the individual characteristics that were not directly targeted, such as Black ethnicity and voting behavior, as can be seen from columns 4 through 6 of Table 2.[14]

We note that the imbalance above did not arise from a failure of randomization. Randomly matching individuals coming from a population consisting of two groups of different sizes into pairs is expected to yield such an imbalance. Despite not arising from a failure of randomization, the imbalance would bias the interpretation of the results in the absence of reweighting. Specifically, in the absence of reweighting, the comparison between cross-partisan and co-partisan pairs would disproportionately reflect the comparison between cross-partisan and Democrat-Democrat pairs. For robustness, Appendix Table A.5 reports our main analyses with an alternative (inverse probability) weighting procedure, as well as with the addition of controls.

3. Main results

The results section is organized as follows. In Section 3.1, we leverage our willingness-to-pay measure to document a preference for self-selecting into echo chambers, before exploring its plausible drivers. In Section 3.2, we explore the expected and actual learning in the conversations and decompose learning into the potential for learning and knowledge extraction. Section 3.3 examines the hedonic effects of cross-partisan conversations, including a lasting reduction in affective polarization.

Following our preregistration, our analyses primarily compare cross-partisan with co-partisan interactions. In Appendix C, we revisit our main results looking separately at Republicans and Democrats. The results from the split sample show a pattern similar to the aggregate comparisons.

[14]The reason why our reweighting procedure addresses the small imbalance on Black ethnicity is arguably because Black ethnicity is highly correlated with self-identifying as a Democrat.

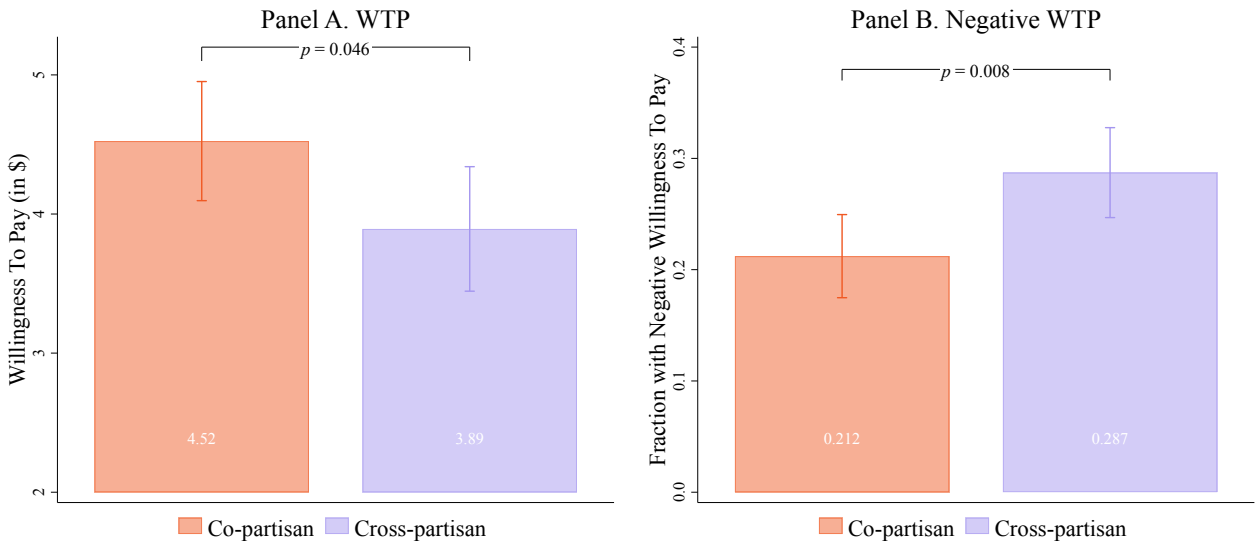

Figure 3: Willingness to Pay

Notes: The figure shows predicted values from regressions where Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals use robust standard errors.

3.1. Self-selection into Echo Chambers

3.1.1. Treatment Differences in Participants' Willingness to Interact

Our first result is that participants are significantly less likely to want to interact with counter-partisans than with co-partisans. Figure 3A shows that the average willingness to pay for interacting with one's partner is significantly lower in cross- than in co-partisan pairs. Moreover, Figure 3B shows that participants in cross-partisan pairs are significantly more likely to have a negative willingness to pay to interact, thus indicating a strict preference against interacting. Appendix Table A.6 shows that our results are robust to different specifications.[15] In a world of constrained time and attention, a greater willingness to interact with members of one's party implies the formation of echo chambers, personal networks or groups in which ideologically like-minded individuals are over-represented.

[15] Appendix Table A.7 uses open-ended responses by participants to show that, across treatments, participants are not differentially apprehensive about technological or other disturbances, differentially suspicious of a conversation actually being implemented, or differentially struggling to understand the willingness-to-pay instructions. The table also suggests that these considerations do not loom large in the participants' minds to begin with.

3.1.2. Barriers to Cross-Partisan Interactions

What motivates the observed gap in willingness to interact? In the context of our study, participants' willingness to pay for the conversation can be driven by both instrumental and hedonic motives. Specifically, the conversation has instrumental value because it might help participants improve their scores on the quiz, and has hedonic value because it can be more or less pleasant, more or less interesting, etc.[16] The fact that $\sim$ 25 percent of participants have a strict preference against interacting is highly suggestive of the presence of hedonic motives (see Figure 3). Specifically, since information can always be ignored, its instrumental value is non-negative, implying that negative willingness to pay has to stem from hedonic factors.[17]

To better understand the drivers of participants' willingness to interact, we leverage data from an open-ended question asking participants for the rationale behind their choices in the willingness-to-pay elicitation. Participants' open-ended responses were hand-coded by a research assistant blind to treatment status. Figure 4A shows the four most frequently mentioned rationales and how they correlate with willingness to pay. The four most frequently mentioned reasons are the desire to improve on the quiz, curiosity, the expectation that the conversation will be enjoyable, and the worry that the conversation will not be enjoyable.[18] These reasons correlate with willingness to pay for the conversation in intuitive ways. The desire to improve on the quiz, curiosity, and the expectation that the conversation will be enjoyable are all associated with a higher willingness to pay for the conversation. Conversely, worrying that the conversation will not be enjoyable is negatively correlated with willingness to pay.

[16]The conversation can also have instrumental value outside the context of our experiment, though such value is likely to be small. Nonetheless, our willingness-to-pay measure is designed to account for this possibility. Since virtually no participant mentions the instrumental value of the conversation outside the experiment when asked to explain the rationale behind her choices in the willingness-to-pay elicitation, we ignore such value in the remainder of the paper.

[17]Such hedonic factors, for instance, include unwillingness to talk to a stranger, unwillingness to discuss politics, etc.

[18]In Figure 4, we include separate indicators for participants mentioning expecting the conversation to be enjoyable and expecting it to be unenjoyable. This allows us to be consistent in the analysis presented in the figure by always constructing our outcome variables as indicators for mentioning a particular issue in the open response.

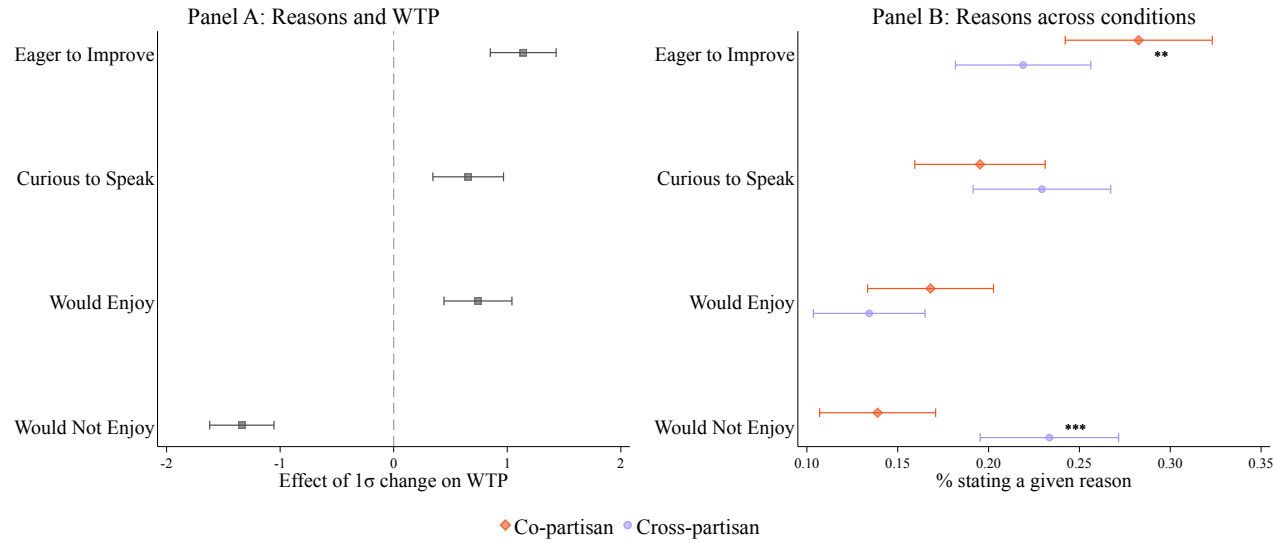

Figure 4: Reasons for WTP Decision

Notes: Panel A shows the coefficients from a regression of WTP (in U.S. dollars) on the four most frequently mentioned rationales for the WTP decision. Panel B shows predicted values from regressing each rationale on a cross-partisan treatment indicator; stars indicate a significant difference between co- and cross-partisan groups. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals use robust standard errors. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Figure 4B shows how the four most frequently mentioned rationales behind the choices in the willingness-to-pay elicitation differ between experimental conditions. Participants poised to participate in cross-partisan interactions are significantly less likely to state that wanting to improve on the quiz factored into their willingness to pay for the conversation and significantly more likely to mention worries about the conversation being unpleasant. There are no statistically significant differences across conditions as to whether participants state that they are curious about the interaction or that they expect to enjoy it.[19]

In summary, participants' willingness to pay reflects both their desire to improve their quiz performance and concerns about not enjoying the conversation. These motives vary systematically between co- and cross-partisan interactions, aligning with the observed differences in willingness to pay.

3.2. The Instrumental Value of Conversations

A detailed investigation of the instrumental motives for selecting into echo chambers is important because, for consequential decisions, these motives might overshadow hedonic factors and be the primary drivers of self-selection into echo chambers.

[19] Appendix Figure A.1 describes less frequently stated considerations about willingness to pay and illustrates how such considerations vary by treatment.

3.2.1. The Gap in Expected Improvement

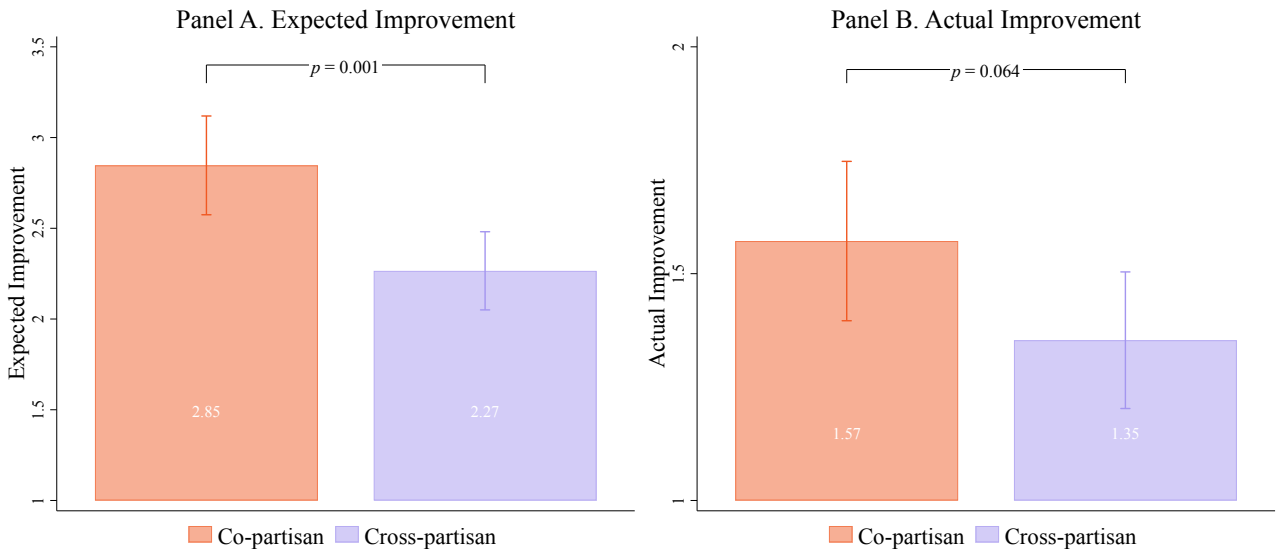

Figure 5: Expected and Actual Improvement

Notes: Panel A shows predicted values from a regression of expected improvement on a cross-partisan treatment indicator. Panel B shows the same for actual improvement. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals use robust standard errors (Panel A) and standard errors clustered at the pair level (Panel B).

We begin by examining respondents' expectations about the extent to which the conversation can help them improve their answers to the quiz. If participants expect to learn less from counter-partisans than from co-partisans, then even absent hedonic factors such as the comfort of interacting with like-minded individuals, the perceived relative instrumental value of co- vs. cross-partisan conversations alone could lead to the formation of echo chambers.

As depicted in Figure 5A, participants expect to improve significantly less when they are assigned to a cross-partisan rather than a co-partisan conversation. The magnitude of the effect is around half a question in the quiz. As shown in Appendix Figure A.2, the conversation itself does not appear to alter these expectations significantly. When we assess participants' expected improvement after the conversation, we find that: a) they continue to expect greater instrumental benefits from co-partisan interactions compared to cross-partisan ones ($p = 0.015$), and b) the size of the expectation gap does not significantly change ($p = 0.250$). Thus, the conversation itself fails to significantly shift participants' beliefs about its instrumental value.

3.2.2. The Gap in Actual Learning

To what extent are participants' expectations about learning from co- and cross-partisan conversations justified? We begin by analyzing the average differences in actual learning between co-partisan and cross-partisan conversations, where actual learning is defined as the increase in the number of correct answers to the quiz following the conversation. As shown in Panel B of Figure 5, we can be relatively confident that cross-partisan conversations are less informative than co-partisan ones. Specifically, we can reject at the 10 percent level the null hypothesis that co-partisan and cross-partisan conversations lead to the same level of improvement ($p = 0.064$). Even in the relatively unlikely event that cross-partisan conversations lead to larger improvements than co-partisan ones, we can rule out improvements greater than 0.01 correct answers on the quiz at the 5 percent significance level.[20]

These results are qualitatively consistent with participants' expectations that co-partisan conversations are more informative than cross-partisan ones. Quantitatively the expected difference in improvements (0.58) is larger than the actual difference in improvements (0.22) between co- and cross-partisan conversations, albeit not significantly so at conventional levels ($p = 0.371$).

Previous work (e.g. Hobolt et al. 2024) finds that cross-partisan contact leads to a convergence of attitudes and beliefs. In Appendix D we describe a similar convergence result. Relative to co-partisan conversations, cross-partisan conversations in our study eliminate the partisan signature of a participant's profile of answers.[21] However, a strength of our design is that we can study convergence to the truth as a phenomenon that may be distinct from convergence of beliefs between groups. Leveraging this distinction, we see that cross-partisan conversations might be worse at achieving learning than at achieving factual

[20] An improvement of 0.01 correct answers on the quiz translates to less than one cent in monetary terms. Thus, even at the upper bound of the confidence interval for the relative expected improvement from cross-partisan conversations, the monetary gain is so minimal that it would not alter the behavior of any participant in our experiment who, according to her willingness to pay, has a strict preference against interacting with a counter-partisan. In other words, even if participants' beliefs about the relative gains from cross- versus co-partisan conversations were mechanically set at the upper limit of the confidence interval, this marginal improvement would not be sufficient to shift their preference away from self-selecting into echo chambers.

[21] In particular, we are able to correctly predict someone's party affiliation by looking at their answers 63 percent of the time. After engaging in a cross-partisan conversation, this fraction drops to 47 percent. Moreover, this factual depolarization persists in the follow-up survey.

depolarization or convergence.

3.2.3. Potential for Learning and Knowledge Extraction

To understand differences in learning from co-partisans and counter-partisans more deeply, we can decompose actual improvement into two key components: *potential for learning* and *difficulties in knowledge extraction*. We define potential for learning as the number of quiz questions for which a participant's partner knows the correct answer and the participant does not. We define difficulties in knowledge extraction as a participant's inability to correctly revise the answer to a question, conditional on her partner having the correct answer.

The decomposition of actual improvement into potential improvement and difficulties in knowledge extraction is helpful for shedding light on the reasons why participants are able to learn more from co-partisans than from counter-partisans. The potential for learning reflects differential knowledge distributions between co-partisan and cross-partisan pairs. Difficulties in knowledge extraction might stem, for instance, from a lack of trust in the credibility of the other side of the political aisle, a tendency of cross-partisan discussions to focus on unproductive questions, or a propensity of such conversations to become heated and hostile.

Formally, the decomposition works as follows. Let $improvement_{i,j}$ denote participant i's improvement on the quiz as a result of having a conversation with participant j. Let $potential_{i,j}$ denote the number of questions that participant i answered incorrectly in the Initial Quiz and participant j answered correctly. By the law of total expectation, we have

$$E(improvement_{i,j}) = \sum_{k=0}^{14} E(improvement_{i,j}|potential_{i,j} = k)P(potential_{i,j} = k)$$

where the expectations indicate population-level quantities.

In order to produce one summary measure of perceived difficulties in knowledge extraction, we impose the assumption that, conditional on player j correctly answering k questions in the Initial Quiz that player i answered incorrectly, player i improves her score on the Revised Quiz by, on average, βk questions. We can thus interpret $\beta \in [0, 1]$ as parametrizing the perceived ease of knowledge extraction, and we can let it vary by treat-

ment $t \in \{co, cross\}$. The assumption allows us to rewrite expected improvement as:

$$E(improvement_{i,j}) = \beta_t E(potential_{i,j}) \tag{3.1}$$

By conditioning the expectations in the expression above on whether participant i is poised to have a conversation with a co- or counter-partisan, we obtain four measures: two measures of potential improvement (one from co- and one from counter-partisans), and two measures of ease of knowledge extraction (again, one from co- and one from counter-partisans).

Panels A and B of Figure 6 show that impairments to learning in cross-partisan interactions are primarily due to difficulties in knowledge extraction. Specifically, Panel A of Figure 6 shows that potential improvement is, if anything, larger in cross-partisan conversations than in co-partisan ones. Appendix Figure A.3 shows that this is because knowledge is distributed across party lines, with Democrats and Republicans being equally informed on average, but differentially informed across quiz questions. Appendix Figure A.3 also shows that while both parties are equally informed, our participants are much more pessimistic about the knowledge of counter-partisans than they are about the knowledge of co-partisans.

The fact that, despite the greater potential for learning in cross-partisan conversations, participants tend to learn less in cross-partisan than in co-partisan interactions must then be due to greater difficulties in knowledge extraction when meeting counter-partisans. Indeed, Panel B of Figure 6 shows that, for every question that a participant's conversation partner answers correctly and the participant does not, the participant's score improves by 17 percent less when the partner is a counter-partisan rather than a co-partisan.

Our definition of potential improvement and our decomposition of learning into potential improvement and knowledge extraction make two implicit assumptions. The first assumption is that the most valuable learning opportunities are instances where exactly one of the two conversation partners has the correct answer, as opposed to instances where neither partner has the correct answer. The second assumption is that unlearning, defined as transitioning from having a correct answer to having an incorrect one as a result of the conversation, is sufficiently uncommon as to be negligible.

Appendix Table A.8 provides support for both assumptions. First, instances in which

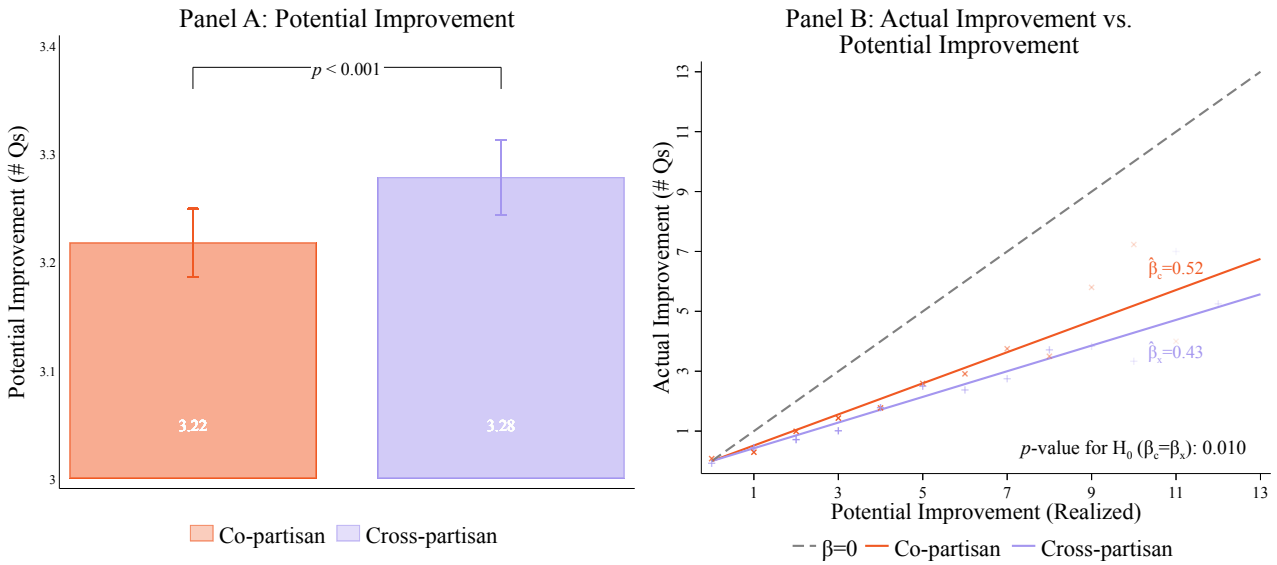

Figure 6: Actual and Potential Improvement

Notes: Panel A shows potential improvement by treatment. Potential improvement is constructed as follows. First, we select two parties i and j from the set $\{Democrat, Republican\}$. Second, for each question, we calculate the potential improvement for the average affiliate of party i meeting an average affiliate of party j as the sample-level probability that the affiliate of party i does not know the answer times the probability that the affiliate of party j does. We then sum these question-level statistics at the quiz level and average across cross- and co-partisan pairs to produce the bars. In Panel B, we relate potential improvement at the level of individual pairs to actual improvement in those pairs. We refer to our measure of potential improvement in Panel B as potential improvement (realized), because it refers to the number of questions, for each participant and for the partner she is randomly matched to in conversation, where the participant gave an incorrect answer and her partner gave the correct one. Panel B presents regressing actual improvement for each participant on potential improvement (realized), in a model where the intercept is fixed at 0 as required by Equation 3.1. In the regression, potential improvement (realized) is interacted with treatment assignment. β_c and β_x represent the estimated ease of knowledge extraction for the co- and cross-partisan treatments respectively. In the overlayed scatterplot of Panel B, opacity is proportional to the square root of the number of observations in each bin. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. All outcomes are in number of correct quiz answers (out of 14). The 95 percent confidence intervals and p-value use bootstrapped standard errors (Panel A) and standard errors clustered at the pair level (Panel B).

conversation partners transition from one correct answer between them to two correct answers are 4 to 5 times more common than instances in which the partners transition from zero correct answers between them to one correct answer.[22] Moreover, transitioning from one correct answer between the two partners to two correct answers is 16 times more likely than transitioning from zero correct answers to two correct answers. Second, we do not find evidence that unlearning plays a significant role in our experiment. Specifically, the fraction of instances in which at least one partner transitions from a correct answer to an incorrect one is less than 6 percent.[23]

Mirroring our results on treatment differences in knowledge extraction, Appendix Ta-

[22]Here an "instance" is a particular question in which a particular participant pair transitions from one knowledge state to another.

[23]These results suggest that people who give the correct answer in the Initial Quiz are not simply guessing randomly. In fact, the results presented in Panel B of Figure 6 are robust to different specifications of potential learning in which one's conversation partner is considered to know the correct answer only if she also reports a high level of confidence in her answer (see Appendix Figure A.4).

ble A.8 also shows treatment differences in exactly two of the nine possible pairwise transitions: cross-partisan conversations are significantly less likely to feature transitions from one correct to two correct answers and they are significantly more likely to lead to stagnant transitions from one correct answer to one correct answer.

3.2.4. What Drives Treatment Differences in Knowledge Extraction?

We investigate two broad candidate explanations for why it may be harder to extract knowledge from counter-partisans. The first is that cross-partisan conversations may be less conducive to information sharing, for example because information is communicated less effectively, because participants focus on less productive topics, or because conversations become more heated. The second candidate explanation is that prior prejudice about, or lack of trust in, a counter-partisan's knowledge impedes learning and cannot be overcome during the conversation.

To study differences in the conversations themselves, we leverage a dataset consisting of the video recordings of the conversations. In particular, we instructed research assistants, who were blind to treatment status, to watch each conversation and to code it according to a pre-specified codebook. Furthermore, we supplemented these hand-coded measures with machine-extracted audiovisual measures. The human-coded variables capture interpersonal and communicative behavior, including qualities like openness to listening, friendliness, assertiveness, and the type of justification offered for answers. The machine-coded variables are derived algorithmically from video and audio using computer vision and speech models, extracting nonverbal signals such as body posture, facial action, gaze direction, and predicted emotional states such as happiness, contempt, and anger from voice (see Appendix Table B.1 for the full list of variables).

To get further traction on the conversations themselves, we also transcribed them and represented each transcript using a high-dimensional text-embedding model for general-purpose retrieval (`voyage-4-large`). The embedding model maps text into a 1,024-dimensional vector that summarizes its semantic content, with the aim of capturing subtle features of meaning and discourse that may be missed by our codebook.

To study the second candidate explanation, which relies on participants' expectations and interpretations of the conversations, we leverage variables elicited from participants

before and after the conversation.

Differences in the conversation. We begin by looking for differences between co- and cross-partisan conversations that may plausibly drive differences in knowledge extraction. Figure A.6 shows that, if anything, cross-partisan pairs spend more time on productive questions, defined as questions for which exactly one of the two partners has the correct answer. Moreover, the answers participants communicate in both co- and cross-partisan conversations have above-90-percent fidelity to their actual answers, are expressed with similar confidence, and are supported by similar levels of justification. Thus, there appear to be no differences in the transmission of information that could rationalize differences in knowledge extraction.

More generally, observable differences between co- and cross-partisan conversations are barely detectable. In Appendix Figure A.7, we check for treatment differences in the broader set of variables coded by the research assistants and in the audio-visual variables we extracted. Overall, the conversations look strikingly similar. The only exception to this rule is that cross-partisan conversations are more likely to involve explicit rejection of the partner's answer.

Of course, our hand-coding summarizes only part of the information contained in a conversation. We therefore turn to the full transcripts and ask whether an embedding-based classifier can distinguish cross-partisan from co-partisan conversations. The classifier performs better than chance ($p = 0.04$), indicating that treatment status leaves a detectable signature in the language of the conversations. However, the magnitude of this signal is small: the out-of-sample AUC is only 0.559 (see Appendix Table B.3).[24] Thus, the transcripts contain a real but substantively limited treatment signal. Co- and cross-partisan conversations are somewhat distinguishable based on their transcripts, but far from cleanly separable.

To understand what the embedding-based classifier is detecting, we relate the transcript embeddings back to the hand-coded features. Only one feature correlates significantly with the embeddings: the *explicit rejection of the partner's answer* (see Appendix Figure B.2). This suggests that cross-partisan conversations are somewhat more likely to involve

[24]The AUC, or area under the ROC curve, measures how well a classifier separates two groups across all possible classification thresholds. An AUC of 0.5 corresponds to chance performance; an AUC of 1 corresponds to perfect separation.

direct pushback, but are void of other significant differences.

Differences in participants' trust. We next turn to the role of expectations and interpretations in shaping knowledge extraction. We have already seen that, although Democrats and Republicans are equally knowledgeable on average, both sides believe the other side to be less knowledgeable and neglect the fact that knowledge is distributed across the aisle (see Appendix Figure A.3). Consistent with this pattern, Appendix Table A.10 shows that, when asked how knowledgeable they deem their partner specifically, participants are substantially more pessimistic about counter-partisans. Moreover, this pessimism persists after the conversation. We also observe similar relative pessimism in an incentivized post-conversation belief about the partner's number of correct answers. Finally, it is noteworthy that the only substantive difference between treatments that we glean from the transcripts is an expression of distrust, in the form of pushback. Suggestive of trust driving differences in knowledge extraction, Appendix Table A.10 shows that lower trust in conversation partners *and* greater pushback during the conversation are predictive of lower levels of knowledge extraction.[25]

Taken together, the evidence suggests that the central obstacle to cross-partisan learning is neither the absence of useful information, nor a breakdown in the conversation itself, but rather a failure to credit counter-partisan information. Participants enter cross-partisan conversations expecting their partners to be less knowledgeable even though Democrats and Republicans are similarly informed. Moreover, even though co- and cross-partisan pairs discuss similarly productive questions, transmit answers with similar fidelity and confidence, justify their answers in similar ways, and look strikingly similar in observable content and tone, the trust gap persists and predicts failures of knowledge extraction.

[25]Only post-conversation measures of trust significantly predict knowledge extraction. Crucially, these measures continue to exhibit lower levels in cross-partisan pairs. The fact that the correlation between trust and knowledge extraction increases and becomes significant after participants experience the conversation suggests that participants learn about their partner's knowledgeability and are better able to predict it ex-post. The fact that the partisan gap in trust persists suggests that this learning is imperfect and does not lead to a sufficient reduction in undue prejudice.

3.3. The Hedonic Value of Conversations

3.3.1. Expected and Realized Hedonic Value

We already saw that, facing cross-partisan conversations, participants were more likely to voice the concern that the conversation would not be enjoyable. Panel A of Figure 7 confirms this result in a test that subsumes positive and negative mentions of predicted enjoyment (as well as the absence of any mentions) in an index.

Panel B of Figure 7 reveals that participants' worries about the relative unpleasantness of cross-partisan contact are largely unwarranted. The figure depicts agreement with the statement "I had a good time during the conversation" on a Likert scale from 1 (strongly disagree) to 4 (strongly agree). We see that the average participant in both conversations experienced conversations as rather pleasant and that participants were equally likely to report having had a good time in co-partisan and cross-partisan conversations.

We note that predictions and self-reports are not incentivized and that they are measured on different scales. Nonetheless, panels A and B of Figure 7 exhibit qualitatively different patterns: the ex-ante difference in the likelihood of voicing concerns about not enjoying the conversation is significantly negative, whereas we see no ex-post difference in self-reported experiences.

3.3.2. Affective Polarization: Short-term Effects

Next, turning to affective polarization, we examine whether cross-partisan conversations improve how participants feel about individuals affiliated with the opposing political party relative to their co-partisans. Given the recent rise in affective polarization and its well-documented negative effects on democratic processes and social cohesion, any reduction in affective polarization carries significant normative weight (Iyengar et al., 2019; Boxell et al., 2024).

Consistent with prior findings on the pronounced levels of affective polarization in the U.S. political landscape (e.g. Druckman and Levy, 2022), our primary measure of affective polarization, the feeling thermometer, illustrates substantial polarization at baseline. Participants rated their feelings toward individuals from the opposing party 39 points colder (on a scale from 0 to 100) compared to those from their own party.

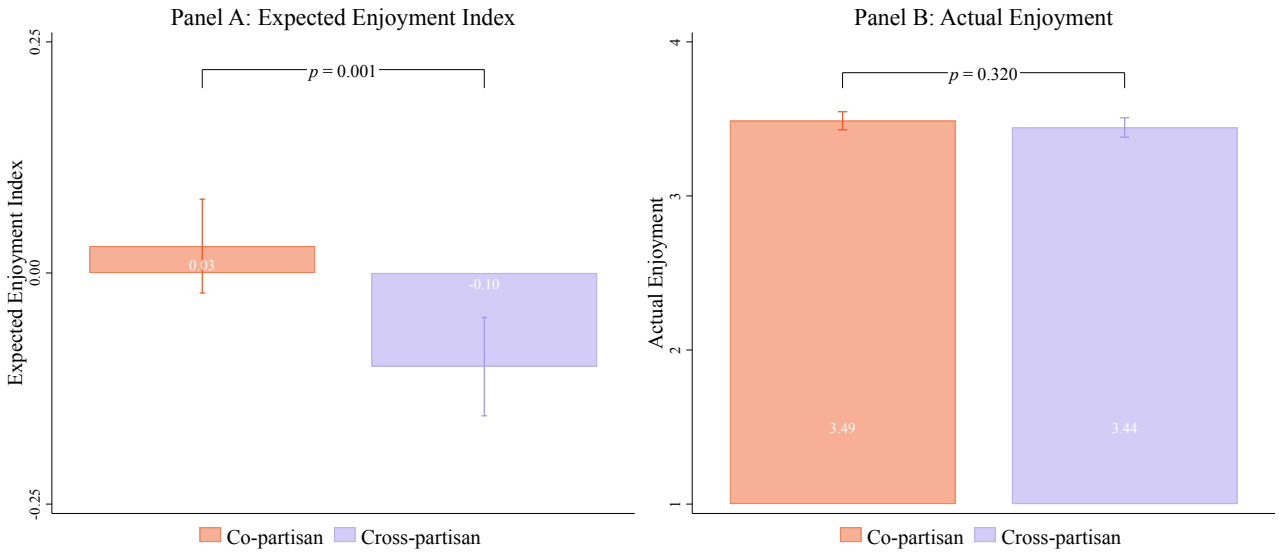

Figure 7: Expected and Actual Enjoyment

Notes: Panel A shows predicted values from a regression of an expected enjoyment index on a cross-partisan treatment indicator. The index is coded on $\{-1, 0, +1\}$: -1 if the participant expects not to enjoy the interaction, $+1$ if they expect to enjoy it, 0 if they did not mention enjoyment in the open-ended explanation of their WTP decision. Panel B shows predicted values from regressing reported enjoyment after the interaction on a cross-partisan treatment indicator; the outcome uses a 1–4 Likert scale (1 = "strongly disagree", 4 = "strongly agree"). Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals and p-values use robust standard errors (Panel A) and standard errors clustered at the pair level (Panel B).

We find that our intervention substantially mitigates this polarization. First, Figure 8A shows, in a before-after comparison, that cross-partisan conversations reduce the feeling thermometer gap by $\sim$ 20 percent, indicating a notable improvement in attitudes toward the opposing party. Second, we employ five distinct measures of affective polarization elicited after the conversations to estimate the differential effects of cross-partisan versus co-partisan interactions. Panel B of Figure 8 demonstrates that cross-partisan conversations produce significant and substantial reductions in affective polarization across all five measures, with effect sizes ranging from 0.14 to 0.33 standard deviation units.

Because we elicit participant's baseline willingness to interact, our study can address a threat to external validity that usually plagues experimental studies on intergroup contact, i.e. that individuals who select into contact might be more receptive to the depolarizing effects of intergroup contact than those who choose to opt-out. We find that the correlation between willingness to interact and affective depolarization is not statistically different from zero ($r = 0.013$, $p = 0.772$), assuaging concerns that the positive effects of cross-partisan contact on affective polarization are more pronounced for individuals who are more willing to engage with counter-partisans.

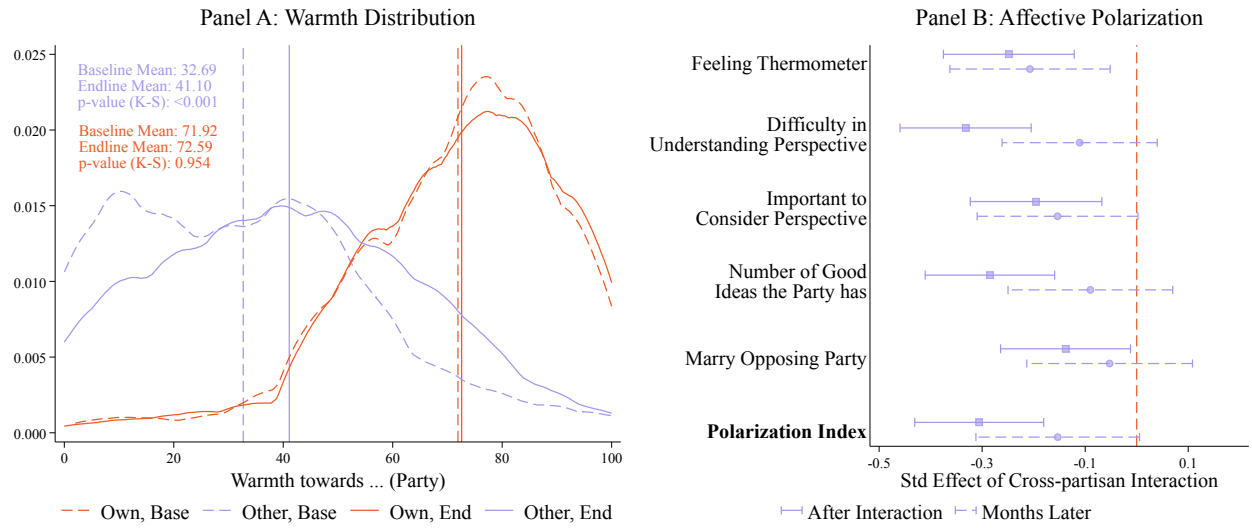

Figure 8: Affective Polarization

Notes: Panel A shows the distributions of baseline and endline warmth toward one's own and the other party, restricted to the cross-partisan treatment group. Kolmogorov-Smirnov test p-value for own-party warmth: 0.954; for other-party warmth: <0.001. Panel B plots coefficients on a cross-partisan treatment indicator from regressions of the 5 outcome variables and an index of affective polarization (construction in Section 2.2), measured both immediately after the interaction and again ∼100 days later. We deviate from the pre-registration by including the "Marry opposing party" variable in the polarization index — a more comprehensive and conservative choice, as discussed in the main text. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals use standard errors clustered at the pair level.

3.3.3. Affective Polarization: Persistence of Treatment Effects

To test for the persistence of the depolarizing effects of cross-partisan conversations, we conducted a follow-up survey approximately 100 days after the main study.[26] As previously discussed, the study was fielded and designed to appear unrelated to the original experiment, minimizing the chance that participants associate the follow-up survey with the initial study and give rise to experimenter demand effects.

The primary outcome measure specified in our pre-analysis plan, the feeling thermometer, indicates that the treatment effect of cross-partisan interactions remains positive and statistically significant in the follow-up survey ($p = 0.009$). This long-term effect is 92 percent of the immediate effect, and the difference between the immediate and follow-up effects is not statistically significant at conventional levels ($p = 0.541$). Figure 8B shows that the persistence of the effects varies across different outcomes, but that the overall pattern remains consistent. Specifically, on a pre-specified equally weighted index of our five standardized affective polarization measures, we observe a 0.15 standard deviation reduction in polarization in the follow-up survey ($p = 0.058$) when comparing participants in

[26] Appendix Table A.2 shows no differential response rates across treatments.

cross-partisan pairs to those in co-partisan pairs.[27]

For additional context, Appendix E presents a meta-analysis that compares our results to those in the experimental literature on partisan contact.[28] Our study is among the largest and one of a few that involve an obfuscated follow-up study to measure the persistence of treatment effects (see Appendix Figure E.1). Our standardized short-run treatment effects are broadly comparable to the meta-analytic effects of interpersonal contact documented in the literature.

3.3.4. What drives treatment differences in affective polarization?

As we have shown, participants expect cross-partisan conversations to be less enjoyable than co-partisan ones. Ex post, however, participants report that cross-partisan conversations are just as pleasant as co-partisan conversations. This pattern mirrors the evidence from the audiovisual data: along a wide range of observable dimensions, co- and cross-partisan conversations look strikingly similar (see Appendix Figure A.7). But unlike in the case of knowledge extraction, participants appear to draw a positive lesson from this similarity. They enter cross-partisan conversations with relatively pessimistic expectations, are positively surprised by the interaction, and update their views of their conversation partner's group.

Consistent with this mechanism, Appendix Table A.9 shows that depolarization is substantially larger among participants who were more affectively polarized at baseline.[29] These are precisely the participants for whom a pleasant cross-partisan conversation should be most surprising.

---

[27] In the pre-analysis plan we indicated that this index would exclude the fifth polarization outcome, already found to be less malleable in Levy et al. (2022). If we followed the pre-analysis plan and excluded this outcome, the gap at follow-up would be larger (0.17 standard deviations) and significant at the 5 percent level ($p = 0.040$). On a separate note, Appendix Figure A.5 shows qualitatively similar results when restricting the analysis to participants observed both in the main study and the follow-up survey.

[28] The literature on contact has traditionally emphasized interactions between different racial, ethnic, or social groups. See Paluck et al. (2021) for an older, comprehensive survey of this literature and Lowe (2024) for a selective survey of studies that had been pre-registered on the AEA-RCT Registry and the EGAP Registry.

[29] This finding cannot be the result of mean reversion, since mean reversion should apply equally in co-partisan and cross-partisan conversations.

3.3.5. What features of conversations are depolarizing?

We conclude this section by asking which features of cross-partisan conversations are particularly depolarizing. This question is of interest to the designers of information environments. Compared to previous work, our study is uniquely situated to answer it because of our rich audio-visual data and conversation transcripts. Appendix Table A.9 reports LASSO regressions using the full set of hand- and machine-coded features of the conversations. Emotional engagement emerges as the key predictor of depolarization in cross-partisan conversations. This variable captures RA-coded measures of whether a participant shared something personal, expressed their feelings, or validated their partner's feelings. Thus, conversations that feature greater intimacy and bonding affect not only participants' views of their conversation partner, but also their views of their partner's party.

This pattern is closely related to mechanisms emphasized in the contact-hypothesis literature. Classic accounts argue that intergroup contact is most likely to reduce prejudice when it creates conditions for meaningful, cooperative, and equal-status interaction (Allport, 1954). Subsequent work emphasizes that contact can reduce prejudice by lowering intergroup anxiety and increasing empathy and perspective-taking (Pettigrew, 1998; Pettigrew and Tropp, 2006b). Using rich process data, we show that these mechanisms are pivotal. In particular, our measure of emotional engagement captures both vulnerability, which is indicative of lower anxiety, and validation of the other person, which is indicative of greater empathy.

In co-partisan conversations, emotional engagement weakly predicts increased polarization ($r = 0.082$, $p = 0.067$).[30] This pattern is also intuitive: in these conversations, bonding occurs with a member of one's own party and increases warmth toward the ingroup. A corollary of our two results is that echo chambers are especially detrimental to social cohesion when there is scope for friendly and emotionally engaged conversations across party lines.

Taken together, the results in this section suggest that the effects of cross-partisan conversations on affective polarization are substantial and largely persistent. The follow-up study also helps abate potential concerns about experimenter demand, as the effects are

[30]This effect is driven by increased warmth towards co-partisans ($p = 0.026$), and not by changes in warmth towards cross-partisans ($p = 0.545$).

present in an obfuscated and seemingly unrelated elicitation. Finally, our results suggest that contact depolarizes through a form of positive surprise, with more emotionally engaged cross-partisan conversations leading to greater depolarization.

4. Conclusion

Our experiment delivers new evidence on the drivers and consequences of co- and cross-partisan conversations centered on facts about politics. We identify a preference for co-partisan over cross-partisan conversations that might contribute to the prevalence of echo chambers even where geographical sorting offline and algorithmic sorting online are not at play. Our finding that both hedonic and instrumental motives play a role in driving the preference for co-partisan interactions likely makes self-selection into such interactions more robust to the exact nature and objective of the interaction than it would otherwise be.

A nuanced picture emerges when we look at the consequences of cross-partisan contact. On information aggregation, we document significant difficulties in extracting knowledge from counter-partisans: our participants struggle to harness the benefits of knowledge that is distributed across the aisle. As a result, policies that encourage cross-partisan interactions may not significantly enhance cross-partisan information sharing. Because the conversations do little to change broadly correct beliefs about overall learning and mistaken beliefs about the informedness of counter-partisans, policies that induce cross-partisan interactions are also not likely to reduce future self-selection on informational grounds. At the same time, our results suggest the intriguing possibility that policies that correct prejudice about the other side's informedness can both increase the willingness to engage in cross-partisan contact and improve how much individuals learn from cross-partisan contact. Future work could put the potential of such policies to foster cross-partisan learning to an experimental test.

On social cohesion, our findings suggest that encouraging cross-partisan interactions may durably reduce affective polarization. This stands in contrast to Santoro and Broockman (2022), who find that conversations about political ideology broadly do not depolarize. Our results suggest that neither avoiding disagreement nor avoiding politics is a necessary condition for cross-partisan contact to work. What matters is positive emotional engage-

ment, and conversations about politics can readily provide it.

References

Acemoglu, Daron, Cevat Giray Aksoy, Ceren Baysan, Carlos Molina, and Gamze Zeki, "Misperceptions and Demand for Democracy under Authoritarianism," Technical Report, National Bureau of Economic Research 2024.

Allport, Gordon Willard, *The Nature of Prejudice*, Addison-Wesley, 1954.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson, "The Diffusion of Microfinance," *Science*, 2013, *341* (6144), 1236498.

_ , Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson, "Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials," *The Review of Economic Studies*, 02 2019, *86* (6), 2453–2490.

Banerjee, Abhijit V., "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 1992, *107* (3), 797–817.

Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya, "Facts, alternative facts, and fact checking in times of post-truth politics," *Journal of Public Economics*, 2020, *182*, 104123.

Bauer, Kevin, Yan Chen, Florian Hett, and Michael Kosfeld, "Group identity and belief formation: a decomposition of political polarization," 2023.

Bazzi, Samuel, Arya Gaduh, Alexander D. Rothenberg, and Maisy Wong, "Unity in Diversity? How Intergroup Contact Can Foster Nation Building," *American Economic Review*, November 2019, *109* (11), 3978–4025.

Becker, Gordon M., Morris H. Degroot, and Jacob Marschak, "Measuring utility by a single-response sequential method," *Behavioral Science*, 1964, *9* (3), 226–232.

Beknazar-Yuzbashev, George, Rafael Jiménez-Durán, Jesse McCrosky, and Mateusz Stalinski, "Toxic content and user engagement on social media: Evidence from a field experiment," *CESifo Working Paper*, 2025.

Belot, Michèle and Guglielmo Briscese, *Bridging America's Divide on Abortion, Guns and Immigration: An Experimental Study*, Centre for Economic Policy Research, 2022.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 1992, *100* (5), 992–1026.

Bishop, B. and R.G. Cushing, *The Big Sort: Why the Clustering of Like-minded America is Tearing Us Apart*, Houghton Mifflin, 2008.

Blattner, Adrian and Martin Koenen, "Does Contact Reduce Affective Polarization? Field Evidence from Germany," SSRN Electronic Journal July 2023.

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles, "Empathy or Antipathy? The Impact of Diversity," *American Economic Review*, December 2006, *96* (5), 1890–1905.

Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro, "Cross-country trends in affective polarization," *Review of Economics and Statistics*, 2024, *106* (2), 557–565.

Braghieri, Luca, "Political Correctness, Social Image, and Information Transmission," *American Economic Review*, December 2024, *114* (12), 3877–3904.

_ , Sarah Eichmeyer, Ro'ee Levy, Markus Möbius, Jacob Steinhardt, and Ruiqi Zhong, "Article-Level Slant and Polarization of News Consumption on Social Media," *Working Paper*, 2024.

Brown, Jacob, Enrico Cantoni, Ryan Enos, Vincent Pons, and Emilie Sartre, "The Increase in Partisan Segregation in the United States," *Working Paper*, 2024.

Burnitt, Christopher, Jared Gars, and Mateusz Stalinski, "Politics of Food: An Experiment on Trust in Expert Regulation and Economic Costs of Political Polarization," 2024. Mimeo.

Chandrasekhar, Arun G, Esther Duflo, Michael Kremer, João F. Pugliese, Jonathan Robinson, and Frank Schilbach, "Blue Spoons: Sparking Communication About Appropriate Technology Use," Working Paper 30423, National Bureau of Economic Research September 2022.

Chen, Daniel L, Martin Schonger, and Chris Wickens, "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

Chopra, Felix, Ingar Haaland, and Christopher Roth, "The demand for news: Accuracy concerns versus belief confirmation motives," *The Economic Journal*, 2024, *134* (661), 1806–1834.

Conlon, John J, Malavika Mani, Gautam Rao, Matthew W Ridley, and Frank Schilbach, "Learning in the Household," Technical Report, National Bureau of Economic Research 2021.

Corno, Lucia, Eliana La Ferrara, and Justine Burns, "Interaction, Stereotypes, and Performance: Evidence from South Africa," *American Economic Review*, December 2022, *112* (12), 3848–3875.

Dahl, Gordon B, Andreas Kotsadam, and Dan-Olof Rooth, "Does integration change gender attitudes? The effect of randomly assigning women to traditionally male teams," *The Quarterly Journal of Economics*, 2021, *136* (2), 987–1030.

Danz, David, Lise Vesterlund, and Alistair J Wilson, "Belief elicitation and behavioral incentive compatibility," *American Economic Review*, 2022, *112* (9), 2851–2883.

DeGroot, Morris H., "Reaching a Consensus," *Journal of the American Statistical Association*, 1974, *69* (345), 118–121.

Druckman, James N and Jeremy Levy, "Affective polarization in the American public," in "Handbook on politics and public opinion," Edward Elgar Publishing, 2022, pp. 257–270.

Dustmann, Christian, Kristine Vasiljeva, and Anna Piil Damm, "Refugee Migration and Electoral Outcomes," *The Review of Economic Studies*, October 2019, *86* (5), 2035–2091.

Enos, Ryan D., "Causal effect of intergroup contact on exclusionary attitudes," *Proceedings of the National Academy of Sciences*, March 2014, *111* (10), 3699–3704.

Fang, Ximeng, Sven Heuser, and Lasse S Stötzer, "How in-person conversations shape political polarization: Quasi-experimental evidence from a nationwide initiative," *Journal of Public Economics*, 2025, *242*, 105309.

Fehr, Dietmar, Johanna Mollerstrom, and Ricardo Perez-Truglia, "Listen to her: Gender differences in information diffusion within the household," *Journal of Public Economics*, 2024, *239*, 105213.

Flaxman, Seth, Sharad Goel, and Justin M. Rao, "Filter Bubbles, Echo Chambers, and Online News Consumption," *Public Opinion Quarterly*, 2016, *80*, 298–320.

Garcia-Hombrados, Jorge, Marcel Jansen, Ángel Martínez, Berkay Özcan, Pedro Rey-Biel, and Antonio Roldán-Monés, "Ideological Alignment and Evidence-Based Policy Adoption," 2024.

Gentzkow, Matthew and Jesse M. Shapiro, "Ideological Segregation Online and Offline *," *The Quarterly Journal of Economics*, 11 2011, *126* (4), 1799–1839.

Golub, Benjamin and Evan Sadler, "Learning in Social Networks," in "The Oxford Handbook of the Economics of Networks," Oxford University Press, 04 2016.

_ **and Matthew O. Jackson**, "Naïve Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, February 2010, *2* (1), 112–49.

González-Bailón, Sandra, David Lazer, Pablo Barberá, Meiqing Zhang, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Gentzkow, Andrew M. Guess, Shanto Iyengar, Young Mie Kim, Neil Malhotra, Devra Moehler, Brendan Nyhan, Jennifer Pan, Carlos Velasco Rivera, Jaime Settle, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker, "Asymmetric ideological segregation in exposure to political news on Facebook," *Science*, 2023, *381* (6656), 392–398.

Graeber, Thomas, Shakked Noy, and Christopher Roth, "Lost in Transmission," CESifo Working Paper 10903 2024.

Guess, Andrew, Benjamin Lyons, Brendan Nyhan, and Jason Reifler, "Avoiding the echo chamber about echo chambers: Why selective exposure to like-minded political news is less prevalent than you think," White Paper, Knight Foundation 2018.

Guess, Andrew M., "Almost Everything in Moderation," *American Journal of Political Science*, 2021, *4* (65), 1007–1022.

Guriev, Sergei, Emeric Henry, Théo Marquis, and Ekaterina Zhuravskaya, "Curtailing false news, amplifying truth," *Amplifying Truth (October 29, 2023)*, 2023.

Haaland, Ingar and Christopher Roth, "Labor market concerns and support for immigration," *Journal of Public Economics*, 2020, *191*, 104256.

_ , _ , Stefanie Stantcheva, and Johannes Wohlfart, "Understanding economic behavior using open-ended survey data," *Journal of Economic Literature*, 2025, *63* (4), 1244–1280.

Henry, Emeric, Ekaterina Zhuravskaya, and Sergei Guriev, "Checking and Sharing Alt-Facts," *American Economic Journal: Economic Policy*, 2022, *14* (3), 55–86.

Hobolt, Sara B., Katharina Lawall, and James Tilley, "The Polarizing Effect of Partisan Echo Chambers," *American Political Science Review*, August 2024, *118* (3), 1464–1479.

Hossain, Tanjim and Ryo Okui, "The binarized scoring rule," *Review of Economic Studies*, 2013, *80* (3), 984–1001.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J Westwood, "The origins and consequences of affective polarization in the United States," *Annual Review of Political Science*, 2019, *22* (1), 129–146.

Jackson, Matthew O. and Leeat Yariv, "Diffusion of Behavior and Equilibrium Properties in Network Games," *American Economic Review*, May 2007, *97* (2), 92–98.

Jo, Donghee, "Better the devil you know: An online field experiment on news consumption," *documento de trabajo de la Universidad Northeastern, consultado el*, 2017, *20*.

Kashner, Daniel and Mateusz Stalinski, "Preempting polarization: An experiment on opinion formation," *Journal of Public Economics*, 2024, *234*, 105122.

Levy, Jonathan, Moran Influs, Shafiq Masalha, Abraham Goldstein, and Ruth Feldman, "Dialogue intervention for youth amidst intractable conflict attenuates neural prejudice response and promotes adults' peacemaking," *PNAS Nexus*, November 2022, *1* (5), pgac236.

Levy, Ro'ee, "Social media, news consumption, and polarization: Evidence from a field experiment," *American Economic Review*, 2021, *111* (3), 831–870.

Lin, Winston, Donald P. Green, and Alexander Coppock, "Standard Operating Procedures for Don Green's Lab at Columbia," 2016.

Lowe, Matt, "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration," *American Economic Review*, 2021, *111* (6), 1807–1844.

_ , "Has Intergroup Contact Delivered?," 2024. Mimeo.

Morris, Stephen, "Political Correctness," *Journal of Political Economy*, 2001, *109* (2), 231–265.

Mousa, Salma, "Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq," *Science*, August 2020, *369* (6505), 866–870.

Nelson, Jacob L. and James G. Webster, "The myth of partisan selective exposure: A portrait of the online political news audience," *Social Media + Society*, 2017, *3* (3).

Ortoleva, Pietro and Erik Snowberg, "Overconfidence in political behavior," *American Economic Review*, 2015, *105* (2), 504–535.

Paluck, Elizabeth Levy, Roni Porat, Chelsey S Clark, and Donald P Green, "Prejudice reduction: Progress and challenges," *Annual Review of Psychology*, 2021, *72*, 533–560.

_ **, Seth A. Green, and Donald P. Green**, "The contact hypothesis re-evaluated," *Behavioural Public Policy*, November 2019, *3* (2), 129–158.

Perego, Jacopo and Sevgi Yuksel, "Media competition and social disagreement," *Econometrica*, 2022, *90* (1), 223–265.

Pettigrew, Thomas F., "Intergroup Contact Theory," *Annual Review of Psychology*, 1998, *49*, 65–85.

_ **and Linda R. Tropp**, "A meta-analytic test of intergroup contact theory," *Journal of Personality and Social Psychology*, 2006, *90* (5), 751–783.

_ **and** _ , "A Meta-Analytic Test of Intergroup Contact Theory," *Journal of Personality and Social Psychology*, 2006, *90* (5), 751–783.

Rao, Gautam, "Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools," *American Economic Review*, 2019, *109* (3), 774–809.

Robbett, Andrea, Lily Colón, and Peter Hans Matthews, "Partisan political beliefs and social learning," *Journal of Public Economics*, 2023, *220*, 104834.

Rossiter, Erin L., "The Similar and Distinct Effects of Political and Non-Political Conversation on Affective Polarization," January 2023.

_ **and Taylor N. Carlson**, "Cross-Partisan Conversation Reduced Affective Polarization for Republicans and Democrats Even after the Contentious 2020 Election," *The Journal of Politics*, October 2024, *86* (4), 1608–1612.

Santoro, Erik and David E. Broockman, "The promise and pitfalls of cross-partisan conversations for reducing affective polarization: Evidence from randomized experiments," *Science Advances*, 2022, *8* (25), eabn5515.

Scacco, Alexandra and Shana S. Warren, "Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria," *American Political Science Review*, August 2018, *112* (3), 654–677.

Schindler, David and Mark Westcott, "Shocking Racial Attitudes: Black G.I.s in Europe," *The Review of Economic Studies*, January 2021, *88* (1), 489–520.

Sunstein, Cass R, *Republic.com*, Princeton university press, 2001.

_ , *Going to extremes: How like minds unite and divide*, Oxford University Press, 2009.

_ , *# Republic: Divided democracy in the age of social media*, Princeton university press, 2017.

Thaler, Michael, "The fake news effect: Experimentally identifying motivated reasoning using trust in news," *American Economic Journal: Microeconomics*, 2024, *16* (2), 1–38.

Yuksel, Sevgi, "Specialized learning and political polarization," *International Economic Review*, 2022, *63* (1), 457–474.

Online Appendix for "Talking across the Aisle"

Luca Braghieri Peter Schwardmann Egon Tripodi

Instructions and Protocols

You can find a complete set of experimental instructions and other detailed protocols in the replication package, which is available at https://osf.io/r56b8/.

A. Additional Tables and Figures

Table A.1: Quiz Questions and Answers

(1) #	(2) Label	(3) Question	(4) Options
Q1	Inflation	What was the path of inflation over the last three years?	a) It was low throughout b) It was high throughout c) It first increased and then decreased d) It increased throughout e) It decreased throughout
Q2	Declare War	Which part of government has the power to declare war?	a) Congress b) The Senate c) The President d) The Department of Defense e) The Secretary of State
Q3	Spend Least	What does the US government currently spend the least on?	a) National Security b) Healthcare c) Social Security d) Foreign Aid e) Education
Q4	Enforce Laws	Which branch of government is responsible for carrying out and enforcing laws?	a) The Legislative branch b) The Judicial branch c) The Executive branch d) The Deliberative branch e) The Enforcing branch
Q5	Gun Checks	According to a survey of US gun owners, what percentage of guns is obtained without background checks?	a) Exactly zero percent b) Between 0 and 25 percent c) Between 25 and 50 percent d) Between 50 and 75 percent e) More than 75 percent
Q6	House Speaker	Who among the following was never a speaker of the US House of Representatives?	a) Nancy Pelosi b) Mike Johnson c) Mitch McConnell d) Kevin McCarthy e) John Boehner
Q7	Healthcare	Compared to countries like Colombia, Finland and Italy, how much of their GDP do Americans spend on health care?	a) A quarter as much b) Half as much c) About the same d) Twice as much e) Four times as much

Q8	Gun Deaths	What is the biggest contributor to gun-related deaths in the United States?	a) Suicides b) Hunting accidents c) Police shootings d) Murders e) Shooting range accidents
Q9	Trump Vaccines	Which of the following statements best describes Donald Trump's most recent view of vaccines, such as those for measles and COVID?	a) They tend to be ineffective b) They are dangerous and can cause autism c) They are very important and people should get vaccinated d) We need more research on whether certain vaccines work e) Everybody should be forced to get vaccinated
Q10	Deported Most	Which administration deported the most immigrants?	a) George Bush b) Bill Clinton c) George W. Bush d) Barack Obama e) Donald J. Trump
Q11	Biden Police	Which of the following statements best describes Joe Biden's position on policing and the 'defund the police' movement?	a) Biden supports "defund the police" unequivocally b) Biden supports "defund the police" unequivocally, but has not condemned police violence against African Americans c) Biden does not support "defund the police", but has condemned police violence against African Americans d) Biden does not support "defund the police" and has not condemned police violence against African Americans e) Biden has never commented on police violence or the "defund the police" movement
Q12	Food Stamps	What fraction of the US population is on food stamps?	a) Between 0% and 10% b) Between 11% and 20% c) Between 21% and 30% d) Between 31% and 40% e) Between 41% and 50%
Q13	Senate Majority	Who is the current senate majority leader?	a) John Fetterman b) Mitch McConnell c) Paul Ryan d) Newt Gingrich e) Chuck Schumer
Q14	Immigrants	Roughly, how many unauthorized immigrants resided in the US in 2021 (including those who overstayed their visas)?	a) 100 000 b) 500 000 c) 1 million d) 5 million e) 10 million
B1	Lowest Tax	Which of the following countries has the lowest corporate income tax?	a) Switzerland b) United States c) Germany d) Mexico e) Ireland
B2	Fewest Police	Which of the following countries has by far the fewest police officers per capita?	a) France b) Spain c) Russia d) United States e) Argentina
B3	Obama Sec. of State	Who was Secretary of State during Barack Obama's second term as president (2013-2017)?	a) Al Gore b) Hillary Clinton c) John Kerry d) Joe Biden e) Ash Carter

Notes: The correct options are highlighted in green. The questions and options are presented to the participants in the order displayed above. The three questions denoted "B" in Column (1) are surprise bonus questions that were asked in the post-experiment survey to detect participants that use google or AI to find correct answers.

Table A.2: Attrition

	(1) Attrition in Main Experiment	(2) Attrition in Main Experiment	(3) Completed Follow-Up	(4) Completed Follow-Up
Crosspartisan Interaction	0.0389	0.0353	0.0346	0.0274
	(0.0320)	(0.0323)	(0.0302)	(0.0306)
Player Party: Democrat		-0.0247		-0.0465
		(0.0226)		(0.0298)
Sample mean	0.202	0.202	0.687	0.687
Observations	1245	1245	993	993
R^2	0.002	0.003	0.001	0.004

Notes: Columns (1) and (2) report attrition regressions among participants who reached the treatment screen and were not assigned the decider role: 1,245 randomly assigned, 993 completing the main experiment. Columns (3) and (4) report regressions for successfully completing both the main experiment and the follow-up survey (administered ∼98 days later); of the 993, 682 completed the follow-up. The regressions are not reweighted. Standard errors clustered at the pair level in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table A.3: Incentivized Expected Improvement Elicitation

	(1) Improvement	(2) Bias
Incentive for E[Improv]	-0.00344	-0.272
	(0.128)	(0.225)
Sample mean	1.447	1.065
Observations	993	993
R^2	0.000	0.002

Notes: This table reports regressions of actual improvement and estimate bias (expected improvement minus actual improvement) on an indicator for whether the expected-improvement question was incentivized. To check whether incentives induced participants to game the expected-improvement question, we incentivized this question for only half of our participants; they learned of the incentive after stating their expectation but before answering the Revised Quiz. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. Standard errors clustered at the pair level in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table A.4: Balance Robustness

Participants who...	reached random matching,			completed the conversation,			completed follow-up.		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Co (weight)	Cross (weight)	p-value (1)-(2)	Co (weight)	Cross (weight)	p-value (4)-(5)	Co (weight)	Cross (weight)	p-value (7)-(8)
Age	42.485	42.48	0.998	42.107	42.581	0.594	44.275	44.716	0.678
	(0.613)	(0.588)		(0.630)	(0.629)		(0.753)	(0.749)	
Female	0.469	0.448	0.492	0.475	0.458	0.589	0.481	0.437	0.266
	(0.022)	(0.021)		(0.023)	(0.023)		(0.029)	(0.027)	
White	0.794	0.778	0.506	0.781	0.784	0.893	0.798	0.801	0.923
	(0.017)	(0.018)		(0.018)	(0.019)		(0.022)	(0.022)	
Race: Black	0.143	0.132	0.608	0.149	0.123	0.242	0.135	0.114	0.402
	(0.015)	(0.015)		(0.016)	(0.015)		(0.018)	(0.017)	
Race: Asian	0.084	0.110	0.138	0.090	0.113	0.224	0.089	0.114	0.274
	(0.012)	(0.013)		(0.013)	(0.014)		(0.016)	(0.017)	
Latino Identity	0.077	0.062	0.36	0.077	0.066	0.49	0.073	0.047	0.15
	(0.012)	(0.010)		(0.012)	(0.011)		(0.015)	(0.011)	
Graduated College	0.220	0.220	0.995	0.205	0.211	0.808	0.206	0.185	0.495
	(0.018)	(0.018)		(0.018)	(0.019)		(0.022)	(0.021)	
Household Income over 50k	0.713	0.721	0.764	0.705	0.715	0.742	0.708	0.704	0.905
	(0.020)	(0.019)		(0.021)	(0.020)		(0.026)	(0.025)	
Urban Residence	0.545	0.517	0.366	0.535	0.520	0.628	0.526	0.487	0.322
	(0.022)	(0.021)		(0.023)	(0.023)		(0.029)	(0.027)	
Republican	0.518	0.495	0.471	0.501	0.499	0.940	0.519	0.513	0.892
	(0.022)	(0.021)		(0.023)	(0.023)		(0.028)	(0.027)	
Voted for Trump	0.430	0.393	0.221	0.412	0.402	0.770	0.439	0.425	0.736
	(0.023)	(0.021)		(0.024)	(0.022)		(0.029)	(0.027)	
Voted for Biden	0.461	0.495	0.253	0.473	0.491	0.578	0.462	0.478	0.689
	(0.022)	(0.021)		(0.023)	(0.023)		(0.028)	(0.027)	
Affective Polarization (baseline)	39.283	39.769	0.781	38.986	39.967	0.597	39.940	39.370	0.803
	(1.249)	(1.222)		(1.318)	(1.306)		(1.671)	(1.564)	
Confidence in Initial Quiz	64.305	63.792	0.597	64.632	63.935	0.490	65.427	64.931	0.679
	(0.687)	(0.683)		(0.718)	(0.708)		(0.886)	(0.809)	
Score in Initial Quiz	6.478	6.552	0.682	6.448	6.583	0.482	6.474	6.900	0.071*
	(0.127)	(0.129)		(0.135)	(0.137)		(0.166)	(0.167)	
Observations	567	545	1,112	510	487	997	341	341	682

Notes: This table reports covariate balance tests for participants reaching three different stages of the experiment, complementing the attrition results in Appendix Table A.2. Each block reports means and robust standard errors across treatment groups. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. Columns (1)–(2) report participants reaching the random assignment stage: 1,245 randomly assigned, 1,112 reporting all covariates. Columns (4)–(5) report the conversation sample: 999 randomly matched, 993 completed the main survey, 2 excluded for incomplete covariates. Columns (7)–(8) report participants who completed the follow-up survey. Columns (3), (6), and (9) report p-values from Kruskal-Wallis tests. * p<0.1, ** p<0.05, *** p<0.01.

Table A.5: Specification Checks (p-Values)

	(1) Weights only	(2) Weights and controls	(3) Controls only	(4) No weights, no controls	(5) IP weights only	(6) IP weights and controls
WTP (Fig. 3A)	0.046	0.040	0.031	0.063	0.063	0.029
Negative WTP (Fig. 3B)	0.008	0.007	0.005	0.008	0.008	0.004
Expected improvement (Fig. 5A)	0.001	0.001	0.001	0.003	0.003	0.001
Actual improvement (Fig. 5B)	0.064	0.100	0.087	0.101	0.100	0.076
Feeling thermometer (Fig 8B, top row)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Affective polarization index (Fig 8B, top row)	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Feel. therm. at follow-up (Fig 8B, bottom row)	0.009	<0.001	<0.001	<0.001	<0.001	<0.001
Aff. pol. index at follow-up (Fig 8B, bottom row)	0.058	0.007	0.009	0.003	0.003	0.010

Notes: This table shows p-values for our main results under different regression specifications. Rows correspond to outcome variables, with the relevant figure number in parentheses. Column (1) is our main specification, with Republican/Democrat-only pairs reweighted to balance partisan composition across treatment groups. Column (2) adds controls. Column (3) keeps controls but drops the reweighting. Column (4) drops both. Column (5) estimates the average treatment effect via inverse-probability weighting using a probit propensity model; column (6) is the same with controls. Controls: age, gender, partisan affiliation, baseline polarization. Standard errors are robust for outcomes measured before the interaction and clustered at the pair level for outcomes measured during or after.

Table A.6: WTP Robustness Check

	(1) WTP ($)	(2) WTP ($)	(3) WTP ($)	(4) WTP ($)	(5) WTP negative	(6) WTP negative
Cross-partisan	-0.631**	-0.650**	-0.836***	-0.836***	0.0751***	0.0754***
	(0.316)	(0.316)	(0.282)	(0.281)	(0.0280)	(0.0279)
Controls		✓		✓		✓
Sample	Full	Full	WTP < 16	WTP < 16	Full	Full
Sample mean	12.185	12.185	11.026	11.026	0.250	0.250
Observations	993	993	867	867	993	993
R^2	0.004	0.021	0.011	0.030	0.008	0.029

Notes: This table reports regressions of various WTP measures on a cross-partisan treatment indicator. Columns (3) and (4) restrict the sample to participants whose reported WTP was strictly below 16 (the maximum possible value). The "sample mean" in columns (5) and (6) is the fraction of the sample reporting negative WTP. Controls: age, gender, partisan affiliation, baseline polarization. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table A.7: WTP Robustness Check, Based on Open-Ended Responses

	Disturbance Afraid	Suspicious	Complicated
Cross-partisan	-0.00287	0.00825	-0.00789
	(0.00696)	(0.00755)	(0.0106)
Sample mean	12.185	12.185	12.185
Observations	993	993	993
R^2	0.000	0.001	0.001

Notes: This table reports regressions of hand-coded explanations of participants' WTP decisions on a cross-partisan treatment indicator. "Disturbance afraid" means the participant indicated being afraid of being interrupted (e.g., by a pet or a child) during the conversation; "Suspicious" means the participant believed the conversation would not actually happen; "Complicated" means the participant did not understand. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table A.8: Transition Frequencies, in %

	(1) Co-partisan	(2) Cross-partisan	(3) p-value
Improvement			
Both wrong to both right	1.34	1.19	0.638
Both wrong to one right	4.25	4.65	0.473
One wrong to both right	22.38	19.74	0.029
Worsening			
Both right to both wrong	0.11	0.06	0.505
One right to both wrong	5.43	5.87	0.477
Both right to one wrong	1.11	1.42	0.249
No change			
Both right to both right	22.11	22.18	0.972
Both wrong to both wrong	25.54	24.04	0.302
One right to one right	17.74	20.85	0.011
Observations	7097	6748	
Participants	509	484	

Notes: Columns (1) and (2) report relative frequencies of transition types between the Initial and Revised Quiz; Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The unit of observation is the individual-question level. Column (3) reports p-values from regressions of each transition type on cross-partisan treatment, with the same reweighting and standard errors clustered at the pair level.

Table A.9: Depolarization Predictors and Baseline-to-Posterior Affective Polarization

Panel A. Cross-partisan depolarization			Panel B. Posterior affective polarization gap	
	(1)	(2)		(1)
	β_{LASSO}	β_{OLS}		
Emotional engagement	0.109	0.177***	Cross-partisan	-4.530***
		(0.045)		(1.166)
Interim: partner knew more	0.015	0.078	Prior affective gap	0.916***
		(0.053)		(0.017)
Partner speech on chitchat	0.008	0.059	Cross-partisan × prior affective gap	-0.090***
		(0.063)		(0.029)
Social potency	-0.005	-0.082*		
		(0.044)		
Speech on own correct questions	-0.003	-0.048		
		(0.048)		
Observations		450		993
R^2		0.063		0.792

Notes: Panel A shows the five largest absolute nonzero coefficients selected by LASSO for cross-partisan depolarization, followed by post-LASSO OLS estimates on the selected predictors. Depolarization is coded as the change from prior to posterior affective polarization gap, so positive coefficients indicate greater depolarization. Panel B shows a full-sample weighted OLS regression of posterior affective polarization gap on treatment, prior gap, and their interaction. Standard errors clustered at the pair level in parentheses below each OLS coefficient. * p<0.1, ** p<0.05, *** p<0.01.

Table A.10: Improvement Predictors: Treatment and Interaction Effects

	(1) Treatment Effects	(2) Interaction Effects
Pre-conversation: Partner knows more	-0.176***	0.040
	(0.043)	(0.039)
During conversation: Rejection count	0.128**	-0.042**
	(0.065)	(0.021)
Post-conversation belief about partner score	-0.799***	0.026**
	(0.167)	(0.013)
Post-conversation: Partner knew more	-0.136***	0.132***
	(0.043)	(0.036)

Notes: Column (1) reports the cross-partisan treatment effect on each row variable. Column (2) reports the interaction-term coefficient in separate regressions of actual improvement on each row variable, interacted with potential improvement. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. Standard errors clustered at the pair level in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

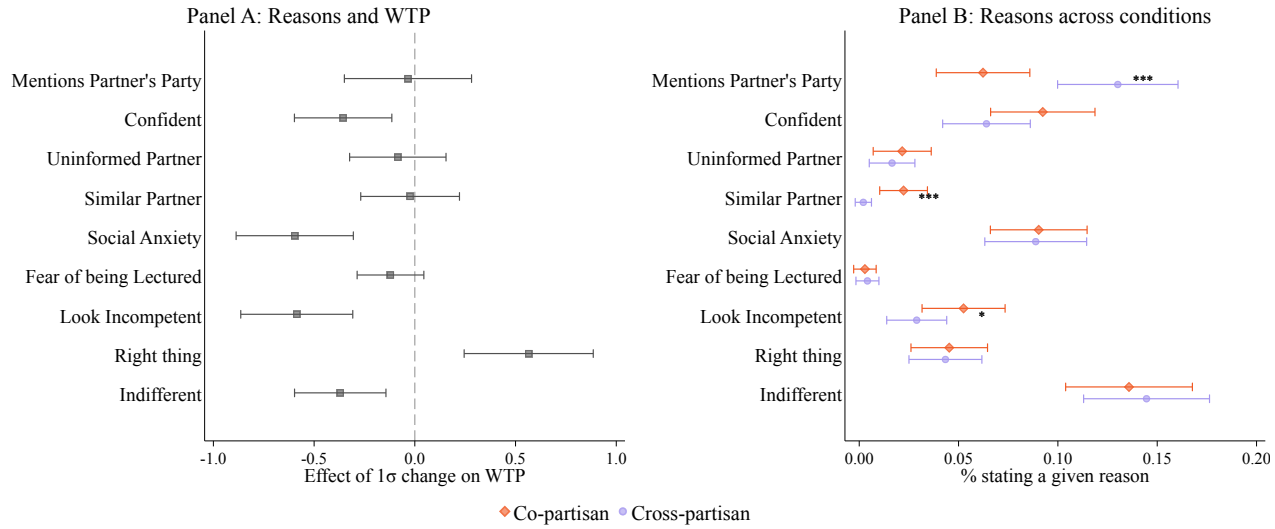

Figure A.1: Reasons for WTP Decision (other)

Notes: Panel A shows the coefficients from regressing WTP on hand-coded explanations participants gave regarding their WTP decision. Row variables: *Mentions partner's party* — participant mentions the partner's political affiliation; *Confident* — confident in their own answer, nothing to learn; *Uninformed partner* — partner unlikely to be well-informed, nothing to learn; *Similar partner* — partner likely knows the same answers; *Social anxiety* — worried due to shyness or social anxiety; *Fear of being lectured* — does not want to be lectured or confronted; *Look incompetent* — worried about appearing incompetent; *Right thing* — talking to the other person is the right thing to do; *Indifferent* — explicit indifference. Panel B shows predicted values from regressions of each hand-coded explanation on a cross-partisan treatment indicator; stars indicate a significant difference between co- and cross-partisan groups. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals use robust standard errors. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

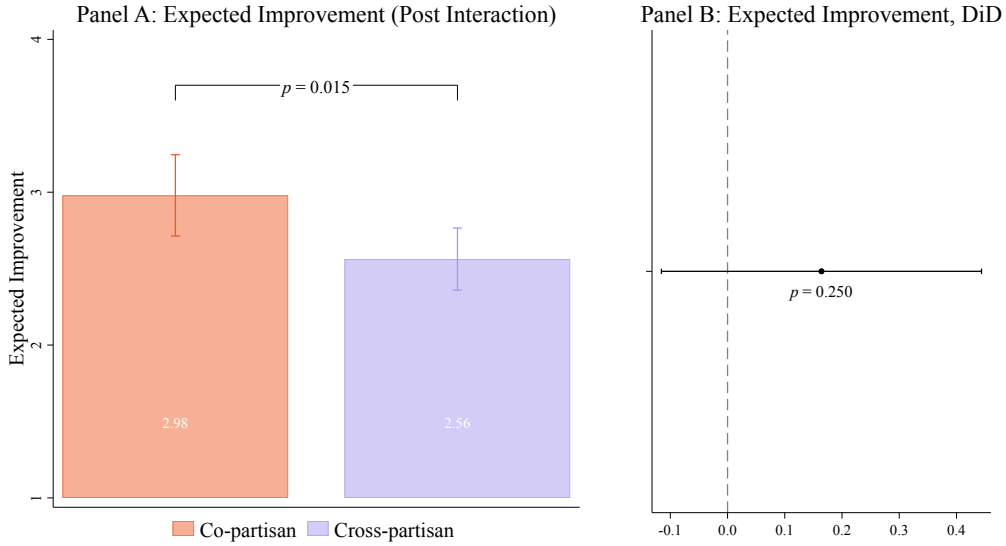

Figure A.2: Expected Improvement (Post Interaction) by Partisan Composition

Notes: Panel A shows predicted values from regressing expected improvement after the interaction on a cross-partisan treatment indicator. Panel B shows the difference-in-differences estimator of expected improvement before and after the intervention. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals use standard errors clustered at the pair level.

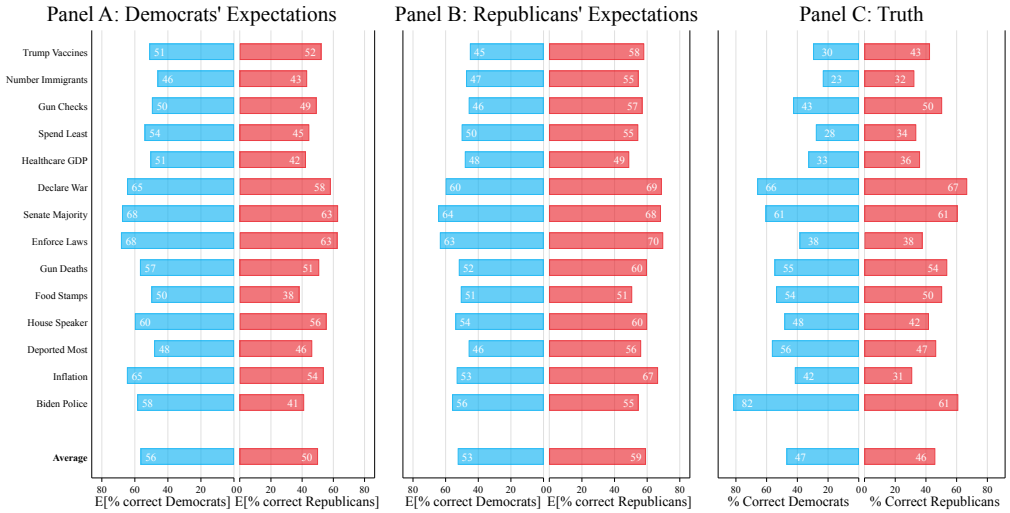

Figure A.3: Expected and Actual Shares of Correct Responses

Notes: The figure shows the expected and actual shares of correct responses by respondents affiliated with different parties. Each row is a question from the factual quiz (crosswalk in Appendix Table A.1). Rows are ordered by the Republican–Democrat difference in correct-answer shares on the Initial Quiz, in decreasing order. Panel A: Democrats' expectations about Democrats and Republicans. Panel B: Republicans' expectations about Democrats and Republicans. Panel C: actual shares of correct responses to the Initial Quiz, by party, in our main sample.

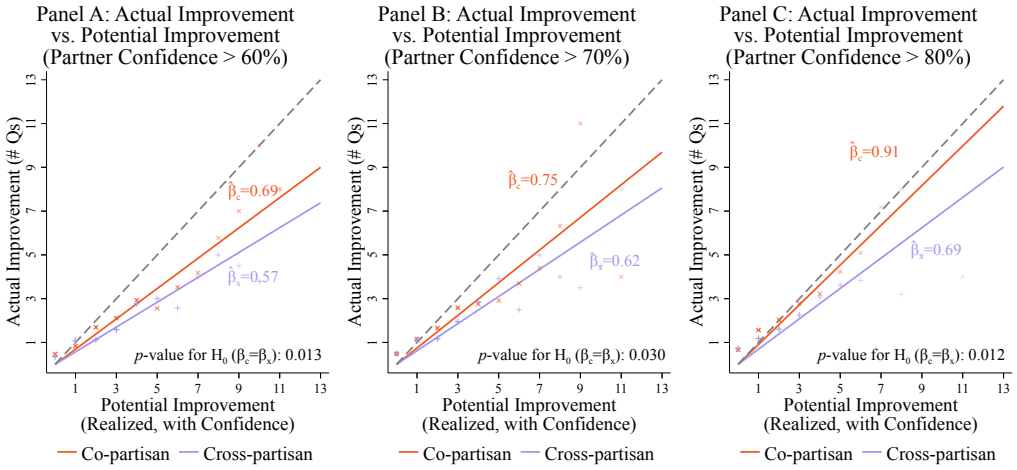

Figure A.4: Actual Improvement and Potential Improvement with Confidence Thresholds

Notes: The figure regresses actual improvement on potential improvement, conditional on the partner's confidence in their Initial Quiz answer exceeding threshold k, in a no-intercept model (Equation 3.1). Potential improvement is interacted with treatment, with β_c and β_x estimating the ease of knowledge extraction in the co- and cross-partisan treatments. Panels correspond to different confidence thresholds: Panel A ($k = 60$ percent), Panel B ($k = 70$ percent), and Panel C ($k = 80$ percent). In the overlaid scatterplots, opacity is proportional to the square root of the bin sample size. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. All outcomes are in number of correct quiz answers (out of 14). The 95 percent confidence intervals and p-values use standard errors clustered at the pair level.

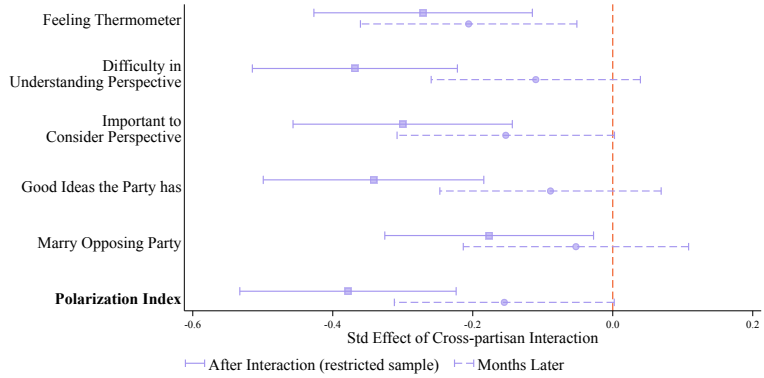

Figure A.5: Affective Polarization (Restricted Sample)

Notes: The figure plots coefficients on the cross-partisan interactions indicator from regressions of the 5 outcome variables and a standardized index of their means. The sample is restricted to follow-up respondents ($N = 682$). All outcomes are standardized. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals use standard errors clustered at the pair level.

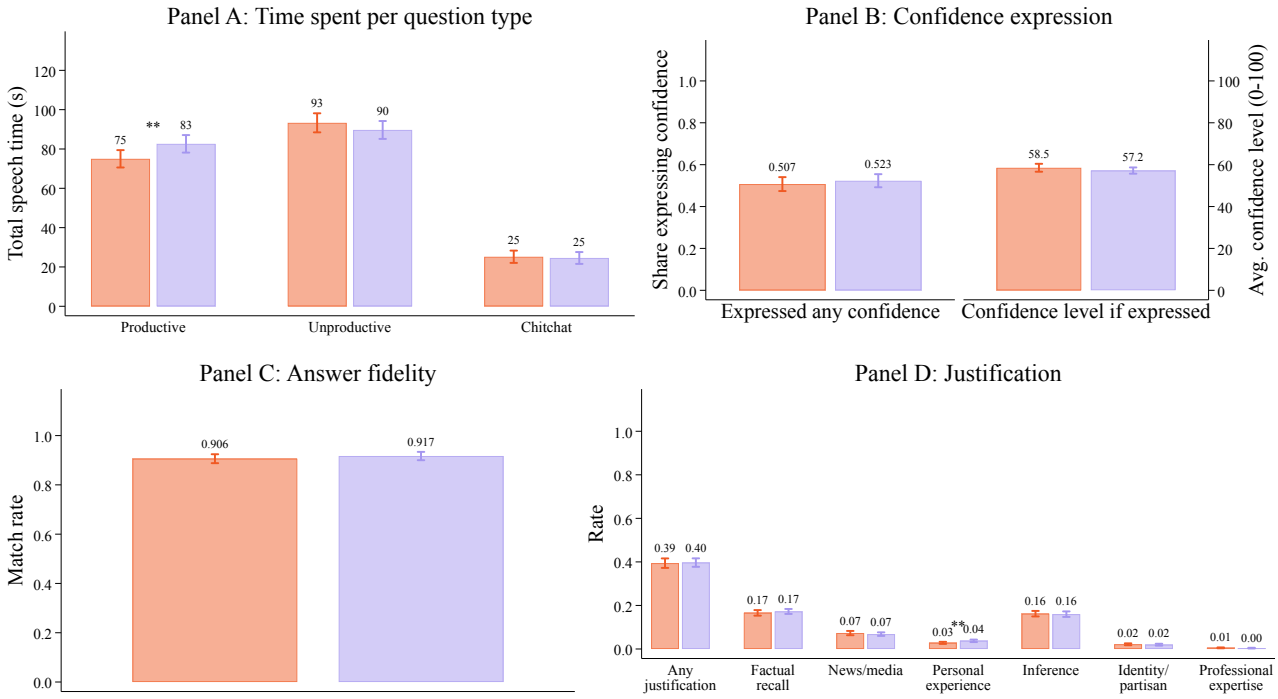

Figure A.6: Answers and Confidence during the Conversation

Notes: The figure shows participant behavior during the quiz discussion. Panel A presents total speech time in seconds by participant across question types — productive questions are those where exactly one partner has the correct answer; unproductive questions are those where either both or neither do. Panel B presents confidence expression and confidence levels of stated answers — first the sample mean of having expressed any confidence around an answer, then the confidence level conditional on expressing any. Panel C presents match rates for answer choices across treatments — match rate is 1 if the answer the participant gives in the call matches their answer in the Initial Quiz, 0 if they state a different answer, and blank if they did not share their answer. Panel D presents the rates at which each justification type was given. Data on in-call behavior (Panels B–D) was hand-coded by research assistants blind to treatment. Panels B–D include question fixed effects. Significance levels come from regressions with standard errors clustered at the pair level, * p<0.1, ** p<0.05, *** p<0.01.

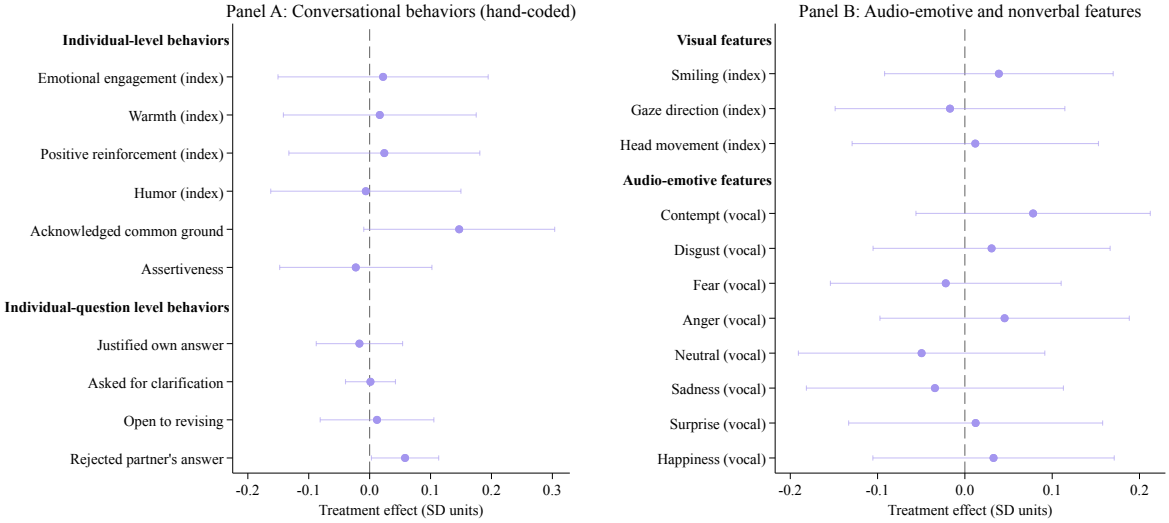

Figure A.7: Treatment Effects on In-Conversation Behavior

Notes: Panel A shows the effect of cross-partisan treatment on different behaviors observed during the call; in-call behaviors were hand-coded by research assistants blind to treatment. Panel B shows the same treatment effect on audio-visual indicators during the conversation: facial features were elicited using MediaPipe Pose, and audio-emotional features were extracted using Microsoft's WavLM Large model for categorical emotion detection. All outcomes are standardized.

B. Analysis of the Videos of the Conversations

The audio tracks from each pair's interaction videos were transcribed using the pre-trained transcription model provided by Vosk, an offline open-source speech recognition toolkit. Specifically, we employed vosk-model-en-us-0.42-gigaspeech, a "generic U.S. English model trained by Kaldi on Gigaspeech" (Shmyrev et al., 2024). The transcription output included the recognized words along with their start and end times.

To attribute transcribed words to individual speakers and determine the temporal order of their speeches (a process known as diarization), we used NVIDIA's NeMo framework (Harper et al., 2024). This framework leverages vocal pitch and other audio features to identify speech zones and assign speaker labels. We utilized the pre-configured settings for telephone recordings from the diar_infer_telephonic.yaml file on NeMo's GitHub page and fixed the number of speakers to two. To construct a speaker-level speech dataset, we mapped the diarization output to the transcribed words.[1]

To ensure accuracy, our team manually reviewed each interaction video to match participant identifiers with the speakers recognized by NeMo. We also verified the quality of the diarization output by checking whether speech from different speakers was properly distinguished.[2] Additionally, two research assistants compared all transcripts against the original video calls to resolve any discrepancies in speech recognition.

For categorizing the processed transcriptions, we employed the OpenAI ChatGPT-3.5 Turbo API. The API was used to classify each transcribed segment by determining which quiz question was being discussed or identifying digressions coded as "chitchat." The algorithm analyzed each segment in the context of surrounding text and metadata. To ensure accuracy, two research assistants manually reviewed all categorizations and resolved any errors in the automated classifications.

In addition to transcription, we extracted fundamental pitch (F0) and other acoustic features from the diarized speech segments using Parselmouth (Jadoul et al., 2018), a Python library for Praat software (Boersma and Weenink, 2021). The final dataset is structured at

[1]Occasionally, a single speaker's continuous speech was split into multiple zones. To address this, we merged consecutive segments belonging to the same speaker if the gap between them was less than a second.

[2]In rare cases where the NeMo framework failed to differentiate speakers, we manually corrected the diarization and appended the revised transcription to the dataset.

the speech-segment level. For instance, a sample transcribed segment might read, "and then there's the thing it's deliberative or enforcing." For analyses conducted at the category level, we concatenated the transcribed text in chronological order.

Conversation behavior during the video call was coded by five research assistants in the Spring of 2026. Research assistants were assigned videos at random, they were blind to treatment, and were not involved in any other coding related to this study. They coded a set of variables at both the individual and individual-question level; the full list, together with the scale, source, and definition of each variable, is presented in Appendix Table B.1. We further standardize some of these hand-coded variables and aggregate them into behavioral indices. We construct a "warmth index" which includes the level of being friendly and polite, the level of being open to listening, a binary indicator for letting the conversation partner finish their sentence (Rapport 7), and a binary indicator for giving the partner visual affirmation (Rapport 1). We also construct a "humor index," composed of binary indicators for both non-threatening (Rapport 8) and threatening humor (Rapport 9). The "positive reinforcement index" aggregates binary indicators for complimenting the partner, having a generally positive conversation, and showing interest in the conversation partner (Rapport 10). Finally, we construct an "emotional engagement index," which uses binary indicators for sharing feelings and emotions (Rapport 5), validating the partner's feelings (Rapport 6), and sharing something personal (Rapport 3).

We supplemented this hand-coding with automated extraction of nonverbal and paralinguistic features from the recordings. From the video, we used MediaPipe Bazarevsky et al. (2020) to obtain body-pose landmarks — from which we derive posture and gesture measures such as head height and wrist movement. We used OpenFace Hu et al. (2025) to extract facial action-unit activations and gaze direction. From the audio, we isolated each participant's speech segments and applied a WavLM-large model fine-tuned for categorical emotion recognition Feng et al. (2025), which assigns each segment a probability distribution over eight interpretable emotion classes. All automated features were computed per participant over the full call and can be found in Appendix Table B.1. We also construct standardized nonverbal behavioral indices from these measures. We construct a "smiling index" using the cheek raiser (AU06) and lip corner puller (AU12) action units. We also construct a "gaze direction index" which uses gaze pitch and gaze yaw measures. Lastly,

we construct a "head movement index," using head sway and shoulder movement.

The conversation data offers valuable insights into how interactions unfold. Figure B.1 examines what participants focus on during their conversations. Although participants were not explicitly required to discuss the quiz, they devoted most of their time to quiz-related topics, with only a small portion of the conversation dedicated to chitchat. Interestingly, the distribution of time across topics shows only minor differences between treatments.

One might hypothesize that participants in cross-partisan interactions avoid discussing contentious issues where they disagree. If this were the case, such avoidance could contribute to explaining the lower knowledge extraction observed in cross-partisan conversations. Table B.2, however, suggests otherwise: participants spend more time discussing points of disagreement, with this tendency being even more pronounced in cross-partisan interactions. Specifically, column (1) shows that questions involving disagreement between conversation partners are discussed in greater depth, while column (2) reveals that cross-partisan interactions place an even stronger emphasis on disagreements compared to co-partisan ones.

In the next appendix section, we delve further into the conversation data to identify the factors contributing to the treatment gap in knowledge extraction.

Table B.1: Variable Overview

Variable	Scale	Source	Level	Description
Handcoded Variables				
Open to listening	1–5	Video	Individual	Rating of how open the participant is to listening to their partner
Friendly and polite	1–5	Video	Individual	Rating of how friendly and polite the participant is
Hedging vs. assertive	1–5	Video	Individual	Rating of tone, from hedging (1) to assertive (5)
Positive conversation evaluation	0–1	Video	Individual	Whether the participant evaluates the conversation positively
Compliment partner in general	0–1	Video	Individual	Whether the participant compliments their partner in general terms
Rapport 1: Affirmation	0–1	Video	Individual	Consistently gives affirmation (e.g. nodding, verbal, smiling)
Rapport 2: Self-introduction	0–1	Video	Individual	Introduced themselves
Rapport 3: Personal disclosure	0–1	Video	Individual	Shared something more personal

Continued on next page

Table B.1 – continued from previous page

Variable	Scale	Source	Level	Description
Rapport 4: Common ground	0–1	Video	Individual	Acknowledged and remarked on common ground in response to the partner, excluding simply giving the same answer
Rapport 5: Emotional sharing	0–1	Video	Individual	Shared feelings or emotions
Rapport 6: Emotional validation	0–1	Video	Individual	Validated the partner's feelings or emotions
Rapport 7: Non-interruption	0–1	Video	Individual	Mostly lets the partner finish their thoughts and does not consistently interrupt
Rapport 8: Non-threatening humor	0–1	Video	Individual	Non-threatening humor (e.g. self-deprecating, politically correct)
Rapport 9: Threatening humor	0–1	Video	Individual	Threatening humor (e.g. sarcasm, edgy jokes)
Rapport 10: Showed interest	0–1	Video	Individual	Showed interest in the partner (e.g. asked open, person-centred questions)
Initial answer in call	0–5	Video	Individual-question	Initial answer choice for the question in the call
Revised answer in call	0–5	Video	Individual-question	Revised answer choice for the question in the call
Expressed confidence	0–1	Video	Individual-question	Whether participant explicitly expresses a confidence level
Expressed confidence level	0–100	Video	Individual-question	Numeric confidence level expressed (only when Expressed confidence $= 1$)
Open to revise question	0–1	Video	Individual-question	Whether the participant is open to revising their answer
Rejected answer	0–1	Video	Individual-question	Whether the participant rejects the proposed/-partner's answer
Requested justification	0–1	Video	Individual-question	Whether the participant requests a justification
Justified answer	0–1	Video	Individual-question	Whether the participant justifies their own answer
Justification 1: Factual recall	0–1	Video	Individual-question	Justification based on factual recall
Justification 2: News/media	0–1	Video	Individual-question	Justification based on news or media
Justification 3: Personal experience	0–1	Video	Individual-question	Justification based on personal experience
Justification 4: Inference	0–1	Video	Individual-question	Justification based on inference or reasoning
Justification 5: Identity/partisan	0–1	Video	Individual-question	Justification based on identity or partisanship
Justification 6: Professional expertise	0–1	Video	Individual-question	Justification based on professional expertise
Machinecoded Variables				
MediaPipe Pose				
Mean shoulder angle	$[-\pi, \pi]$	Video	Individual	Mean angle of the shoulder line (radians); body orientation/lean
Mean head height	0–1	Video	Individual	Mean normalised vertical nose position; proxy for upright posture
Hand-near-face proportion	0–1	Video	Individual	Proportion of frames with a wrist close to the face
Mean wrist movement	≥ 0	Video	Individual	Mean frame-to-frame wrist displacement (normalised); gesture activity
OpenFace				
AU06 cheek raiser	Cont.	Video	Individual	Mean activation of AU06 (cheek raiser)

Continued on next page

Table B.1 – continued from previous page

Variable	Scale	Source	Level	Description
AU12 lip corner puller	Cont.	Video	Individual	Mean activation of AU12 (lip corner puller; smiling)
Gaze pitch	Cont.	Video	Individual	Mean vertical gaze angle (radians)
Gaze yaw	Cont.	Video	Individual	Mean horizontal gaze angle (radians)
Microsoft WavLM speech-emotion model				
Anger	0–1	Audio	Individual	Mean predicted probability of anger across the participant's speech segments
Contempt	0–1	Audio	Individual	Mean predicted probability of contempt across speech segments
Disgust	0–1	Audio	Individual	Mean predicted probability of disgust across speech segments
Fear	0–1	Audio	Individual	Mean predicted probability of fear across speech segments
Happiness	0–1	Audio	Individual	Mean predicted probability of happiness across speech segments
Neutral	0–1	Audio	Individual	Mean predicted probability of neutral across speech segments
Sadness	0–1	Audio	Individual	Mean predicted probability of sadness across speech segments
Surprise	0–1	Audio	Individual	Mean predicted probability of surprise across speech segments

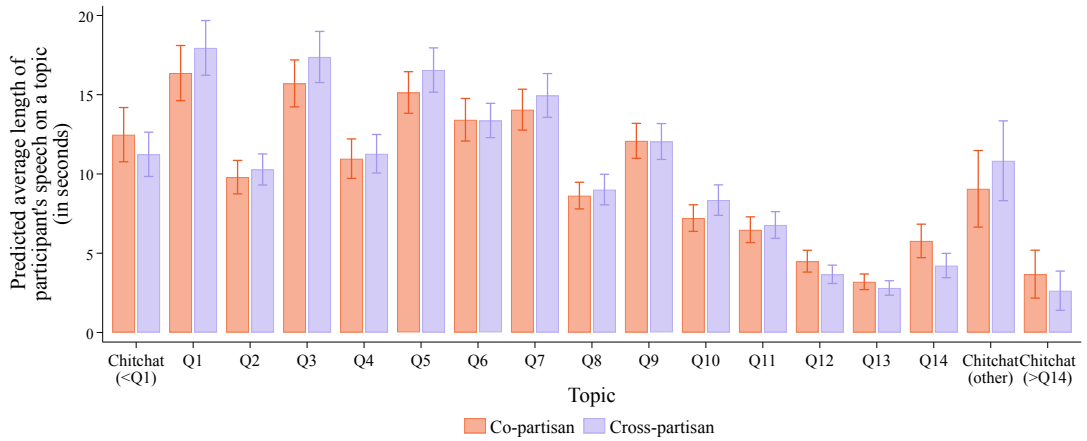

Figure B.1: Speech Length by Question

Notes: The figure shows the time participants spend discussing each question and engaging in small talk (chitchat). Questions are denoted Q1, Q2, ..., Q14 and listed in Table A.1. The dependent variable is the total duration of a participant's speech in each x-axis category. The figure shows predicted values from a regression on the participant-question level dataset. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals use standard errors clustered at the pair level. The speech analysis excludes 4 participants: 3 with audio issues who relied on hand gestures, and 1 whose video was not saved due to a technical glitch.

Table B.2: Share of Speech and Disagreement

	(1) Nb Words	(2) Nb Words
Disagreement	18.44***	17.00***
	(0.755)	(1.044)
Cross		0.655
		(1.143)
Disagreement × Cross		2.953**
		(1.494)
Social Potency		4.392**
		(1.969)
Confidence on Initial Quiz		-0.00597
		(0.0136)
Question FE	✓	✓
Participant RE	✓	✓
Observations	11192	11192
Participants	984	984

Notes: This table examines the relationship between the time conversation partners spend on a question and whether they disagree about its answer. The dataset is at the participant-question level; we include participant-level random effects and question fixed effects. The speech analysis excludes 9 participants: 3 with audio issues who relied on hand gestures, 1 whose video was not saved, and 5 who engaged only in small talk (chitchat). Standard errors clustered at the pair level in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table B.3: Classifier Performance: Can Algorithms Tell Co- and Cross-Partisan Conversations Apart?

Representation of the conversation	(1) Number of features	(2) Out-of-sample AUC	(3) Permutation p
TF-IDF (bag of words)	vocab	0.502	0.365
Hand-coded dialogue moves (RA)	66	0.503	0.073
LLM-rated conversation features	135	0.545	0.082
Full-text embeddings (`voyage-4-large`)	1,024	0.559	0.040

Notes: Each row reports the out-of-sample AUC of a classifier trained to predict whether a conversation is cross- or co-partisan from a different representation. $N = 499$ conversations (243 cross-partisan, 256 co-partisan). AUC is estimated via 5-fold stratified cross-validation. Permutation p-values come from 300–1,000 random permutations of the treatment label. AUC of 0.50 is chance performance. TF-IDF results use L1-regularized logistic regression on bag-of-words features, selected from the best-performing unigram, bigram, or trigram specification with truncated SVD. Hand-coded dialogue moves use L1-regularized logistic regression on 66 RA-coded dialogue-move rates, including friendliness, hedging, justifications, rapport, rejection, and related conversational behaviors. LLM-rated conversation features include warmth, assertiveness, reasoning quality, and disagreement patterns, using the best-performing classifier among LASSO, random forest, and gradient boosting. Full-text embeddings use L1-regularized logistic regression on 1,024-dimensional `voyage-4-large` embeddings of the full transcripts, using an 8,192-token context window. The best representation—full-text neural embeddings—exceeds chance only modestly, and the hand-coded LASSO sets all 66 coefficients to exactly zero.

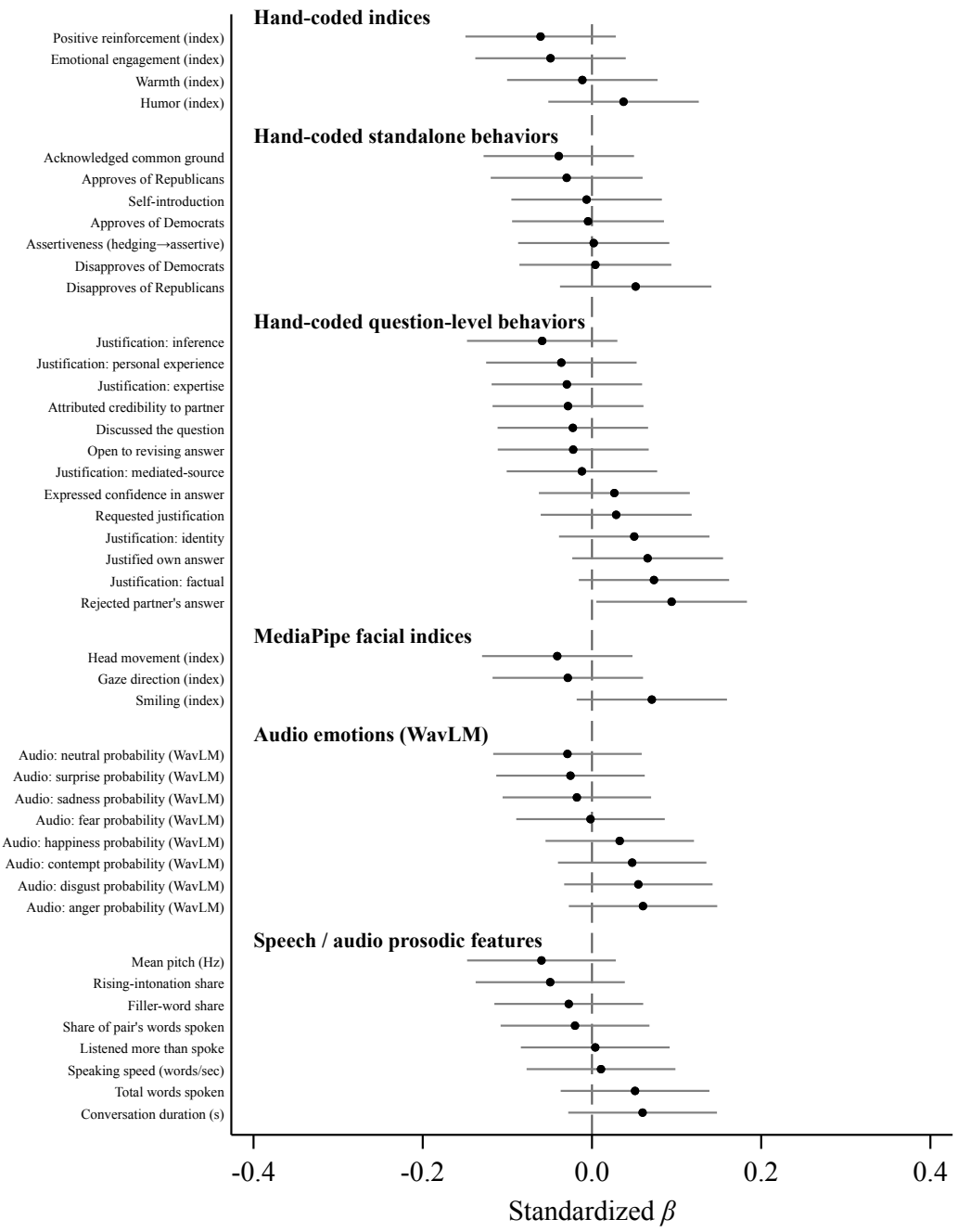

Figure B.2: Correlations of Conversation Features with the Embedding Classifier's Discriminant Direction

Notes: Each dot is the standardized coefficient from a separate univariate OLS of the embedding classifier's discriminant-direction score on the listed conversation feature; horizontal lines are 95 percent confidence intervals. Positive coefficients indicate features more prevalent in conversations identified as cross-partisan by the classifier (full-text `voyage-4-large` embeddings; see Table B.3). Hand-coded and MediaPipe indices follow Appendix Figure A.7: rapport components are first taken as the max across the 14 quiz questions for each participant, then averaged across the two participants in each conversation, z-scored, and combined via row-mean. Features are grouped by category and sorted within category by signed β. $N = 486$–499 conversations depending on feature availability.

C. Partisan Differences in Cross-partisan Contact

Our main analyses compared cross-partisan and co-partisan conversations without differentiating between Democrats and Republicans. Here, we revisit the key findings separately for each party. The results are qualitatively similar across groups and align with the aggregated analysis presented in the main text. However, splitting the sample reduces statistical power, limiting our ability to detect partisan differences with confidence.

Figure C.1 shows that Democrats exhibit a significantly lower willingness to pay for cross-partisan conversations compared to co-partisan ones. They are also considerably more likely to express a strict preference against interacting with counter-partisans than against co-partisan interactions. For Republicans, these preferences are less pronounced; the gaps are smaller and not statistically significant.

Both parties expect to learn less from counter-partisans, but this effect is more pronounced and statistically significant only for Republicans (Figure C.2, Panels A and C). When it comes to actual learning, participants from both parties learn less from counter-partisans, but neither effect is statistically significant at conventional levels (Figure C.2, Panels B and D).

Regarding the hedonic experience of the conversation (Figure C.3), both parties enter cross-party interactions with greater pessimism. This pessimism appears stronger among Democrats. Post-conversation, Republicans report enjoying cross-partisan conversations as much as co-partisan ones, while Democrats show a slight and marginally significant preference for co-partisan interactions.

Finally, as illustrated in Figure C.4, cross-partisan conversations lead to substantial reductions in affective polarization for both parties. Since cross-partisan contact requires mutual willingness to engage — given that either participant can choose to walk away — the observed decrease in affective polarization among both Republicans and Democrats underscores the potential of such interactions to foster social cohesion. These findings bolster optimism that policies encouraging cross-partisan contact may promote greater harmony in future exchanges.

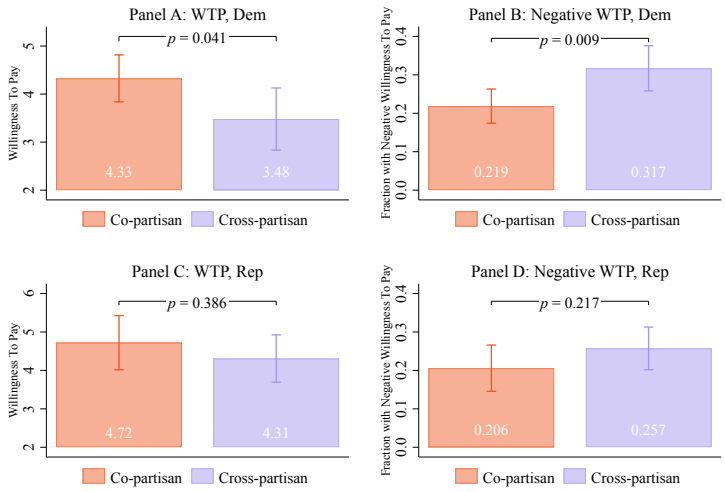

Figure C.1: Willingness to Pay, by Party

Notes: The figure shows predicted values from regressions where Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. Panels A and B: Democrats. Panels C and D: Republicans. The 95 percent confidence intervals use robust standard errors.

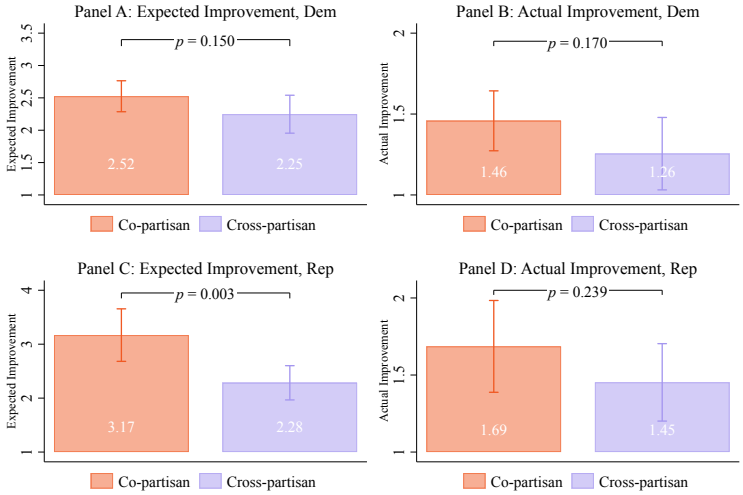

Figure C.2: Expected and Actual Improvement, by Party

Notes: Panel A [C] shows predicted values from a regression of expected improvement on a cross-partisan treatment indicator, for Democrats [Republicans]. Panel B [D] shows the same for actual improvement. Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals use robust standard errors (Panels A, C) and standard errors clustered at the pair level (Panels B, D).

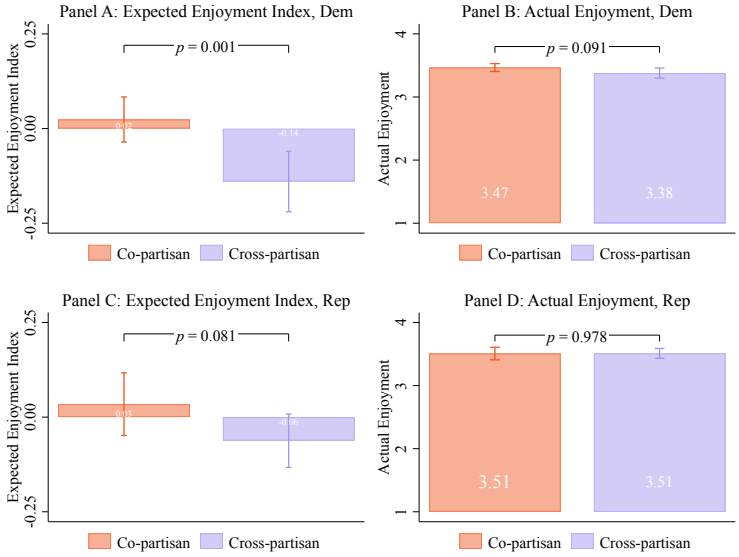

Figure C.3: Expected and Actual Enjoyment, by Party

Notes: Panel A [C] shows predicted values from a regression of an expected enjoyment index on a cross-partisan treatment indicator, for Democrats [Republicans]. The index is coded as in Figure 7. Panel B [D] shows predicted values from regressing reported enjoyment after the interaction on a cross-partisan treatment indicator, for Democrats [Republicans], on a 1–4 Likert scale (1 = "strongly disagree", 4 = "strongly agree"). Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals and *p*-values use robust standard errors (Panels A, C) and standard errors clustered at the pair level (Panels B, D).

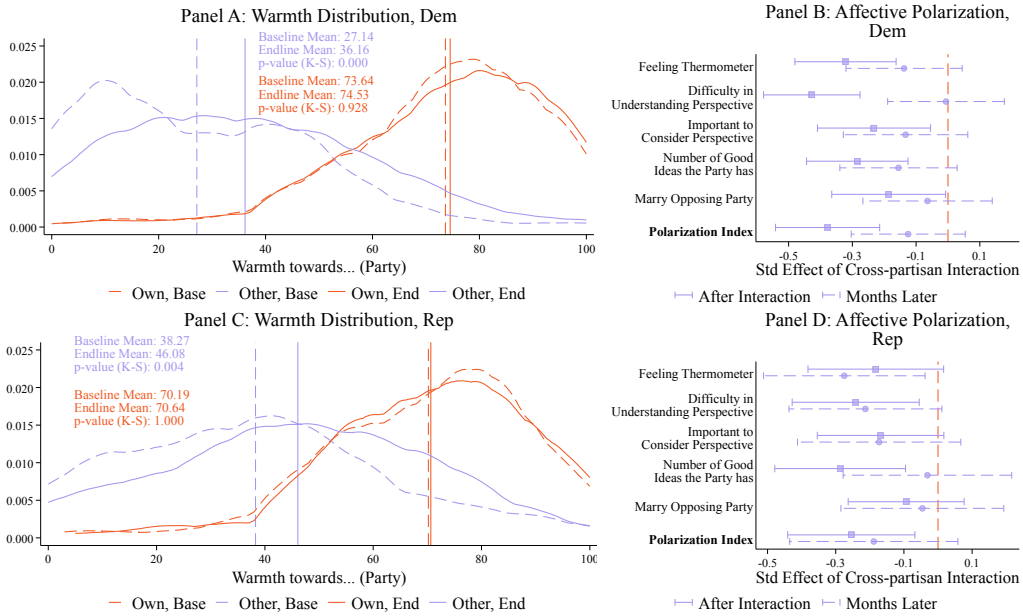

Figure C.4: Affective Polarization, by Party

Notes: Panel A [C] shows the distributions of baseline and endline warmth toward one's own and the other party, restricted to the cross-partisan treatment group and to Democrats [Republicans]. Panel B [D] plots coefficients on a cross-partisan treatment indicator from regressions of the 5 outcome variables and an equally weighted index, for Democrats [Republicans]. Outcomes are measured immediately after the interaction and again ∼100 days later, and standardized. We deviate from the pre-registration by including the "Marry opposing party" variable in the polarization index (more comprehensive and conservative, as discussed in the main text). Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The 95 percent confidence intervals use standard errors clustered at the pair level.

D. Factual Polarization

D.1. Correlates of Factual Polarization

What drives factual polarization? To explore this question, we construct a measure of baseline factual alignment, which captures the extent to which participants' responses to the baseline quiz align with Democratic or Republican perspectives. This measure ranges from very Democratic (1) to very Republican (5).[3]

Table D.1 shows that baseline factual alignment is significantly correlated with two key factors: self-reported ideological intensity, which ranges from very liberal (-3) to very conservative (3), and news consumption slant, measured as the average slant of the self-reported media consumed, ranging from very liberal (-1) to very conservative (1).

Table D.1: Correlates of Baseline Factual Alignment

	(1) Factual Alignment	(2) Factual Alignment
News Slant	0.156***	
	(0.0597)	
Reported Ideology		0.0268***
		(0.00735)
Observations	417	993
Sample mean	2.538	2.567
R^2	0.015	0.013

Notes: Baseline factual alignment is constructed using ChatGPT-4 to rank all possible quiz answers from most Democratic-aligned (1) to most Republican-aligned (5); for each participant we then average across the seven questions where the share of correct responses differs significantly between Democrats and Republicans. Ideology intensity ranges from very liberal (-3) to very conservative ($+3$). News consumption slant is the average slant of the media a participant reports consuming, ranging from very liberal (-1) to very conservative ($+1$). Republican/Democrat-only pairs are reweighted to balance partisan composition across treatment groups. The lower N in Column (1) reflects that not everyone consumes news. Robust standard errors in parentheses, * p<0.1, ** p<0.05, *** p<0.01.

D.2. Treatment Effect on Factual Polarization

To analyze the degree of factual polarization among Democrats and Republicans within each treatment condition, we use linear discriminant analysis (LDA). LDA is particularly well-suited for this analysis because it performs well with smaller sample sizes. Specifically, we employ a leave-one-out classification method, predicting each participant's partisan affiliation based on their responses to individual quiz questions. We then calculate the

[3]Specifically, using ChatGPT-4, we rank all possible answers to the quiz questions on a scale from most aligned with Democratic views (1) to most aligned with Republican views (5). For each participant, we then calculate their average alignment across the seven quiz questions where the share of correct responses differs significantly between the parties.

share of participants whose party affiliation is correctly predicted. A higher share of correct predictions indicates that participants' responses are more closely aligned with their party affiliation, suggesting greater factual polarization. To assess statistical significance, we perform chi-squared tests for identical distribution across conditions and compute p-values.

Figure D.1 shows that the degree of factual polarization, as proxied by the share of correctly predicted responses, is similar across the two conditions in the Initial Quiz. However, in the cross-partisan condition, the share of correctly predicted responses decreases significantly, indicating a reduction in factual polarization. Although the differences between conditions become less pronounced in the follow-up survey, they remain statistically significant.

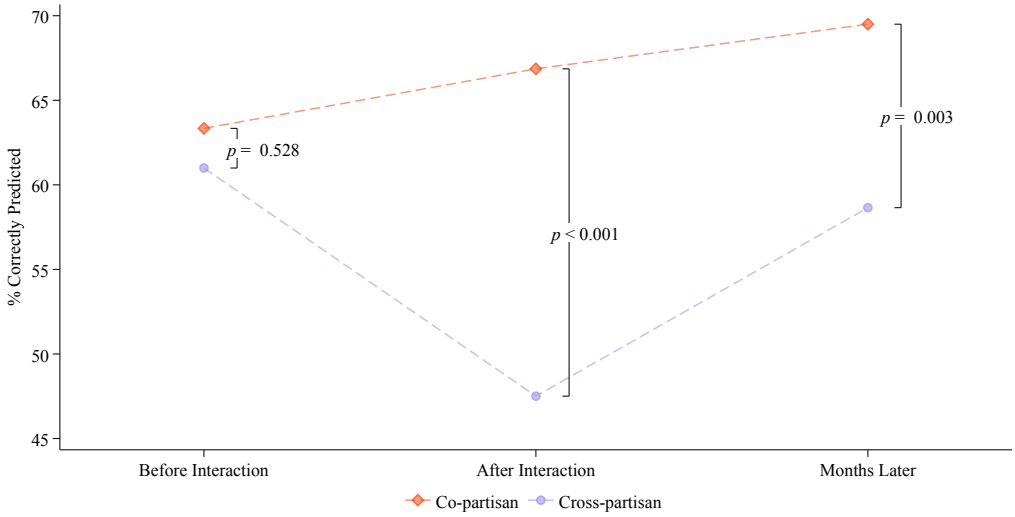

Figure D.1: Factual Polarization, Linear Discriminant Analysis

Notes: We predict participants' political party affiliation using leave-one-out linear discriminant analysis (LDA) on their responses to the Initial, Revised, and follow-up quizzes. Accuracy can fluctuate below 50 percent when no systematic difference between classes is detected. The p-values come from Pearson's chi-squared tests for whether the distribution of correctly classified observations is the same across cross- and co-partisan treatments. The before- and after-interaction analyses use the 993 main-study participants; the follow-up analysis uses the 682 follow-up respondents. Results are qualitatively unchanged when restricting throughout to the 682 follow-up respondents.

E. Interpersonal Contact: Meta-analysis

We conduct a meta-analysis to examine the short- and long-term effects of interpersonal contact interventions on reducing prejudice and increasing social cohesion. By aggregating evidence, we contextualize our quantitative results within the broader literature and compare design features across studies. This exercise illustrates that our study is the largest to date in estimating the long-term effects of contact interventions and that the effect sizes of cross-partisan contact that we estimate on our pre-registered measures of social cohesion (affective polarization) align with the publication-bias-adjusted meta-analytic effects of this literature (both for short- and for long-term effects).

As detailed in the protocol of this meta-analysis in the replication files, for each study, we identify the outcome of interest and categorize the interventions by type of contact, intensity, mode of interaction (virtual vs. in-person), whether long-term outcomes are measured (and the time frame), and whether beliefs about the expected value of contact were assessed prior to the interaction. Figure E.1 plots the standardized short- and long-term effects of these interventions against study sample sizes. Similar to findings by DellaVigna and Linos (2022) on nudge interventions and Goette and Tripodi (2024) on social recognition interventions, we observe that larger studies tend to report smaller effect sizes — a pattern potentially attributable to publication bias. For both short- and long-term effects, we estimate publication-bias-corrected meta-analytic effects using the methodology of Andrews and Kasy (2019) and report these estimates in the figure.

In Table E.1, we address concerns that specific features of our intervention might limit its comparability to other studies. Notably, the effect size of cross-partisan interactions in our study is comparable to those of studies with similar characteristics (e.g., cross-partisan, pre-registered, and virtual). Moreover, it is not substantially different from the effect sizes observed in other types of interventions.

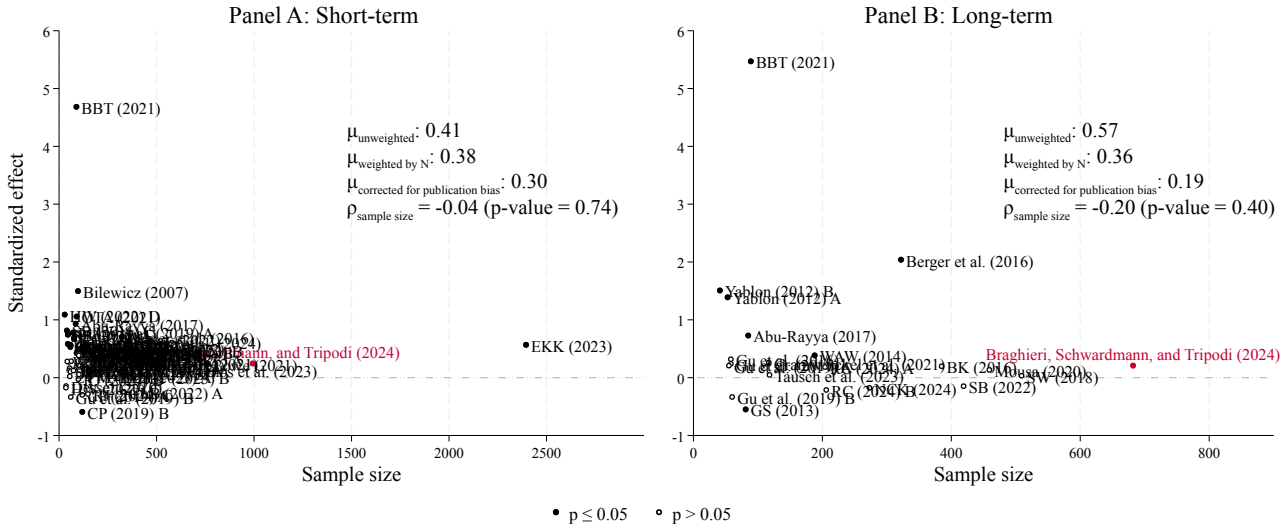

Figure E.1: Meta-Analysis

Notes: Effect-level details are in the replication files. Long-term effects are those measured after the date of intervention. Standardized effects are the ratio of the mean difference to the control-group standard deviation. *p*-values come from t-tests for equality of means. Publication-bias-corrected estimates follow Andrews and Kasy (2019), assuming symmetry and t-distribution of effect sizes. When sample sizes are not reported separately by condition, we impute by halving the total. ρ is the correlation between standardized effect and sample size; the *p*-value refers to the significance of that correlation.

Table E.1: Meta-Analytic Effects, Short-Term

Study sample	**N**	**Unweighted**	**Publication bias-adjusted**
By category			
Ethnic	19	0.71	0.57
Cross-partisan	15	0.28	0.22
LGBT	12	0.29	0.24
Racial	11	0.25	0.22
Refugee	3	0.68	0.53
Miscellaneous	13	0.33	0.10
Pre-registered			
Yes	18	0.30	0.29
No	55	0.46	0.36
Virtual	27	0.32	0.18
In-person	46	0.47	0.40
Full sample	73	0.42	0.30

Notes: The sample is 73 estimates from 52 papers measuring short-term effects of contact interventions. Publication-bias-adjusted effects follow Andrews and Kasy (2019), assuming symmetry and t-distribution of effect sizes. "Miscellaneous" covers categories with fewer than 3 studies (neurodivergent, schizophrenic, overweight, religious, generational, sectarian, urban-rural, ex-offender contact). Studies with both in-person and virtual contact are coded as in-person. Studies reporting only long-term outcomes are excluded.

Appendix References

Andrews, Isaiah and Maximilian Kasy, "Identification of and Correction for Publication Bias," *American Economic Review*, August 2019, *109* (8), 2766–2794.

Bazarevsky, Valentin, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann, "BlazePose: On-device Real-time Body Pose Tracking," in "CVPR Workshop on Computer Vision for Augmented and Virtual Reality" 2020.

Boersma, Paul and David Weenink, "Praat: doing phonetics by computer [Computer program]," Version 6.1.38, retrieved 2 January 2021 http://www.praat.org/ 2021.

DellaVigna, Stefano and Elizabeth Linos, "RCTs to scale: Comprehensive evidence from two nudge units," *Econometrica*, 2022, *90* (1), 81–116.

Feng, Tiantian, Jihwan Lee, Anfeng Xu, Yoonjeong Lee, Thanathai Lertpetchpun, Xuan Shi, Helin Wang, Thomas Thebaud, Laureano Moro-Velazquez, Dani Byrd et al., "Vox-Profile: A Speech Foundation Model Benchmark for Characterizing Diverse Speaker and Speech Traits," *arXiv preprint arXiv:2505.14648*, 2025.

Goette, Lorenz and Egon Tripodi, "The limits of social recognition: Experimental evidence from blood donors," *Journal of Public Economics*, 2024, *231*, 105069.

Harper, Eric, Somshubra Majumdar, Oleksii Kuchaiev, Li Jason, Yang Zhang, Evelina Bakhturina, Vahid Noroozi, Sandeep Subramanian, Nithin Koluguri, Jocelyn Huang, Fei Jia, Jagadeesh Balam, Xuesong Yang, Micha Livne, Yi Dong, Sean Naren, and Boris Ginsburg, "NeMo: a toolkit for Conversational AI and Large Language Models," 2024.

Hu, Jiewen, Leena Mathur, Paul Pu Liang, and Louis-Philippe Morency, "OpenFace 3.0: A Lightweight Multitask System for Comprehensive Facial Behavior Analysis," *arXiv preprint arXiv:2506.02891*, 2025.

Jadoul, Yannick, Bill Thompson, and Bart de Boer, "Introducing Parselmouth: A Python interface to Praat," *Journal of Phonetics*, 2018, *71*, 1–15.

Shmyrev, Nickolay V. et al., "Vosk Speech Recognition Toolkit: Offline speech recognition API for Android, iOS, Raspberry Pi and servers with Python, Java, C# and Node," 2024.