Breaking the Bubble - The Determinants and Effects of Cross-partisan Contact

Click **Here** for Latest Version

Adrian Blattner* Vlasta Rasocha

(Job Market Paper)

January 10, 2025

Abstract

Should policymakers intervene to reduce social segregation? The answer depends on the effect of increased contact among those who would not seek it out on their own. We conduct a field experiment in the context of political segregation in Brazil to measure the joint distribution of avoidance of cross-partisan interactions and the treatment effects of actual interactions on outgroup hostility. To do so, we randomize participants into interactions providing strong financial compensation for compliance. We find evidence of a *segregation trap*: the most hostile partisans avoid interactions because they incorrectly believe they will be unpleasant and unproductive. Actual interactions correct these beliefs and lead to large, long-lasting reductions in hostility. Importantly, these effects are *largest* among partisans with *high* levels of baseline hostility and outgroup avoidance, and lead to an increase in demand for future interactions. In a cost-benefit analysis, we find that current policy interventions targeting volunteer participants are outperformed by incentive-based interventions at any level of budget. Finally, we show that a lower-cost intervention that substitutes a video for actual interactions can produce similarly beneficial effects.

*Blattner: Stanford University (email: ablatt@stanford.edu); Rasocha: Stanford University (email: rasocha@stanford.edu). We thank My Country Talks and Instituto Sivis, particularly Sara Cooper and Fernanda Santos, for their dedication and trust in our project. We thank Ana Resende, Pedro Oliveira, and Gustavo Chacha for outstanding research assistance. Blattner thanks Matthew Gentzkow, Muriel Niederle, Nicholas Bloom, and Pascaline Dupas for their invaluable guidance. We also thank Hunt Allcott, Doug Bernheim, Arun Chandrasekhar, Marcel Fafchamps, Caroline Hoxby, Al Roth, Dmitry Taubinsky, Noam Yuchtman, and seminar audiences at Stanford and FGV São Paulo. We also thank our freelancers for their work. The Stanford IRB and the Ethics Committee of Insper approved the research protocols. The experimental designs were pre-registered in the AEA RCT registry (AEARCTR-0016625). We thank the National Science Foundation under grant #2315456, the Weiss Fund at the University of Chicago, J-PAL JOI Brazil, J-PAL GI, J-PAL CVI, the Stanford King Center, the Stanford Center on Philanthropy and Civil Society (PACS), Stanford Institute for Research in the Social Sciences (IRiSS), the Stanford Institute for Economic Policy Research (SIEPR), the Konrad Adenauer Stiftung, and the Institute for Humane Studies for financial support for this project.

1 Introduction

Segregation by political ideology, race, or socioeconomic status is widespread (Brown et al. 2025; Chetty et al. 2014; Cutler and Glaeser 1997). Across contexts, segregation has been shown to coincide with hostility between groups, manifesting in adverse outcomes ranging from economic discrimination to intergroup violence (Field et al. 2008; Iyengar et al. 2019).

Segregation is in many cases the result of individual choices, with individuals exhibiting the greatest degree of hostility being those most likely to avoid interactions with outgroup members (Becker 1957; Bertrand and Duflo 2017; Paluck et al. 2019; Schelling 1969). This has led to the proposal to promote intergroup contact to overcome hostility between segregated groups (Allport 1954). However, assessing the societal implications of segregation and the impact of policy responses that aim to address it through contact requires examining who chooses to avoid outgroup interactions and why.

If individuals with high levels of hostility correctly anticipate interactions to be unproductive and antagonistic, endogenous selection into interactions may prevent further escalation of conflict and intergroup hostility. In this scenario, desegregation policies enforcing intergroup contact may not have the intended effects or even backfire (Enos 2014).

On the other hand, individuals with high levels of hostility and outgroup avoidance may hold overly pessimistic expectations about their costs and benefits. Such misperceptions may fuel a vicious cycle, where endogenous segregation reinforces negative stereotypes and intergroup hostility, which in turn stimulate further segregation (Currarini et al. 2009). In this case, endogenous selection into interactions generates societal costs, motivating policy interventions to counter segregation and foster intergroup contact, in particular among those with high degrees of hostility who currently avoid interactions.

This paper investigates the dynamics and consequences of selection into intergroup contact in the context of political segregation. We examine the demand for cross-partisan interactions and the consequences of selection for policy design through a field experiment in Brazil, a country that has seen a strong increase in partisan segregation, partisan intergroup hostility, and political violence in the past decade (Nunes and Traumann 2023).[1]

Our paper aims to address four questions: (1) What drives individuals to avoid cross-partisan interactions, and are people's beliefs about these interactions accurate? (2) What is the impact of cross-partisan interactions on intergroup hostility, interest in future interactions, and beliefs about their costs and benefits? (3) How does endogenous selection affect the impact of interactions? (4) What do these patterns imply for the design of policy interventions?

To study these questions, we conduct a field experiment in Brazil in partnership with three non-profit organizations that have previously facilitated political conversations with over 300,000

[1] In a 2024 survey conducted by PEW research center, 80% of surveyed respondents agree that there are strong conflicts between people who support different political parties in Brazil.

volunteer participants in more than 30 countries. Different from prior work on intergroup contact, our experiment combines incentivized measures of avoidance of cross-partisan interactions and incentivized beliefs about their costs and benefits with causal evidence on the impact of the interactions on individuals with large variation in baseline hostility and avoidance of interactions. This allows us to directly measure whether individuals misperceive the costs and benefits of intergroup contact, study the implications for steady state intergroup hostility and segregation, and assess the impact of current policy interventions to address segregation and hostility based on political ideology.

For our experiment, we recruit a sample of more than 5,000 Brazilians through neutral, apolitical social media ads. In the baseline survey, we elicit incentivized measures of avoidance of cross-partisan interactions and of the perceived costs and benefits of such interactions. Using sizable payments of twice the daily minimum wage, we subsequently randomize more than 2,400 participants into either a control group, or one of two interventions we co-developed with our partner organizations: Participants either (i) participate in a cross-partisan political conversation on an online video call platform (Conversation Treatment), (ii) watch a video with testimonials of partisans who completed a cross-partisan conversation, correcting beliefs about cross-partisan political conversations and differences in cross-partisan characteristics (Video Treatment), or (iii) watch a short apolitical video (Control).

We measure the impact of treatment on intergroup hostility, proxied by an incentivized cash-transfer decision and survey questions, in a short-term endline survey, administered immediately after the intervention. We study longer-term impacts of treatment on intergroup hostility, avoidance of cross-partisan interactions, and beliefs about the costs and benefits of these conversations in three additional endline surveys administered (i) within the first four weeks, (ii) within two months, and (iii) within three months of the intervention.

The baseline survey provides evidence for strong levels of political segregation in Brazil, and that individual segregation levels strongly predict individual hostility. Among baseline respondents, 40% do not have a single member of the political outgroup among their five closest social network contacts, compared to 25% for racial outgroups. Similarly, 39% have not had a conversation about politics with an outgroup member in the last six months, compared to only 17% reporting that they did not have a conversation about politics with an ingroup member. We also confirm that in our context, individuals' degree of segregation strongly correlates (ρ=0.3) with levels of partisan hostility (affective polarization) and ideological extremity (ideological polarization). That is, Brazilians with the highest degrees of partisan hostility are those least likely to engage with the outgroup in their day-to-day lives.

Next, we demonstrate that participants exhibit substantial avoidance of political conversations with the outgroup, which is strongly correlated with pessimistic expectations about the conversations' costs and benefits. As part of the baseline survey, we surprise participants with a paid opportunity to engage in a political conversation with an outgroup member. We elicit incentivized

monetary willingness-to-accept (WTA) measures for a political discussion with the outgroup, the ingroup, as well as non-political conversations. The average participant would need to be paid 45 BRL (USD 8.4) to engage in a 30-minute political conversation with the outgroup, more than twice the median hourly wage in our sample and twice the amount required for a political conversation with the ingroup. This avoidance of political conversations with the outgroup is largely accounted for by pessimistic expectations about the conversation's outcomes and experience: in incentivized predictions, 63% of participants predict a political conversation with an outgroup member to end in a hostile confrontation, compared to 9% for a political conversation with an ingroup member. Individual-level predictions strongly correlate with both hostility and outgroup avoidance.

Together, this set of baseline findings suggests a potential "segregation trap": the most hostile partisans expect contact to be counterproductive, so they avoid interactions with the outgroup. To assess the impacts of cross-partisan conversations and whether the beliefs driving this segregation trap are accurate or mistaken, we randomize both hostile and non-hostile participants into a 30-minute political conversation with an outgroup member. To incentivize conversation take-up, we provide sizable financial incentives. Our modal payment equals roughly twice the daily minimum wage, covering about 80% of the WTA distribution. As part of the intervention, participants engage in an unmoderated one-on-one conversation about policy questions on a video call platform we developed for the experiment, allowing us to obtain audio and video recordings of the interactions and to control the matching procedure.

We find that cross-partisan political conversations lead to statistically and economically significant reductions in intergroup hostility that persist over three months. The conversation treatment reduces the share of participants preferring to allocate a cash transfer to an affluent ingroup household instead of an outgroup household in poverty by over 60%. We also find significant effects on participants' dislike of individuals from the other side of the political spectrum on a feeling thermometer, a commonly used measure in the literature on political polarization (Iyengar et al. 2019). The average treatment effect of over 0.4 standard deviations is equivalent to a reduction in hostility of more than 25% relative to baseline values, or fully offsetting the increase in partisan hostility in Brazil and the US over the last 10 years. Based on these results, the conversation treatment outperforms all but three treatments in a recent mega study, evaluating 25 interventions aimed at reducing partisan hostility (Voelkel et al. 2024), and the majority of contact interventions in a recent meta study (Lowe 2025). The treatment effects on hostility remain economically and statistically significant at the time of the long-term endline survey, which respondents completed three months after the intervention.

In line with the observed average treatment effects on partisan hostility, participants report conversation outcomes that strongly exceed their predictions. This indicates that pessimistic baseline beliefs about the costs and benefits of conversations were largely mistaken. Across all conversation outcomes we surveyed, realized outcomes exceed incentivized predictions by at least

0.8 standard deviations.[2] The wedge between predictions and realized outcomes is largest among participants with high levels of baseline hostility and avoidance.

We find that the impact of political conversations is substantially *larger* among those who are hostile and avoidant of outgroup interactions, suggesting that endogenous segregation in the context of political conversations exacerbates hostility. We find that the treatment effect is more than twice as large among participants with above median levels of baseline hostility. Using an index of outgroup avoidance, measuring the difference in interest in a co-partisan versus cross-partisan political conversation, we find that the reduction in hostility is more than 50% larger for those with above median avoidance. Along similar lines, we find that participants with greater degrees of ideological polarization experience a significantly stronger treatment effect. This heterogeneity also matters for persistence, as treatment effects are more persistent for individuals with high levels of baseline hostility.

In line with these findings, we observe that treatment reduced participants' avoidance of cross-partisan political conversations and improved predictions about future interactions, particularly for those with greater levels of hostility and avoidance at baseline. Over 80% of participants indicate interest in meeting with their assigned conversation partner from the outgroup again. The conversation also reduces outgroup avoidance measured via either the difference in payment required to have a second conversation with an outgroup member versus an ingroup member, or stated interest in political conversations outside of the experiment. Moreover, we find that participants have more positive predictions about future interactions both as part of the experiment and in their day-to-day interactions. The effects on beliefs and avoidance are strongest among participants with high levels of baseline hostility and avoidance. This suggests that interventions that reduce segregation have the potential to generate a virtuous cycle, where exposure corrects pessimistic beliefs about interactions, encouraging further contact in the future.

We find that cross-partisan political conversations with partners who are more hostile or ideologically polarized lead to larger reductions in intergroup hostility. The treatment effect sizes are similar for participants who were randomly matched with a participant with high outgroup avoidance compared to a partner with low avoidance. This speaks against the possibility that engaging more avoidant and hostile individuals in cross-partisan political conversation has unintended negative consequences on their conversation partners, at least in our context.

Taken together, our findings provide support for the hypothesis that endogenous selection into intergroup contact causes a vicious cycle in the context of political polarization in Brazil, which may call for policy intervention. The largest effects of intergroup interactions are concentrated among individuals with greater degrees of baseline hostility and avoidance of outgroup contact. However, since these individuals also experience greater pessimism about the costs and benefits

[2]Across the three conversation outcomes that participants ranked as most important on average before the conversation, realized outcomes exceed predictions by 28 percentage points (learning about partner's political positions), 38 percentage points (building and maintaining a personal relationship), and 53 percentage points (partner is respectful), respectively.

of cross-partisan political conversations ex ante, they are unlikely to engage in them voluntarily.

We use our experimental results to conduct a cost-benefit analysis contrasting alternative incentive and recruitment targeting schemes for interventions facilitating cross-partisan interactions. Our framework considers a designer who seeks to achieve the greatest average reduction in society's hostility, subject to a budget constraint. Our results demonstrate that policy interventions targeting volunteer participants are outperformed by interventions offering modest monetary incentives for participation at any level of budget. This is the case, because such interventions attract participants with greater potential treatment effects and reduce the average recruitment cost per converted participant. We also demonstrate that both volunteer and incentivized interventions can substantially increase their impact by recruitment targeting, focusing advertising on individuals with greater baseline hostility. Even though fewer high-hostility individuals choose to participate at any incentive level, their greater average treatment effects offset increased recruitment costs.

Our results, however, also point to an important limitation of programs facilitating intergroup contact: Individuals with the greatest treatment effects may be prohibitively expensive to recruit. To provide evidence on alternative scalable policies, we benchmark the conversation treatment against a short video intervention. We designed the video intervention with an international non-profit organization, aiming to distill the key ingredients of the conversation treatment into a 3-minute video.[3] We find that the video intervention leads to similar impacts on hostility and beliefs in the short run. As is the case for cross-partisan conversations, the greatest impacts are concentrated among participants with greater baseline hostility and avoidance. These effects extend to participants in the top 20% of the WTA distribution who were not randomized into the conversation treatment because their WTA exceeded the incentives we offered. This suggests that the video constitutes a viable alternative policy, in particular when targeting hostile partisans who are unlikely to participate in contact interventions even with high financial incentives.

The findings in our paper suggest that endogenous selection could play an important role in contexts studied by the extensive literature across the social sciences on intergroup contact (Allport 1954). This literature has examined the impact of intergroup contact on prejudice and hostility, finding mixed results across contexts, creating positive results in some cases, while backfiring in others (Lowe 2025). This literature has typically abstracted away from the endogenous decision to engage in contact and the potential implications of this selection for the treatment effect of contact interventions.

We advance this literature by studying how demand for contact and prior attitudes influence the outcomes of interactions with cross-partisans. These findings help contextualize prior work on contact between individuals from different sides of the political spectrum, drawing on volunteer populations in field settings, and studies recruiting participants for short guided conversations

[3]The video treatment includes testimonials by pilot participants and provides information about conversation outcomes and characteristics of the ingroup and outgroup.

(Blattner and Koenen 2023; Braghieri et al. 2025; Heuser and Stötzer 2022; Levendusky and Stecula 2021; Santoro and Broockman 2022). Our findings also contribute to work that has studied contact between individuals with a different ethnicity or race (Boisjoly et al. 2006; Bursztyn et al. 2021; Finseraas et al. 2019), immigrant background (Bailey et al. 2022), religion (Jha 2013), or socioeconomic status (Rao 2019), individuals from a different caste (Lowe 2021), or the effect of contact between voters and activists, and contact between voters and transgender canvassers (Broockman and Kalla 2016; Kalla and Broockman 2022).

By examining the detrimental role of network homophily and social interactions in driving political polarization, we also contribute causal evidence to a central debate in the literature on the causes and consequences of political polarization (Iyengar et al. 2019). Our work extends research in economics and political science on the drivers of polarization—including cable news (Martin and Yurukoglu 2017), social media (Bail et al. 2018; Levy 2020), internet use (Boxell et al. 2017), misperceptions (Ahler and Sood 2018), and trade exposure (Autor et al. 2020)—and connects to studies on the benefits of diverse networks for economic outcomes (Chetty et al. 2022). In addition, our findings complement a growing interdisciplinary literature evaluating interventions to reduce partisan animosity, such as emphasizing common identities (Levendusky 2018), providing information about outgroup characteristics (Ahler and Sood 2018), and testing scalable short online interventions (Voelkel et al. 2024).

The remainder of this paper is structured as follows. Section 2 describes the experimental design. Section 3 presents descriptive results on partisan segregation in Brazil. Section 4 reports the results from the experiment. Section 5 concludes.

2 Experiment Design

We conduct a field experiment in Brazil between August 2025 and January 2026. The experimental design is summarized in Figure 1.

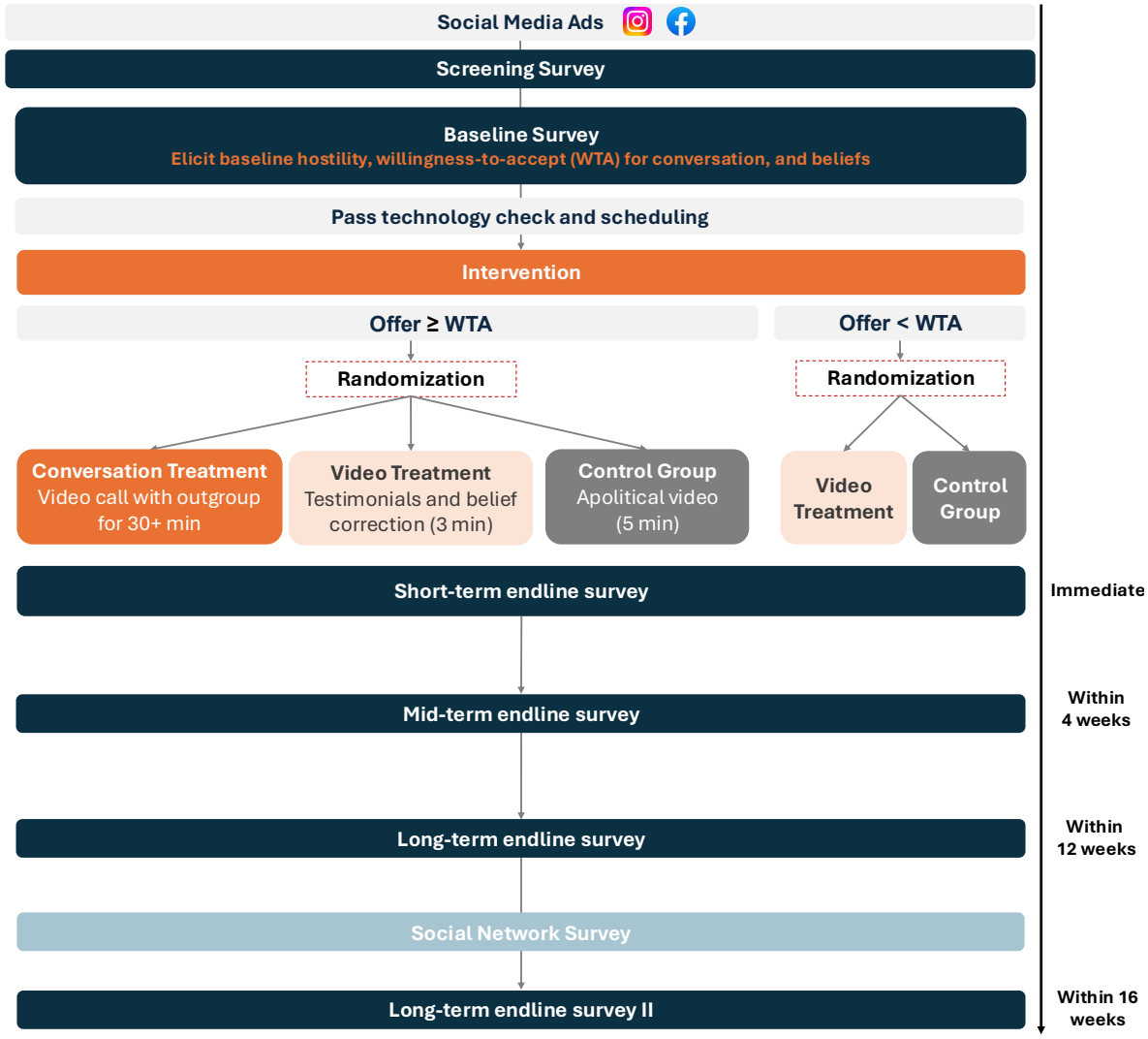
2.1 Experiment Overview

We recruit participants through apolitical advertisements on Instagram and Facebook. We run large-scale advertising campaigns with varying content to target different demographic groups and achieve a balanced sample based on political leaning. More than 4 million Brazilians were shown at least one of our ads. The ads targeted demographic subgroups with the goal of achieving a broadly representative sample. Figure 2 shows the best-performing ads among left and right-wing participants.

At the point of recruitment and screening, participants are not aware that the study focuses on politics or offers the opportunity to participate in cross-partisan conversations.[4] Instead, participants consent to completing a longitudinal study about "Life in Brazil" (4 surveys, weekly

[4]These procedures were approved by Stanford's IRB and a local IRB.

FIGURE 1: Experiment Design



WhatsApp surveys, and 1 short "scheduled" survey on a scheduled date and time) for a gift card compensation of 75 BRL (1.5 times the daily minimum wage).

Upon clicking on the ad, participants are routed to the **Screening Survey**. In the Screening Survey, we confirm participants' eligibility to take part in the study, elicit key demographics and political leaning on the left-right scale, and collect their contact information, including the WhatsApp-connected phone number used for participant communication during the study. Finally, we ask an open-ended question about current challenges in Brazil, where participants found out about the study, and whether they recently participated in other studies.

We screen out participants who do not reside in Brazil, are younger than 18 years old, do not have access to a device with internet access and a camera, do not have a WhatsApp account, fail to pass two-factor authentication for their WhatsApp number, fail to answer one of two attention

FIGURE 2: Best-performing advertisements on Meta platforms

Panel A: Among Left-wing Users | **Panel B: Among Right-wing Users**

Notes: This figure shows ads that had the lowest cost per conversion among left-wing users (Panel A) and right-wing users (Panel B).

TABLE 1: Sample Summary Statistics

	rndzed_any	ps_finished		treat_con	control_con	rndzed_con	treat_nrs	con_nrs	rndzed_nrs
	Full Sample		**Brazilian**	**Eligible for Conversation**			**Eligible for Video Treatment Only**		
	Randomized on Platform	**Completed Endline**	**Adult Population**	**Conv. Treatment**	**Control or Video**	**Balance Test**	**Video Treatment**	**Control**	**Balance Test**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Demographic Characteristics									
Female	0.60	0.60	0.51	0.57	0.56	0.01 (0.67)	0.69	0.72	-0.03 (0.39)
34 years or younger	0.38	0.38	0.35	0.37	0.37	0.00 (1.00)	0.43	0.37	0.06 (0.11)
White	0.57	0.58	0.43	0.57	0.55	0.02 (0.39)	0.61	0.63	-0.02 (0.59)
Protestant or Evangelical	0.26	0.26	0.27	0.26	0.27	-0.01 (0.62)	0.27	0.24	0.03 (0.38)
High School Degree	0.36	0.35	0.31	0.35	0.37	-0.02 (0.37)	0.38	0.33	0.05 (0.18)
University Degree	0.62	0.63	0.25	0.63	0.61	0.02 (0.37)	0.61	0.63	-0.02 (0.59)
Political Characteristics									
Voted Lula	0.47	0.48	0.39	0.45	0.47	-0.02 (0.39)	0.49	0.52	-0.03 (0.44)
Voted Bolsonaro	0.37	0.36	0.37	0.40	0.36	0.04 (0.08)	0.34	0.32	0.02 (0.58)
Left	0.51	0.51	0.29	0.47	0.51	-0.04 (0.09)	0.55	0.56	-0.01 (0.80)
Center	0.08	0.07	0.22	0.08	0.09	-0.01 (0.43)	0.06	0.05	0.01 (0.56)
Right	0.42	0.41	0.49	0.45	0.41	0.04 (0.08)	0.40	0.39	0.01 (0.79)
Ideologically extreme	0.61	0.61		0.61	0.59	0.02 (0.38)	0.67	0.65	0.02 (0.59)
Above Median Hostility	0.49	0.49		0.45	0.46	-0.01 (0.67)	0.56	0.60	-0.04 (0.30)
Above Median Avoidance	0.49	0.49		0.41	0.36	0.05 (0.03)	0.79	0.77	0.02 (0.53)
N	2539	2412		848	1019	1867	336	336	672

Notes: This table shows means, sample sizes, and balance tests for the sample randomized into one of the interventions. Column 1 includes participants who successfully entered the platform on the scheduled date and time, and were subsequently randomized. Column 2 includes participants who completed the short-term endline survey. Column 3 compiles data for the Brazilian adult population from census data and nationally representative surveys. Columns 4 and 5 show means for the sample of participants eligible for the Conversation Treatment. Columns 7 through 8 show means for the sample of participants eligible for the video treatment only. Columns 6 and 9 show the difference in means between treatment and control groups, and the p-value of a t-test between the two samples (in parentheses).

checks correctly, or access the study through a link shared by another participant.[5] About 19,000

[5]Each time a participant opens the survey link, the link is automatically appended with a URL variable indicating

participants passed these screening checks and consented to participating in the study.

Eligible participants received a link to the **Baseline Survey** via WhatsApp within approximately one week of completing the Screening Survey.[6] The baseline survey elicited participants' *political identity*, *outgroup avoidance*, *hostility* towards the political outgroup, and *beliefs about conversations*' costs and benefits (see Section 2.2 for details).

The Baseline Survey also includes a short demographics section, questions about participants' level of political interest, and positions on 10 divisive political topics. Finally, we survey participants' *political segregation*, by asking questions about their five closest network contacts, such as friends, family, and colleagues. Participants indicate their relationship with each network contact, their perceived political affiliation, race and gender, as well as how often they talk in general and discuss politics. Participants then report how long it has been since they discussed politics with the political ingroup and outgroup, and since they last spoke to a member of the political outgroup about any topic. The Baseline Survey included multiple attention checks. Participants who failed more than 3 attention checks were screened out and not allowed to proceed with the rest of the survey.

At the end of the Baseline Survey, participants complete a **Technology Check** to ensure that their device and network are compatible with our video call platform. Upon completing the Technology Check, participants are offered several dates and times to complete a **Scheduled Survey**. Participants are assigned dates for the Scheduled Survey within two days (on average) of baseline completion. The assignment algorithm was designed to maintain balance of left and right participants assigned to each date. If participants cancel or fail to show up on the assigned dates, additional dates are offered via WhatsApp.

At the start of the Scheduled Survey, participants are randomized into one of three treatments for an additional payment: (i) Outgroup Conversation Treatment, (ii) Video Intervention Treatment, and (iii) Control Treatment (see Section 2.3). The treatment assignment procedure depends on participants' *willingness to accept* for a 30-minute political conversation (WTA_{PO}) elicited in the Baseline Survey relative to the *Offer* of R$ 104 for completing the Outgroup Conversation Treatment:[7]

- **Offer** $\geq WTA_{PO}$**:** Participants satisfying this condition are randomized into the Outgroup Conversation Treatment ($p = 0.5$), the Video Intervention Treatment ($p = 0.125$) or the

that the link was previously used. Subsequent potential participants who use a link with this variable are screened out. Additionally, any participant without a valid Facebook click ID (passed on from Meta ads) is screened out.

[6]To ensure balanced representation across the political spectrum in the conversation treatment, we invited participants to the baseline survey based on quotas. In each invite batch, we aimed to invite 45% of the recipients from the right of the political spectrum, 45% from the left, and 10% from the center, as elicited in the screening survey.

[7]To ensure incentive compatibility of the WTA elicitation in the baseline survey, this offer was equal to R$ 104 with probability 0.99. With probability 0.01, it was drawn from the interval between 0 and 200. To incentivize negative WTA values, we select one participant at the start of the baseline survey to receive an additional gift card of R$ 200 after completing the study. For this participant, the offer was drawn from the interval between -200 and 200. Negative offers are subtracted from the additional gift card.

Control ($p = 0.375$).[8] Treatment assignment was stratified by political leaning (left vs. right) and avoidance index (above vs. below median). Participants in the Conversation Treatment receive the offered gift-card of R$ 104 if they complete the treatment. Participants in the other two conditions receive a smaller gift card of R$ 50 for compliance.[9]

- **Offer** $< WTA_{PO}$**:** Participants satisfying this condition are randomized into the Video Intervention Treatment ($p = 0.5$) or the Control ($p = 0.5$).

This assignment procedure is designed to minimize differential attrition by treatment and by baseline outgroup avoidance. In particular, participants are only ever randomized into the conversation treatment for an offer they find attractive according to their WTA. As a result, it is in their best interest to simply indicate their true interest in the conversation treatment in the WTA elicitation, rather than drop out of the study at a later date. Figure A1 shows that this procedure worked as intended - the WTA distribution among participants who answered this question in Baseline is almost identical to the WTA distribution among those who were randomized into treatments.

A second important feature of this assignment procedure is that randomization only takes place after participants join the platform, complete a second technology check, and confirm their understanding of the procedures. This allows us to further minimize differential attrition by treatment.[10]

After participants complete their assigned treatment, they immediately fill out a **Short-Term Endline Survey**. In this survey, we measure participants' *political leaning* and *hostility*.[11] Additionally, we ask participants to recollect details of their intervention experience. In the Conversation Treatment, participants are asked to recall the name and characteristics of their partner, and to report how much of the conversation was spent discussing politics and what topics were discussed. Participants in the Control and Video Intervention Treatment answer questions about the content of the videos.

Participants in the Conversation Treatment report their conversation outcomes. In particular, they report whether they were persuaded, whether they acquired political knowledge, whether they learned about their partner's political positions and reasons for those positions, whether they maintained a good reputation, and whether their partner listened and came across as respectful.

[8] The video intervention was waiting for approval during the first week of experimental roll-out. As a result, in the first week, the probability of the Conversation Treatment was set to $p = 0.6$ and Control to $p = 0.4$.

[9] In the first phase of the experimental roll-out, the gift card was instead R$ 20. This amount was increased to R$ 50 to minimize attrition concerns out of the control group.

[10] To ensure that participants set aside sufficient time to complete the conversation treatment, we inform participants whether they are eligible to complete an outgroup conversation at the end of the Baseline Survey, along with the offer on the table. Participants are also informed that a subset of participants will be selected to watch a 5-minute video and receive a smaller gift card instead (Video Intervention and Control).

[11] We initially planned to only include the *cash transfer decision* in later surveys. After observing that the Short-Term Endline Survey takes less time to complete than anticipated, we added this decision to the survey to capture short-term treatment effects.

Additionally, they report their beliefs about their partner's answers to these questions. Finally, they guess their partner's position on the left-right political spectrum, report feelings towards the partner on a feeling thermometer, and indicate how representative of the political outgroup they believe their partner is in terms of political positions and personal characteristics. In place of this module, participants in the Control and Video Intervention Treatment were asked several questions about their opinion on the video they watched, including how enjoyable, informative, and partisan the video was.

The **Mid-term Endline Survey** was administered up to 4 weeks after the intervention date. The survey tracks the persistence of treatment effects on *hostility* and on *outgroup avoidance* and *beliefs about conversations*. The survey also studies treatment effects on downstream outcomes of hostility, including participants' anti-democratic attitudes, support for political violence, and attitudes regarding political discrimination in the workplace. Additionally, we include several incentivized questions about the political views and characteristics of the political outgroup and ingroup to provide evidence on mechanisms. To track potential network effects, we ask participants several questions about recent conversations with their five closest network contacts and whether they discussed the intervention they completed with them. Finally, we ask participants about their recollection of their conversation, video intervention, or control video.

We track the persistence of treatment effects in two additional **Long-term Endline Surveys**, administered over the course of the next four months. These surveys additionally ask participants about their recent conversations with outgroup members outside of the experiment.

2.2 Main Survey Measures

2.2.1 Political Identity

We use participants' placement on the left-right scale to define their *political ingroup* and *outgroup*. In particular, we assign participants to the "left" political group if they choose to place themselves between 0 (totally left) and 4 (center-left) or if they select 5 (center) and "center-left" in the follow-up question. Otherwise, we assign them to the "right" political group.[12] We choose the left-right scale as the determinant of participants' political groups, because Brazilian participants exhibited low identification with individual parties and politicians but higher identification with the left and right wing of the political spectrum in piloting.

2.2.2 Outgroup avoidance

We collect two main measures of *outgroup avoidance* - the preference for avoiding political conversations with the outgroup relative to the ingroup. For our main analysis, the two measures are

[12] In survey questions, which refer to people from the left and people from the right, participants are only asked to consider individuals who place themselves from 0-4 and 6-10, respectively. We also exclude participants from the main analysis who do not choose a political leaning initially.

standardized and combined in a pre-registered index.

The first measure of *outgroup avoidance* leverages *participant's willingness to accept (WTA)* to complete each of the following two activities:

1. WTA_{PO}: 30-minute conversation about political topics with *outgroup*
2. WTA_{PI}: 30-minute conversation about political topics with *ingroup*

The WTA_{PO} measures participants' inverse demand for cross-partisan political conversations. We define outgroup avoidance as the difference between the WTA for the two tasks above: $WTA_{PO} - WTA_{PI}$.

We elicit WTA using an incentive-compatible dynamic version of the multiple-price list. At each step of the procedure, participants make binary choice between completing the task for a given payment and not completing the task. We repeat the procedure until we identify the lower amount between 0 and 200 BRL for which the participant completes the task. If a participant indicates that they would complete the task for free, we elicit their *willingness to pay* to complete the activity using a similar procedure. If a participant is not willing to complete the study for 200 BRL, we ask them to state their WTA in a box (unincentivized). In addition to the two tasks above, we also elicit participants' WTA to complete either a co-partisan or cross-partisan conversation about non-political topics ($p(WTA_{PO})$) in the baseline survey. The order of the elicitations is randomized and participants undertake extensive training before completing them. Details on how we incentivize the WTA elicitation in each survey are provided below.

The second measure of *outgroup avoidance* is constructed using participants' self-reported interest in political conversations outside the experiment. In particular, participants indicate whether they agree on a scale from 0 (Completely disagree) to 9 (Completely agree) with the statement "I am interested in engaging in conversations about politics with [outgroup/ingroup]". Our second measure of *outgroup avoidance* is defined as the standardized difference between their interest in a cross vs. a co-partisan political conversation.

2.2.3 Hostility

We define *hostility* towards the outgroup as negative attitudes and behavior, ranging from dislike to discrimination in economic decisions. Our surveys contain two main measures of hostility.

Our first measure of *hostility* measures participants' discrimination in an incentivized *cash transfer decision*. In this decision, participants make a series of incentivized binary choices between two individuals, choosing who they would like to receive a sizeable cash transfer. Profiles of the two potential recipients indicate their political affiliation and their monthly per capita household income. The profile of one of the potential recipients is fixed to be a member of the political *ingroup* with a monthly per capita household income of R$4.000 or R$3.000.[13] The other potential recipient

[13] A per capita household income of R$4.000 corresponds to roughly the 90th percentile of the national income dis-

is a member of the political *outgroup* with varying levels of income. In each choice, the participant therefore trades of her preference against allocating transfers to their political outgroup with her preference for allocating transfers to poorer households.[14]

The procedure is repeated until we identify the participant's switching point - the highest level of *outgroup* income, at which the respondent allocates the transfer to the poorer *outgroup* rather than the *ingroup*.[15] Greater *hostility* towards the outgroup implies a lower switching point. Across participants, we randomize three potential cash transfer amounts with equal probability: R$100, R$300, and R$700. The lowest cash transfer equals about 2 days' work at the current minimum wage. The highest cash transfer equals roughly the average monthly payment to households in Bolsa Familia, a large-scale social welfare program received by 20 million families.[16] To incentivize the *cash transfer decision*, we randomly draw several participants and implement their elicited preferences for recipients in an external sample.[17]

This procedure is a novel extension of the dictator game, which is often used to measure economic discrimination. One major advantage of the *cash transfer decision* is that it substantially reduces the salience of the 50-50 norm – the well-documented tendency to allocate money equally between any two individuals, regardless of one's true social preferences.

Our second measure of *hostility* leverages participants' reported attitudes towards individuals on the "left" and "right" of the political spectrum on a feeling thermometer. Following prior literature, we define participants' *hostility* as the difference between the ingroup and outgroup thermometer scores. The feeling thermometer is commonly used in the political science literature to measure *affective polarization* - the dislike of the political outgroup relative to the ingroup. As a result, this measure allows us to benchmark our results to other studies.

All Endline Surveys contain both measures of participants' hostility.[18] To economize survey time, the Baseline Survey only elicits participants' baseline hostility using the feeling thermometer.

tribution and the 80th percentile of the income distribution in the capital São Paulo. A per capita household income of R$3.000 corresponds to roughly the 87th percentile of the national income distribution and the 70th percentile of the income distribution in the capital São Paulo.

[14]We confirm whether participants indeed prefer allocating cash transfers to poorer households by eliciting two rounds of decisions that ask participants to choose between two recipients with varying levels of income who are both from the outgroup.

[15]We search over a grid of R$100 steps, ranging from R$300 to R$6000. If a participant rejects allocating the cash transfer to the outgroup at R$300, we present them with 4 additional choices in which we increase the income of the ingroup member to 5000, 8000, 10000, and 20000. We use the staircase method to reduce the number of choices shown to the participant.

[16]Source: Government of Brazil - *Bolsa Família chega a 20,8 milhões de beneficiários a partir desta quarta*

[17]Namely, we select two potential recipients from the external sample - one from the political ingroup and one from the outgroup. We then compare the incomes of these potential recipients and allocate the transfer according to the switching point estimated in the *cash transfer decision*.

[18]We initially planned to only include the *cash transfer decision* in later surveys. After observing that the Short-Term Endline Survey takes less time to complete than anticipated, we added this decision to the sample to capture short-term treatment effects.

2.2.4 Beliefs about Conversations

To measure participants beliefs about the costs and benefits of political conversations, we elicit participants incentivizied beliefs about the following conversations outcomes: partner behavior (whether the partner listens and shows respect), reputation (whether they will maintain a positive reputation with the partner), learning about the partner's political opinions and the reasons behind them (for self and for partner), acquiring knowledge about politics and current events (for self and for partner), and persuasion (for self and for partner). In particular, we ask participant to guess the average level of agreement with statements about the 9 conversation outcomes participants on the left (right) will report after a conversation with participants on the right (left) in our study. To incentivize this elicitation, several participants are selected to receive an additional bonus of R$ 10 for every question in which they correctly guess the average response of participants after their conversation. To strengthen the external relevance of our findings, we also survey participants' beliefs about the likelihood of each conversation outcome in cross- and co-partisan conversations in their daily lives. These questions are included in Mid-Term and Long-Term Enline Surveys.

We focus on these 9 conversations outcomes because they were either (i) frequently mentioned in free-form questions as important aspects of conversations in pilots, or (ii) explored in related literature (e.g. see Braghieri et al. (2025)). To gauge what relative importance participants place on each of the outcomes, we asked them to rank their importance in the baseline survey.[19] We elicit participants' preference ranking for political conversations with (i) an outgroup stranger, and one of either (ii) an ingroup stranger, (iii) an outgroup network contact, or (iv) an ingroup network contact.

2.2.5 Downstream Outcomes

In Endline Surveys, we measure treatment effects on several downstream outcomes, related to intergroup hostility. First, we elicit participants' attitudes towards discrimination based on political identity in labor markets. In particular, we ask participants whether they find hiring discrimination based on political affiliation acceptable. We also ask several questions, which measure the extent to which participants prefer to work with colleagues from the political ingroup over the outgroup. The Midterm Endline Survey additionally includes an incentivized co-worker choice, in which participants trade off potential coworkers' political affiliation against their productivity.

Second, we measure participants' anti-democratic attitudes and their support for political violence. In line with the political science literature (see Voelkel et al. (2024)), we measure anti-democratic attitudes by asking participants to report the extent to which they agree with ingroup politicians ignoring unfavorable court rulings, prosecuting critical journalists, not accepting elec-

[19] In particular, participants sequentially choose the 2 most important outcomes out of all remaining options, until the list is exhausted.

tion losses, blocking all proposals of other parties, and diverting humanitarian aid away from disaster victims who did not vote for them. We measure support for political violence by asking participants whether they consider it justifiable to send threatening messages to outgroup politicians, to harass ordinary outgroup voters on the internet, to resort to violence to promote ingroup political goals and oppose outgroup election victories, or to use force to suppress outgroup protests. The Midterm Endline Survey additionally includes an incentivized politician choice, in which participants trade off politicians' ideological alignment with antidemocratic attitudes and support for political violence.

2.3 Treatments

Participants are randomized into one of the following three treatments:

- Outgroup Conversation Treatment: A video-call conversation about politics with a member of the political outgroup.
- Video Treatment: A short video presenting testimonials of participants who completed an outgroup conversation, and information correcting misperceptions about the political outgroup.
- Control Group: A short apolitical video about cooking or travel.

2.3.1 Outgroup Conversation Treatment

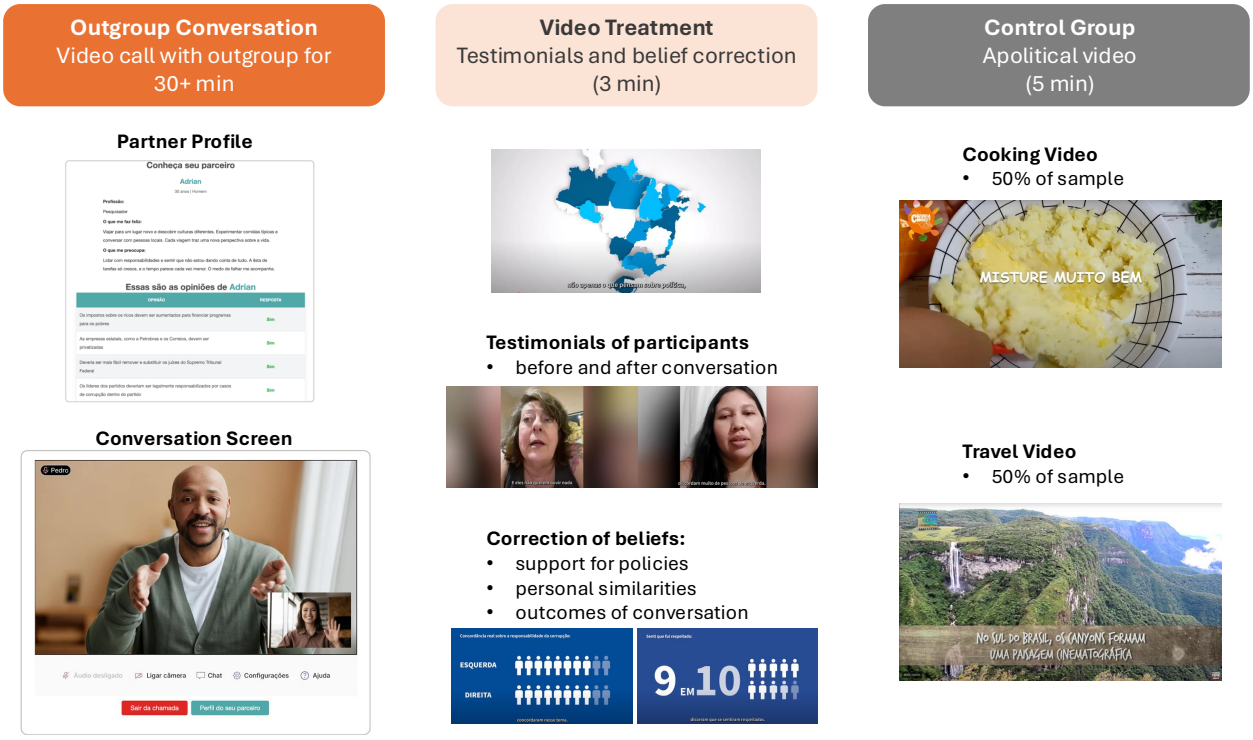
Participants in Conversation Treatment are paired with a conversation partner on the other side of the political spectrum.[20] The matching algorithm matches cross-partisans into pairs randomly, while ensuring maximum possible balance of matches by above and below median avoidance index. Any unassigned participants are reassigned into control after matching. This is the case, in particular, if the number of participants assigned into treatment is odd, or if one side of the political spectrum is overrepresented on a particular event date.

After matching, participants have two minutes to read a profile of their assigned partner, summarizing their demographic background, answers two open-ended questions about what makes participants happy and worried, and 8 political opinions.[21] Next, participants begin the video conversation with their partner. If they wish, they can reopen the profile of their partner at any point in the call by clicking a button. Participants are informed that their conversation needs to last at least 30 minutes to count as completed, but they are free to talk longer. After participants

[20] Participant data is exported into Google Sheets via API requests. The matching algorithm is written as a Google Script that runs 8 minutes after the scheduled time of the activity automatically. Information about matches is then sent back to Qualtrics via API requests. To manage high loads and simultaneous API requests, Firestore is used as an intermediary between Google Sheets and Qualtrics.

[21] These political opinions are drawn randomly from a set of 10 questions asked in the baseline and displayed in random order on the pair level.

FIGURE 3: Treatment Overview



Notes: This figure illustrates the treatments administered in the experiment.

complete their conversation, they click on a button that redirects them to a Short-Term Endline Survey.[22]

2.3.2 Video Intervention Treatment

Participants in the Video Intervention Treatment watch a 5-minute video that we developed jointly with the non-profit More in Common, addressing common misperceptions about political opponents and conversations with them. Participants are informed that we will ask follow-up questions about the content of the video, which participants will need to pass in order to be eligible for payment. The video intervention is modeled after state-of-the-art interventions found effective at reducing partisan hostility (affective polarization) in recent political science research (Voelkel et al. 2024). The Video Intervention Treatment provides a benchmark for the impact of alternative interventions to reduce partisan hostility in our sample.

[22]Participants' progress through the platform and the video conversation are monitored using a live dashboard. The research team reaches out to participants experiencing technical issues via WhatsApp. If a participant experiences technical issues that prevent them from joining the conversation, we offer them and their partner the opportunity to reschedule. Participants who refuse to reschedule are instructed to proceed to the Short-Term Endline Survey. If a participant leaves the platform only after they have entered the video call with their partner, we instruct both participants to proceed to the Short-Term Endline Survey. This conservative procedure lowers the chance that disgruntled participants who had a negative conversation experience drop out of the study at higher rates than other participants.

2.3.3 Control Treatment

Participants in the Control Treatment watch a 5-minute video on a non-political topic. Participants are informed that we will ask follow-up questions about the content of the video, which participants will need to pass in order to be eligible for payment. Participants are randomized into one of two different non-political videos on travel or cooking.

3 Political Segregation and Hostility

In this section, we provide descriptive evidence on political segregation, avoidance of cross-partisan political conversations and its drivers, using data collected from a Baseline Survey administered to over 5,000 Brazilians. We find that participants' social networks are politically segregated and that participants have limited exposure to cross-partisan political conversations, which strongly predicts their degree of partisan hostility. We further document that participants avoid cross-partisan interactions when presented with the opportunity to engage in them. In particular, the average participant needs to be paid twice as much to complete a conversation with the political outgroup relative to a co-partisan. This result is driven by participants' pessimistic beliefs about the outcomes and experience of political conversations with the outgroup.

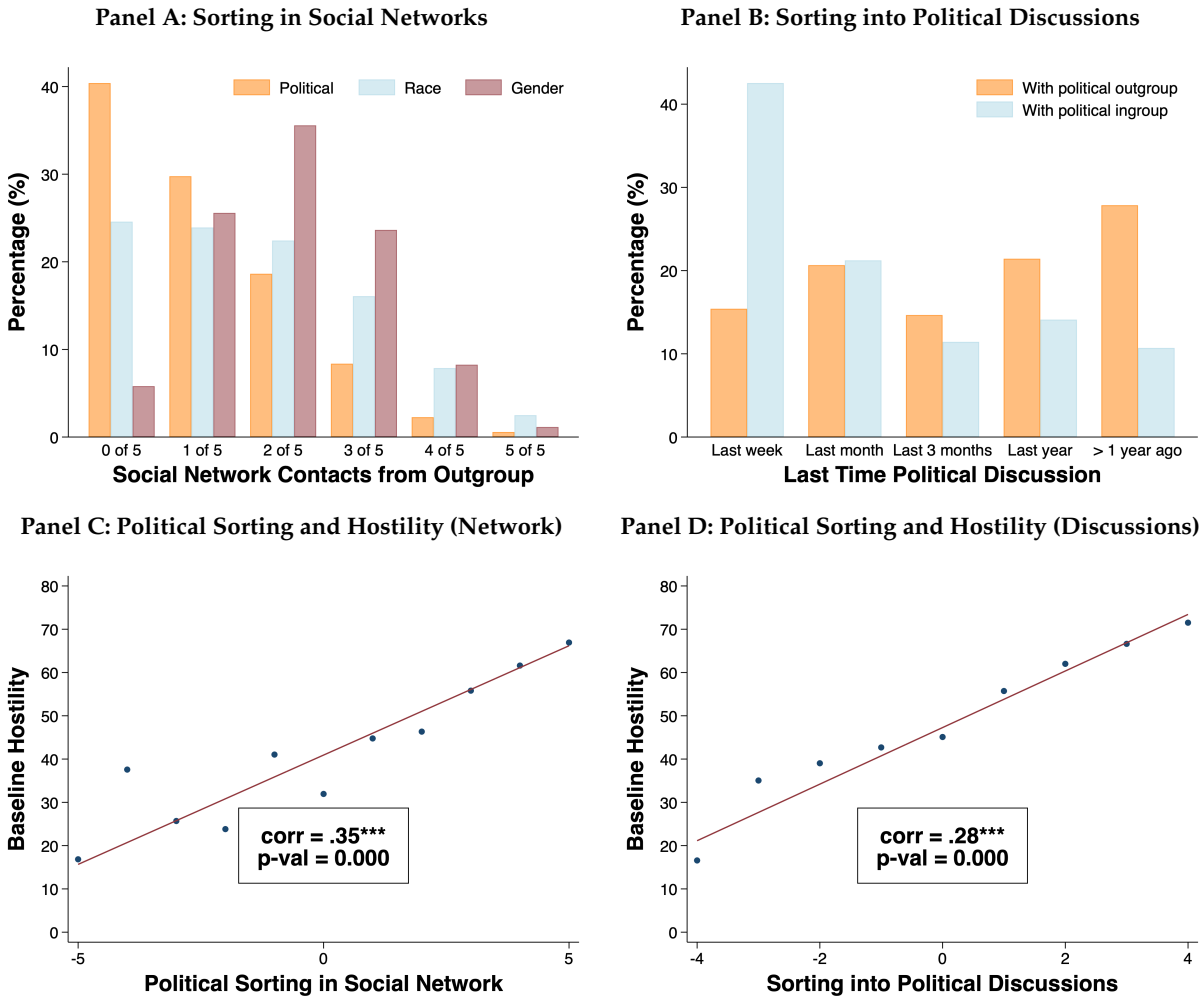
3.1 Descriptive Evidence on Political Segregation

Participants' social networks are politically segregated and their exposure to cross-partisan political conversations is limited. Panel A of Figure 4 plots the segregation of participants' social networks along political lines, race, and gender. The figure demonstrates that segregation along political lines is particularly strong in Brazil - more than 40% of participants indicate that none of their 5 closest contacts is from the other side of the political spectrum. The degree of political sorting outweighs sorting based on other demographic dimensions: only 25% of the sample have no social network contacts from another race. This figure is just 5% for gender.

Panel B of Figure 4 plots when participants last engaged in a political discussion with the outgroup relative to co-partisans. While over 40% of the sample discussed politics with their political ingroup in the last week, only 15% have done so with their political outgroup. Moreover, for approximately 30% of the sample, the last political conversation with the political outgroup took place over a year ago.

Participants' exposure to cross-partisan interactions strongly correlates with their degree of hostility. Panels C and D of Figure 4 present binned scatter plots of participants' baseline hostility against the degree of political sorting of participants' social networks (measured as the difference between the number of ingroup and outgroup friends among their five closest contacts) and the degree of sorting into political conversations (measured as the difference between the time of the last political conversation with an outgroup and an ingroup member). Panel C shows that greater

FIGURE 4: Political Sorting and Hostility



Notes: This figure plots the degree of political, racial, and gender sorting in participants' social networks (Panel A), sorting into political discussions (Panel B), and the correlation with baseline hostility (Panels C and D). Participants were asked to list up to five network contacts with whom they frequently talk about important topics. Ingroups and outgroups are defined based on binary categories of political leaning (left or right), race (white or non-white), and gender (female or non-female). Political sorting in social networks is defined as the difference between the number of ingroup and outgroup friends. Sorting into political discussions is defined as the difference between the time of the last political conversation with an outgroup and an ingroup member. Hostility is defined as the difference in feelings between ingroup and outgroup members.

degrees of political segregation are strongly correlated with baseline hostility ($\rho = 0.35$, p-val < 0.000), measured as the difference in feeling ratings between the ingroup and the outgroup. Similarly, Panel D confirms a similarly strong correlation between participants' level of hostility and the timing of their most recent co-partisan versus cross-partisan political conversation ($\rho = 0.28$, p-val <0.000). Panels C and D of Figure A2 illustrate that these correlations are similarly strong for ideological extremity.

3.2 Outgroup Avoidance, Hostility, and Beliefs

Participants exhibit substantial avoidance of cross-partisan compared to co-partisan political conversations. We demonstrate that participants' avoidance can be accounted for by the fact that participants expect cross-partisan conversations to result in worse outcomes and conversation experience.

The descriptive evidence on existing levels of segregation presented above underscores the relevance of segregation. It, however, does not allow us to directly speak to the drivers of endogenous selection into interactions, as measures of current levels of segregation are driven by both endogenous and exogenous factors.

We thus capture outgroup avoidance by combining two separate measures. First, we draw on a simple survey measure, capturing differential self-reported interest in political conversations in day-to-day interactions. Second, we draw on the incentivized WTA values elicited in the baseline surveys to capture avoidance of cross-partisan conversations in the experiment. Our outgroup avoidance measure combines these two separate measures in a standardized index.[23]

On average, participants prefer discussing non-political topics to politics, and avoid non-political discussion with the political outgroup. Panel A of Figure 5 provides a decomposition of participants' demand for cross-partisan political conversations, measured using an incentivized WTA elicitation for four additional 30-minute activities conducted on an online video platform.

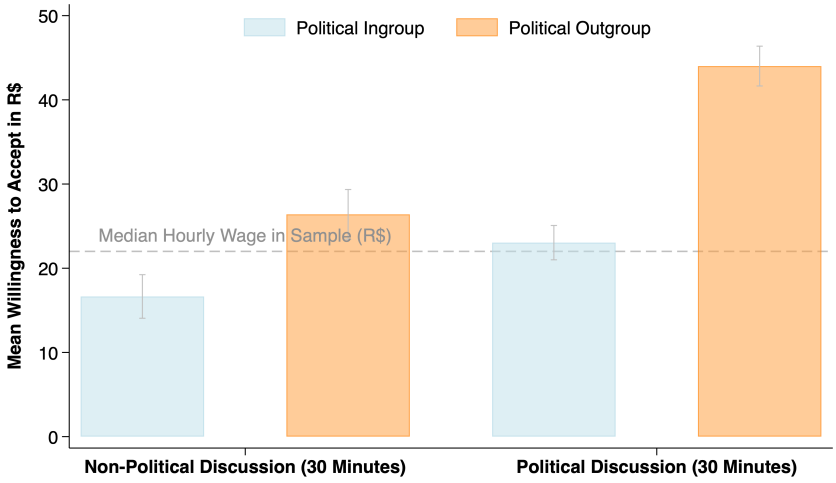
The figure demonstrates that participants, on average, dislike discussing political compared to non-political topics in general. Moreover, participants dislike discussing non-political topics with their political outgroup compared to the ingroup. Participants exhibit a strong avoidance of political conversations with the outgroup relative to alternative activities. On average, the WTA for a cross-partisan political conversation is almost more than twice as large as the WTA for a political conversation with the ingroup and approximately 70% larger compared to a non-political conversation with the outgroup. Notably, participants' average WTA for a 30-minute cross-partisan conversation substantially exceeds the mean hourly wage in the sample, which is not the case for other conversations. Panel B of Figure 5 plots the distribution of the WTA for a 30-minute political discussion.

Figure A4 demonstrates that avoidance of political conversations predicts hostility: polarized individuals are more avoidant of cross-partisan conversations about politics, particularly when compared to co-partisan political conversations. This figure plots the correlation of hostility with interest in discussing politics with the outgroup relative to the ingroup, the differential WTA to

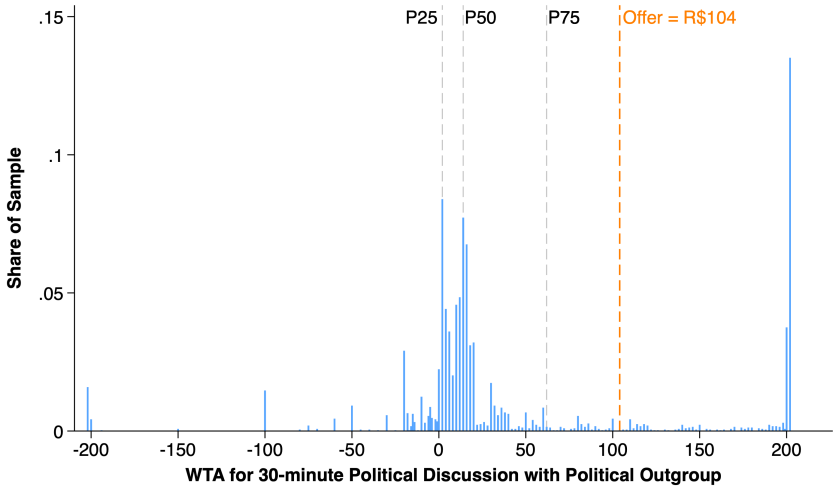
[23] To construct the index, we first calculate the difference between interest in ingroup conversations and interest in outgroup conversations outside of the experiment (both elicited on a scale ranging from 0 to 9). We then calculate the difference between the WTA for a political conversation with a cross-partisan and the WTA for a political conversation with a co-partisan. We standardize each of the difference measures by subtracting the sample mean and dividing by the standard deviation. We then take a simple average betweeen the two standardized difference measures and re-standardize the average.

FIGURE 5: Demand for Conversations in Experiment

Panel A: Decomposition of Demand



Panel B: Distribution of WTA for Political Conversation with Outgroup



Notes: This figure plots the average WTA for non-political and political conversations (Panel A) and a histogram of the WTA for a political conversation with the outgroup (Panel B). Values in both panels are based on WTA distributions, where values below 200 are coded as -202 and values above 200 are coded as 202. The dashed orange line in Panel B plots the modal offer in the experiment, offered to 99% of participants in the Conversation Treatment.

discuss politics with the outgroup compared to the ingroup, and an outgroup avoidance index that combines the standardized values of these measures. This measure of outgroup avoidance has a correlation of $\rho = 0.44$ with baseline levels of hostility (p-val$<$0.000).

To explain why individuals avoid cross-partisan political conversations, we first study participants' preferences over outcomes of political conversations. In the baseline survey, participants rank the subjective importance of nine potential conversation outcomes and experiences: persuasion (for self and for partner), acquiring knowledge about politics and current events (for self

and for partner), learning about the partner's political opinions and the reasons behind them (for self and for partner), maintaining a good reputation with their partner, and partner behavior (whether the partner listens and shows respect). We focus on these motives because they are either proposed as drivers of conversations in the social science literature or they were frequently mentioned in free-form responses in pilot surveys.

In cross-partisan conversations, participants value partners' respect, learning about the partner's positions and the reasons behind them, and building/maintaining a positive reputation. By contrast, they do not value being persuaded or persuading their partners. This can be seen in Panel A of Figure A5, which plots the share of participants indicating that a given motive ranks among their top two most-preferred conversation outcomes in political conversations with the outgroup. Notably, partner's respect is ranked as the most important outcome by over 65% of the participants, almost double the rate of the next most frequently mentioned outcome.

In political conversations with the ingroup, participants are less concerned about how their partner behaves and whether the partner learns or changes their opinion. By contrast, participants are more interested in maintaining a good reputation, acquiring political knowledge, and being persuaded. This can be seen in Panel B of Figure A5.

Across the conversation motives elicited above, participants believe that political conversations with cross-partisans will fare significantly worse compared to conversations with co-partisans. In the baseline survey, we elicited participants' incentivized beliefs about each of the nine outcomes of cross and co-partisan political conversations. In particular, participants were asked to indicate how strongly others agree, on average, with several statements after a 30-minute conversation about politics with a stranger from the political outgroup or political ingroup.

Figure 6 shows that, on average, participants believe that political conversations with a co-partisan stranger are likely to achieve the conversation outcomes we survey. In contrast, participants are much more pessimistic about the outcomes of political conversations with a stranger from the political outgroup.

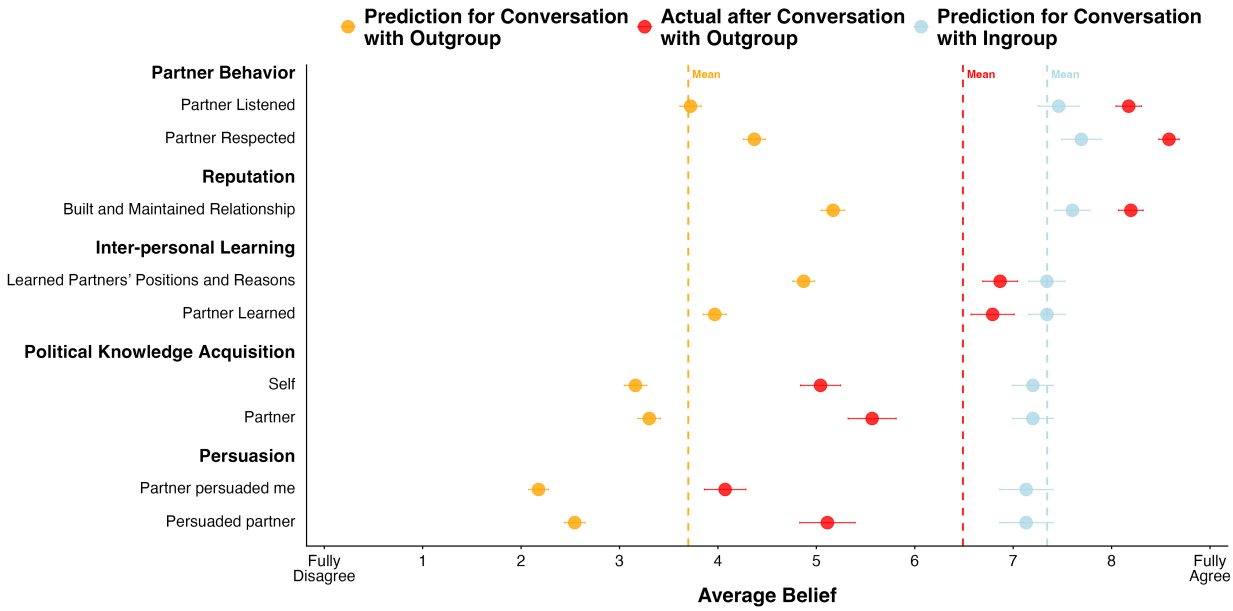
Participants' predictions of conversation outcomes are strongly correlated with both their degree of baseline hostility and of outgroup avoidance. Panel A of Figure A6 plots participants outgroup hostility (on the left) and outgroup avoidance (on the right) against predictions of conversation outcomes in the experiment. Partisans who exhibit strong hostility and outgroup avoidance are most pessimistic about the outcomes of interactions in the experiment. They also are more likely to predict that conversations with the outgroup will not lead to more positive views about the outgroup among ingroup members.

4 The Impact of Cross-partisan Political Conversations

4.1 Qualitative Outcomes of Interactions

Out of the participants assigned to the conversation treatment, 89% complied with their assigned treatment incentive and stayed in the conversation for at least 30 minutes. More than 40% of participants assigned to the conversation treatment group stayed in the conversation for over an hour (twice the duration we incentivized).

FIGURE 6: Beliefs and Actual Outcomes of Political Conversations



Notes: This figure plots incentivized predictions of ingroup members' outcomes of conversations with an outgroup member (orange) and of conversations with another ingroup member (blue). Participants were asked to predict the average response of ingroup members after the respective conversations on a scale from 0 (Fully disagree) to 9 (Fully agree). The red line plots actual average outcomes reported after conversations with an outgroup member. The dashed lines plot the mean belief for each series.

Self-reports indicate that the outcomes of cross-partisan conversations are significantly better than participants anticipated. Figure 6 reports self-reported outcomes of over 760 participants who completed the conversation intervention against the beliefs elicited prior to the intervention. Across all conversation outcomes for which we elicited incentivized beliefs prior to the conversation, realized outcomes significantly exceed prior expectations.

4.2 Empirical Strategy

Our main specification to assess the impact of the two interventions in the sample of participants for whom $\Omega_{PO} \geq WTA_{PO}$ has the following form:

$$Y_{it} = \tau_t^C C_{it} + \tau_t^V V_{it} + \gamma_t \mathbf{X}_i + \epsilon_{it}, \quad (1)$$

where Y_{it} is the outcome of individual i in the endline survey of a given period t (or a WhatsApp survey), C is an indicator for the conversation treatment group, V is an indicator for the video intervention, and X is a vector of control variables, including baseline survey measures of affective and ideological polarization, a demand index, demographic controls, and the event date.[24]

In the sample of participants for whom $\Omega_{PO} < WTA_{PO}$, this equation is simplified to:

$$Y_{it} = \tau_t^V V_{it} + \gamma_t \mathbf{X}_i + \epsilon_{it}, \quad (2)$$

4.3 Average Impact on Hostility

We find that cross-partisan political conversations lead to statistically and economically significant reductions in hostility. Figure 7 plots treatment effect of cross-partisan political conversations on hostility in the short-term endline, filled out immediately after participants complete the intervention. Panel A illustrates that outgroup discrimination is significantly reduced in the incentivized cash transfer decision. The share of participants preferring to allocate a cash transfer to an affluent ingroup household (70th or 80th percentile of the income distribution) instead of an outgroup household in poverty (5th percentile of income distribution) is reduced from 11% to 4%. The average effect amounts to a reduction in discrimination of 0.21 SD.

Panel B demonstrates that conversations also reduce hostility as measured in the feeling thermometer, a survey measure commonly used in the political science literature (Iyengar et al. 2019). Treated participants rate outgroup members significantly higher and give somewhat lower ratings to the ingroup. The average impact on feeling differences between the ingroup and outgroup of 0.4 standard deviations is equivalent to a reduction of more than 25% relative to the control group, or offsetting the increase in affective polarization in Brazil over the last 10 years.

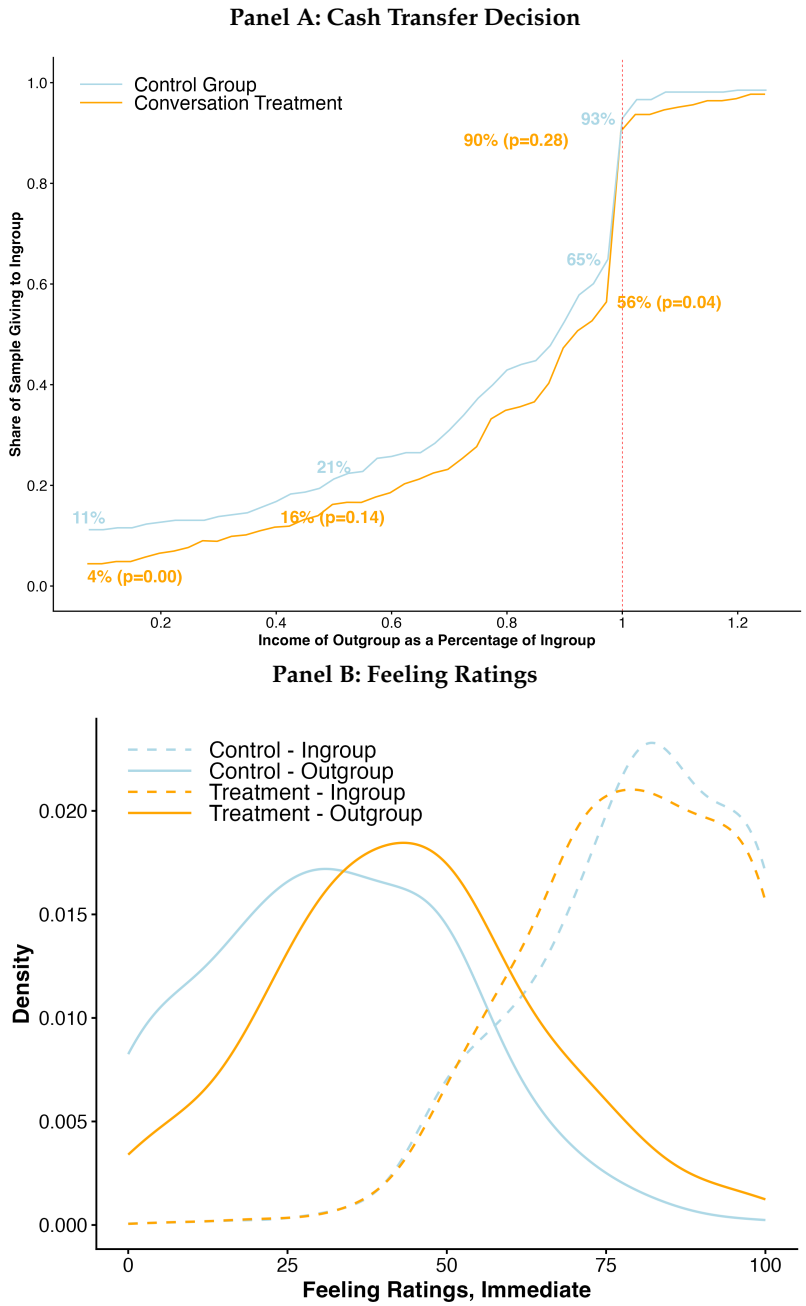
4.4 Effects among High-avoidance and High-hostility Partisans

We find strong evidence that incentivizing contact among those who avoid it is beneficial: the treatment effects of political conversations with the outgroup are strongest among individuals who exhibit *higher* levels of hostility at baseline and are *more* avoidant of cross-partisan political conversations. Moreover, partners of participants with high levels of hostility and avoidance of interaction have similar treatment effects.

Figure 8 plots the effect of conversations on hostility by baseline hostility and outgroup avoidance. Treatment effects are more than twice as large among participants with above median levels of baseline hostility, relative to those less hostile at baseline. Echoing these results, we find that

[24] We include a dummy equal to one for those who were randomly matched to participants who are neither politically left nor politically right. Our findings are robust to including these matches in the treatment group.

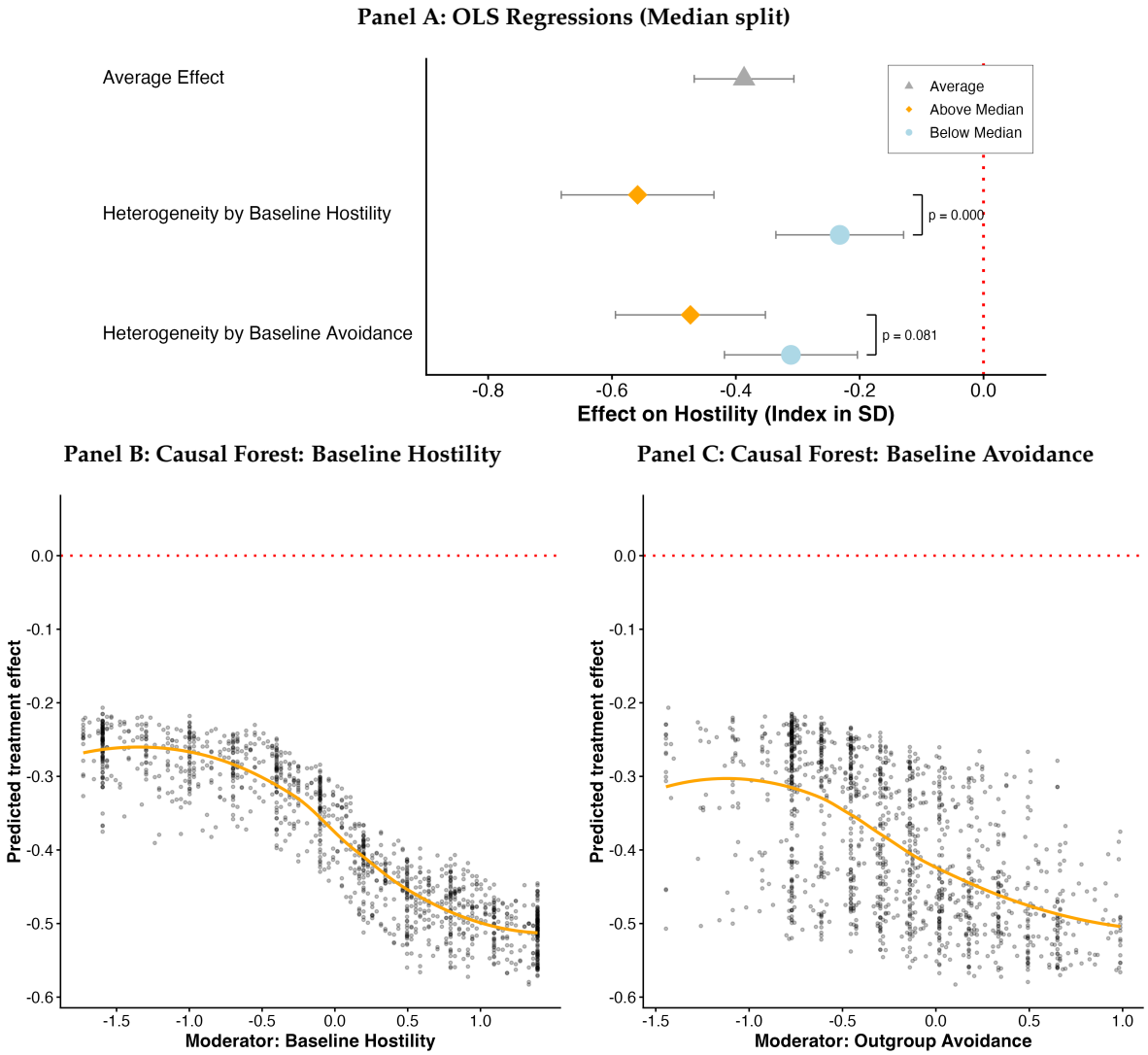
FIGURE 7: Treatment Effects of Conversations on Hostility in Short-Term Endline



Notes: This figure plots the treatment effect of the Conversation Treatment on discrimination in the cash transfer decision (Panel A) and feeling ratings (Panel B). The cash-transfer decision was administered only for about 40% of the sample in the short-term endline.

participants with a greater degree of avoidance of political discussions with the outgroup experience larger reductions in hostility. This can be seen in the bottom of Panel A of Figure 8, which plots the effect of conversations on hostility by baseline levels of outgroup avoidance. We confirm these results in panels B and C which plot predicted treatment effect from causal forest models.

FIGURE 8: Heterogeneous Treatment Effects by Baseline Hostility and Avoidance



Notes: This figure plots treatment effects of the Conversation Treatment by baseline levels of hostility and outgroup avoidance. The outcome measure is the standardized index of hostility in the short-term endline survey. The treatment effects in Panel A are derived from regressions in the subsamples indicated. The p-values are from a regression pooling the subsamples and interacting τ_t^C with a dummy for the respective subsample. Panels B and C plot predicted treatment effects, estimated based on generalized random forest models with the same set of covariates as in panel A.

These findings imply that facilitating cross-partisan political conversations may be a promising policy tool, because the greatest effects on hostility are concentrated among individuals who are most polarized at baseline. However, since these participants are also more likely to avoid political conversations with cross-partisans, they may be less likely to engage voluntarily.

What drives the differential effects for partisans with high degrees of hostility and outgroup avoidance? One potential explanation is that high-hostility and high-avoidance partisans had more positive experiences with their partner than other participants. Panel A of Figure A7 pro-

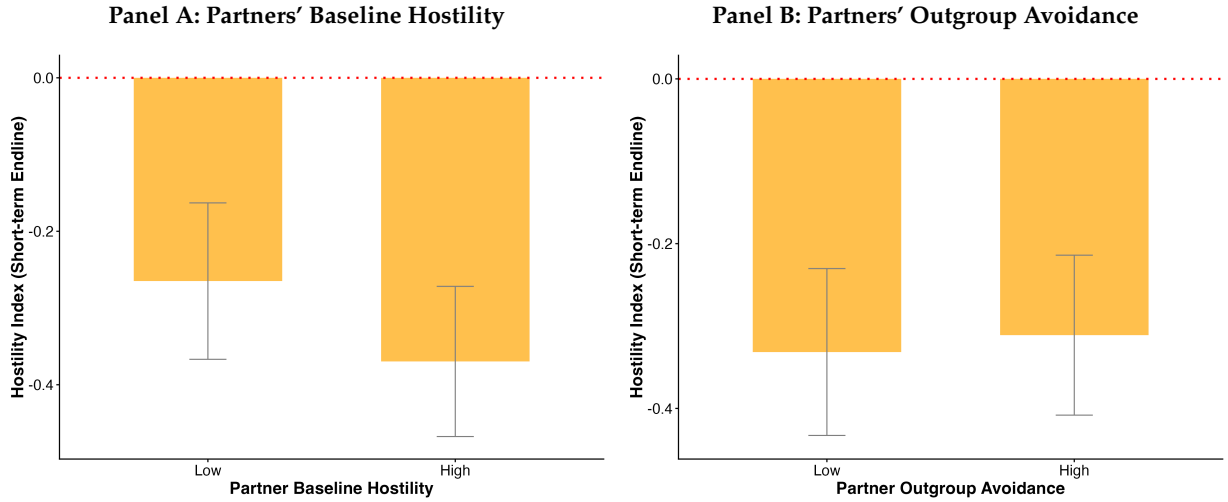
vide suggestive evidence against this hypothesis: if anything, participants with higher levels of hostility and avoidance of interactions report *worse* conversation outcomes. In line with this result, Figure A8 shows that high-hostility and high-avoidance partisans report *less* positive feelings for their conversation partner.

We find evidence for an alternative explanation: conversations lead to stronger belief updating among hostile and avoidant participants. Panel B of Figure A7 shows that high-hostility and high-avoidance partisans were more positively surprised by the conversations' outcomes. Although high-hostility and high-avoidance participants report worse conversation outcomes, the gap between their realized experiences and their ex-ante predictions is substantially larger than that of low-hostility and low-avoidance partisans. In line with this result, we find that hostile and avoidant partisans update more strongly about the outcomes of future interactions, as shown in Figure 13 and discussed below.

4.4.1 Implications for Conversation Partners

Next, we assess the impact of participants' baseline hostility and avoidance on the outcomes of their partners. To do so, we leverage that the assignment to partners is random, conditional on the partner's political group (left versus right). We find that cross-partisan political conversations with partners who are more affectively and ideologically polarized lead to significantly larger treatment effects. Moreover, treatment effects are similarly large for participants matched to partners with low or high degrees of outgroup avoidance.

FIGURE 9: Heterogeneous Treatment Effects by Partner Characteristics



Notes: This figure plots treatment effects of the Conversation Treatment by baseline hostility and outgroup avoidance of the randomly assigned conversation partner. The outcome variable is the hostility index measured in the short-term endline survey.

Figure 9 plots the effect of conversations on hostility by the conversation partner's baseline

hostility (Panel A) and baseline avoidance of cross-partisan political conversation (Panel B). As these figures show, the partner's baseline hostility somewhat moderates the treatment effect, with larger treatment effects for participants matched to partners with higher degrees of baseline hostility.[25] By contrast, conversations with individuals who are more avoidant of outgroup interactions lead to similar reductions in hostility.

These findings highlight that involving high-hostility and high-avoidance individuals in interactions leads to positive impacts both on these individuals and their partners. This suggests that current programs providing opportunities for cross-partisan interactions relying on voluntary enrollment may fail to unlock their full potential.

4.4.2 Persistence

The treatment effect of cross-partisan political conversations on hostility persist over a period of three months. Panel A of Figure 10 shows that treatment effects on hostility remain economically and statistically significant at the time of the long-term endline survey, which respondents completed three months after the intervention. The treatment effect on the feeling thermometer strongly decays relative to the endline survey administered immediately after the intervention and then remains stable. The treatment effect on the incentivized cash-transfer decision is more stable over time. Panels B and C plot heterogeneous treatment effects based on generalized random forest models over time. The heterogeneity in treatment effects by baseline hostility and avoidance persists three months after the intervention.

4.5 Impact on Outgroup Avoidance and Beliefs

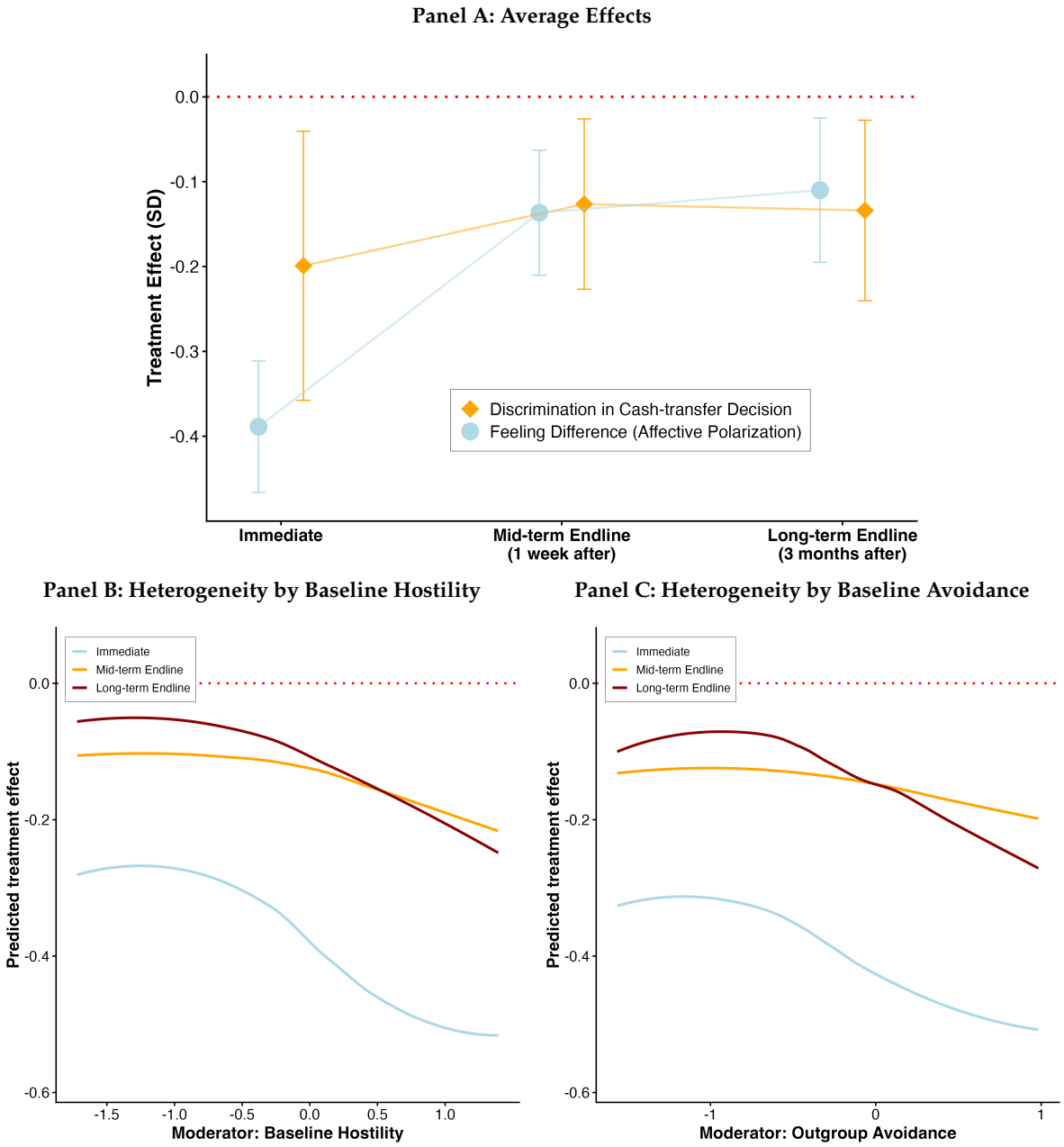
We find that cross-partisan political conversations reduce participants' avoidance of future participation in such conversations and improve their beliefs about their costs and benefits, particularly among participants exhibiting greater avoidance at baseline.

4.5.1 Impact on Outgroup Avoidance

The overwhelming majority of participants is interested in another interaction with the outgroup member they were randomly assigned to. Figure 11 plots the share of participants expressing interest in re-meeting their partner after the conversation: 81% of participants assigned to the Conversation Treatment express at least some interest, while 35% express strong interest in re-meeting their partner. While participants with high levels of hostility and avoidance expressed low initial demand for the outgroup interaction, their stated interest in a follow-up conversation with their assigned partner is as high as among participants with low levels of baseline hostility

[25] We find a similar heterogeneity by ideological extremity, with larger treatment effects for participants matched to partners with higher baseline ideological extremity.

FIGURE 10: Persistence of Treatment Effects on Hostility

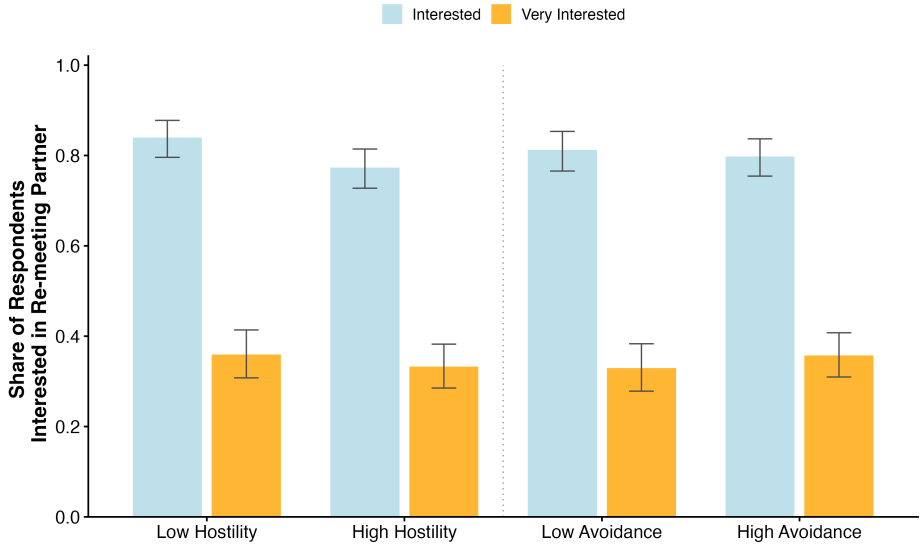


Notes: This figure plots treatment effects of the Conversation Treatment on a standardized index of hostility. Panel A compares treatment effects in the short-run, midterm, and long-term endline survey, separately for the two hostility measures. Panels B and C plot predicted treatment effects, estimated based on generalized random forest models with the same set of covariates as in panel A, separately for the short-run, mid-term, and long-term endline survey.

and avoidance.[26]

[26]We elicited interest on a 3-point scale with the answer options *Not interested*, *Somewhat interested*, *Very interested*. We elicited interest separately for an additional interaction to get to know the assigned partner on a personal level and for an additional interaction to discuss politics. The results across these two types of interactions are very similar. 81%

FIGURE 11: Interest in Re-meeting Outgroup Partner from Experiment



Notes: This figure plots the share of participants indicating interest in re-meeting the assigned conversation partner among participants in the Conversation Treatment. Blue bars plot the share of participants who indicated to be somewhat or very interested. Orange bars plot the share of participants who indicated to be very interested.

Figure 12 plots the impact of the conversation treatment on demand for future interactions with the outgroup. We find that the interactions reduce outgroup avoidance in the experiment (Panel A) and increase interest in outgroup interactions outside of the experiment (Panel B), with effects concentrated among participants with high levels of baseline avoidance and hostility.

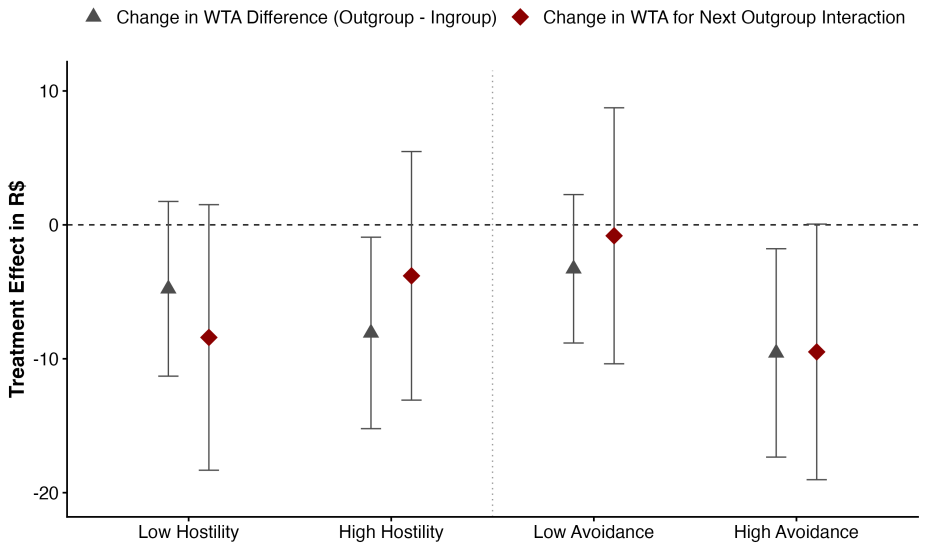
Panel A of Figure 12 plots the impact of the conversation treatment on incentivized measures of changes in demand for outgroup interactions in the experiment. To measure changes in demand, we combine the WTA elicited in the mid-term endline survey with WTA measures elicited in the baseline survey. Importantly, we need to take into account that when re-eliciting the WTA in the endline survey among treated participants, we elicit the WTA for a second conversation with the outgroup. To address this issue, we elicited in the baseline survey both participants' WTA to take part in a first cross-partisan political conversation, and participants' WTA to participate in a second cross-partisan political conversation in the event they are selected to complete the first.

To account for the fact that participants ask for larger payments for repeating the same activity for the second time in the baseline survey, we report treatment effects on changes in incentivized demand for cross-partisan political conversations and incentivized outgroup avoidance. We compute these measures by subtracting the baseline WTA for the second conversation from the endline WTA for the treatment group, while in the control group, we subtract the baseline WTA for the first conversation from the endline WTA, following

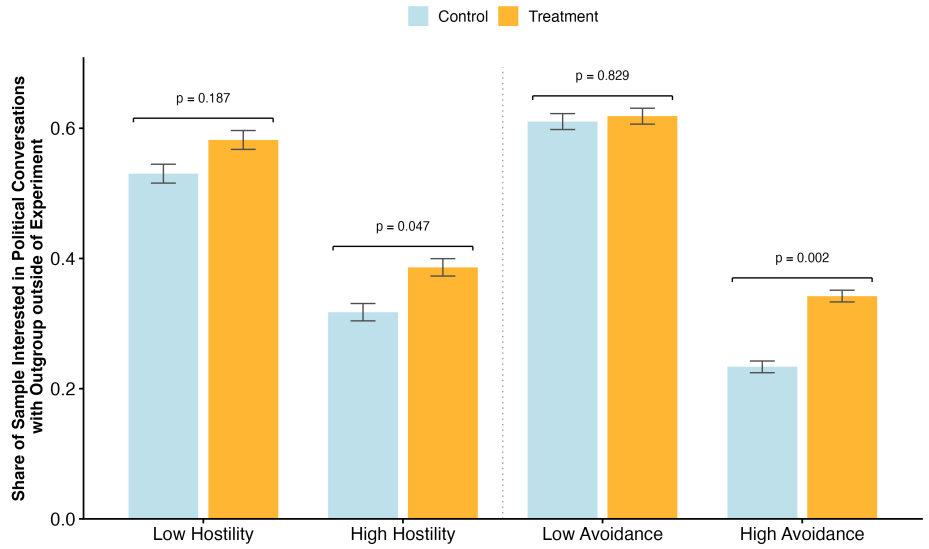
express interest in re-meeting their partner to get to know them on a personal level (shown in Figure 11), compared to 83% expressing interest in an additional conversation about politics.

FIGURE 12: Treatment Effects on Demand for Interactions with Outgroup

Panel A: Treatment Effect on Willingness-to-accept for Interactions in Experiment



Panel B: Treatment Effect on Interest in Outgroup Interactions Outside of Experiment



Notes: This figure plots treatment effects of the Conversation Treatment on demand for future interactions with the outgroup. Panel A plots treatment effects on changes in participants' incentivized willingness-to-accept (WTA) for a future interaction in the experiment (other than the first partner) relative to baseline values. Triangles plot the change in the differential WTA, subtracting the WTA for a political conversation with the ingroup from the WTA for a political conversation with the outgroup. Red diamonds plot the change in the WTA for a political conversation with the outgroup. Panel B plots the share of respondents expressing interest in political conversations with the outgroup outside of the experiment, separately for the Control Group (Blue) and the Conversation Treatment (Orange). Both panels split the results by baseline levels of hostility (left two panels) and baseline levels of outgroup avoidance (right two panels).

$$\Delta WTA_{PO} = \begin{cases} WTA_{\text{PO Endline}} - WTA_{\text{PO1 Baseline}} & \text{if } C_t = 0 \\ WTA_{\text{PO Endline}} - WTA_{\text{PO2 Baseline}} & \text{if } C_t = 1 \end{cases} \tag{3}$$

$$\Delta WTA_{Averse} = \begin{cases} (WTA_{\text{PO Endline}} - WTA_{\text{PO1 Baseline}}) - (WTA_{\text{PI Endline}} - WTA_{\text{PI Baseline}}) & \text{if } C_t = 0 \\ (WTA_{\text{PO Endline}} - WTA_{\text{PO2 Baseline}}) - (WTA_{\text{PI Endline}} - WTA_{\text{PI Baseline}}) & \text{if } C_t = 1 \end{cases} \tag{4}$$

where $WTA_{\text{PO1 Baseline}}$ refers to the WTA for the first and $WTA_{\text{PO2 Baseline}}$ to the WTA for the second cross-partisan political conversation, both elicited in the baseline, $WTA_{\text{PO Endline}}$ refers to the WTA for a cross-partisan political conversation elicited in the mid-term endline, and $WTA_{\text{PI Baseline}}$ and $WTA_{\text{PI Endline}}$ refer to the WTA for a political conversation with the ingroup, elicited in the baseline and mid-term endline.

Panel B of Figure 12 shows that interactions with the outgroup in the experiment also lead to increased interest in interactions outside of the experiment. Treated participants with above-median baseline avoidance and baseline hostility report significantly greater interest in cross-partisan political conversations in day-to-day interactions, compared to participants assigned to the control group. This effect is largest among participants with high degrees of baseline avoidance of interactions: the treatment raises interest by over 40%.[27]

4.5.2 Impact on Beliefs

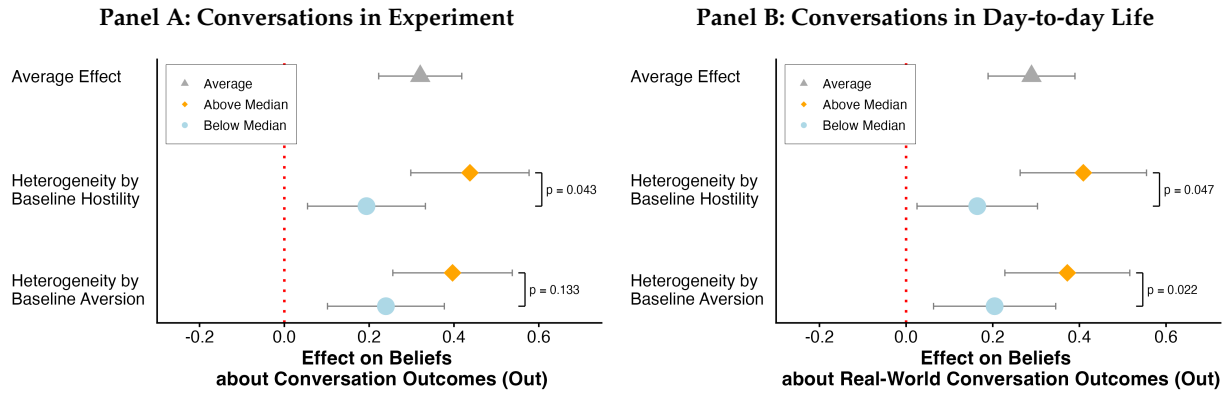
As reported in Figure 6, treated participants report substantially more positive conversation outcomes than they anticipated, suggesting that their pessimistic expectations may have been mistaken. In line with this result, we find that treated participants have more optimistic beliefs about the costs and benefits of future cross-partisan political conversations.

Panel A of Figure 13 plots treatment effects on a standardized index of beliefs about conversation outcomes in the experiment by baseline avoidance and hostility. As shown in this figure, the average treatment effect is large and significant: treated participants report more optimistic beliefs that correspond to an average effect of 0.3 standard deviations relative to the control group. The effect is significantly larger among participants with above-median baseline avoidance and hostility (0.4 standard deviations).

In Panel B of Figure 13, we demonstrate that the conversation also impacted beliefs about conversation outside of the experiment. Similar to the incentivized beliefs about the experiment, the effects are larger among participants with higher avoidance and hostility.

[27] Panel B of Figure A9 displays the treatment effects on the difference between self-reported interest in co-partisan vs. cross-partisan political conversations. Qualitatively, participants experience reductions in outgroup avoidance, though these results are not statistically significant. This result is explained by the fact that conversations also induce an increase in interest in political conversations with the ingroup.

FIGURE 13: Treatment Effects on Beliefs about Outcomes of Cross-partisan Conversations



Notes: This figure plots treatment effects of the Conversation Treatment on a standardized index of beliefs about the outcomes of future interactions with the outgroup. The index averages over the beliefs shown in Figure 6, re-elicited in the mid-term endline survey. Panel A plots effects on incentivized beliefs about future interactions in the experiment. Panel B plots effects on beliefs about future interactions outside of the experiment.

4.6 Impact on Political Outcomes

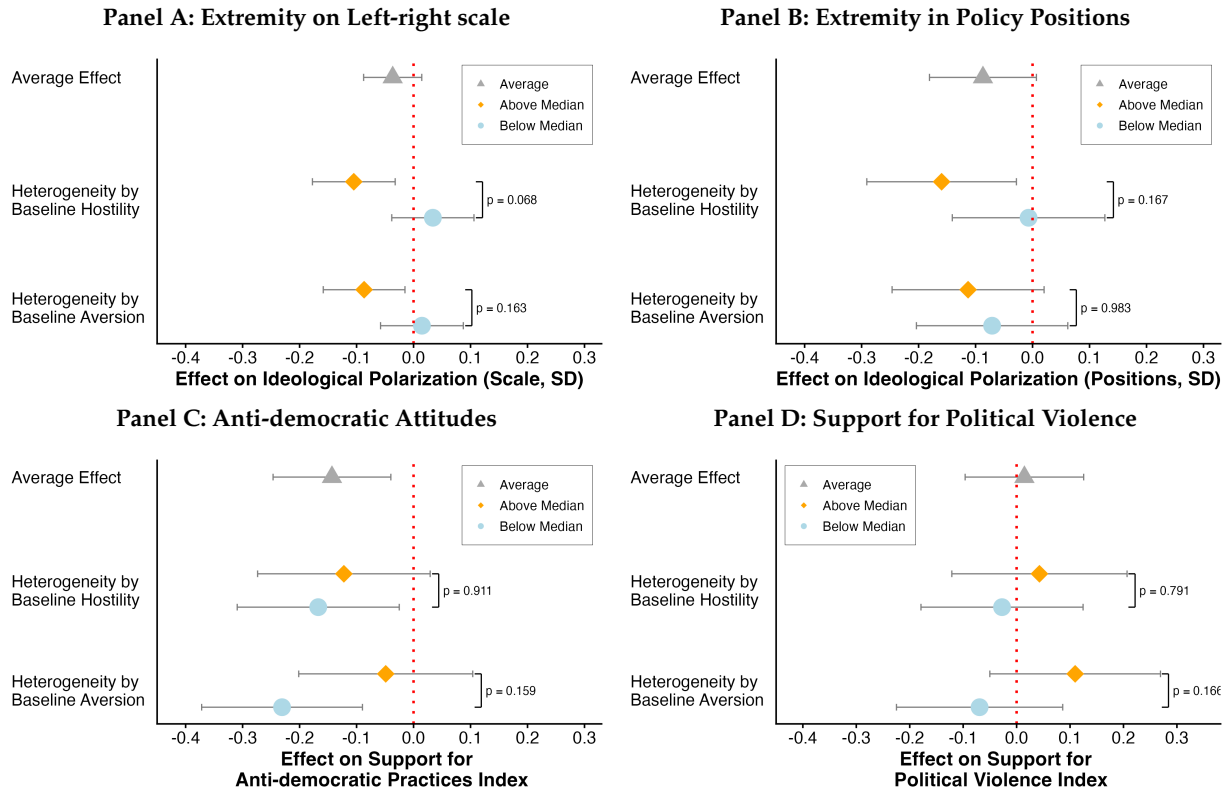
We find that conversations lead to reductions in ideological extremity among participants with high levels of baseline hostility and avoidance. Panel A of Figure 14 plots the treatment effect of the conversations on participants' self-reported extremity on the left-right scale. We find reductions of 0.1 SD among participants with high degrees of hostility and outgroup avoidance at baseline. In Panel B, we plot the treatment effect on extremity in policy positions based on a composite measure of extremity across the 10 policy areas we survey (Allcott et al. 2020). The effects on this measure confirm the finding that the conversations led to reductions in ideological polarization, concentrated among high-hostility and high-avoidance individuals.

We find that conversations causally reduce support for anti-democratic practices but do not shift support for political violence. Panel C of Figure 14 plots the treatment effect of the conversations on an index of measures of anti-democratic practices, following (Voelkel et al. 2024). We find statistically significant treatment effects for the full sample, with no significant heterogeneity between low and high hostility individuals. Treatment effects are more pronounced among participants with lower degrees of baseline avoidance. The treatment effects on an index of support for political violence are close to zero and not statistically significant.

4.7 Impact on Labor Market Outcomes

We find that conversations with outgroup members marginally reduce preferences for labor market discrimination against political outgroups. Figure 15 plots average and heterogeneous treatment effects for four survey measures of labor market discrimination. Panel A plots the effect of outgroup conversations on self-reported preferences to leave current employment upon discover-

FIGURE 14: Treatment Effects on Political Extremity

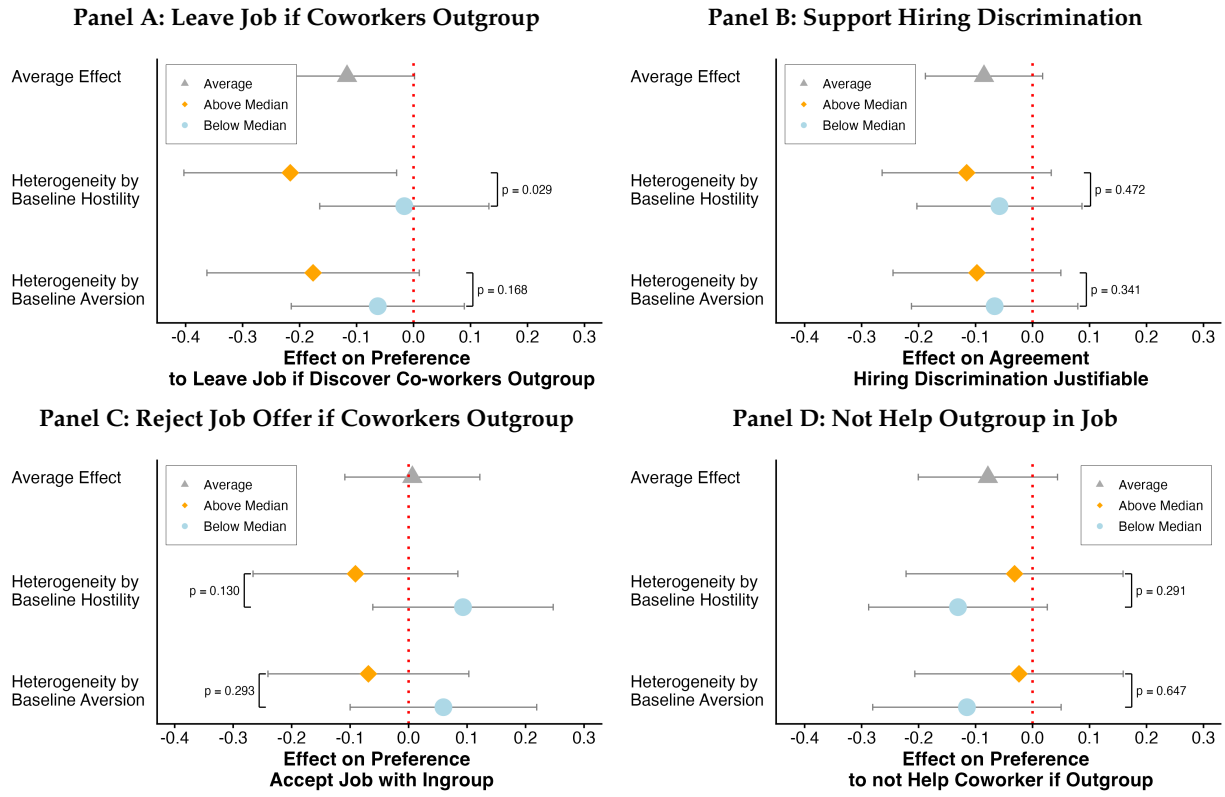


ing that co-workers are from the political outgroup.

We find an average treatment effect of 0.1 SD, with significantly larger effects among participants with higher degrees of baseline hostility (-0.2 SD). We find comparable effects on support for a statement asking whether hiring discrimination based on political affiliation is justifiable (shown in Panel B), and a proxy for collaboration with outgroup workers (Panel D).[28] We do not find effects on self-reported agreement to reject a job offer if the co-workers in the new position were pre-dominantly outgroup members.

[28]Participants were asked whether they would refrain from helping a co-worker if they were from the outgroup.

FIGURE 15: Treatment Effects on Political Discrimination in Labor Market Decisions



5 Implications for Policy Design

Our findings highlight that enforcing desegregation through contact interventions is effective at reducing hostility, in particular among high-hostility individuals who currently avoid interactions. In this section, we present a cost-benefit analysis, demonstrating that contact interventions relying on voluntary self-enrollment, like those from our partner organizations, may fall short of their full potential as they fail to enroll the groups most likely to benefit from the interaction. In the second part of this section, we benchmark the contact intervention against a lower-cost alternative policy.

5.1 Cost-Benefit Analysis

In the cost-benefit analysis, we focus on two key aspects in the design of current policy interventions: should interventions rely on volunteer enrollment or offer financial incentives for participation? How should interventions be targeted in recruitment?

5.1.1 Framework

We consider an intervention design problem, in which an organizer seeks to minimize average hostility in a population of N individuals. Individual i needs to be paid at least WTA_i to participate in the intervention and has a potential treatment effect of τ_i. The organizer has access to observables that partition the population into $J \geq 1$ groups of size N_j. The organizer knows the joint distribution of WTA and τ in each group.

The designer chooses a *recruitment scheme* $\mathcal{R}$ and an *participation incentive scheme* $\mathcal{P}$. The recruitment scheme $\mathcal{R}$ specifies for each group j a number N_j^T of individuals to be targeted via a recruitment ad. The cost of each ad is denoted by c. The incentive scheme $\mathcal{P}$ specifies for each $j \in \mathcal{J}$ a monetary incentive $P_j \geq 0$ offered in exchange for participating.

Given this set-up, the total cost and benefit of intervention $(\mathcal{R}, \mathcal{P})$, can be expressed as follows:

$$C_{\mathcal{R},\mathcal{P}} = \sum_j N_j^T (c + \pi(WTA_i \leq P_j | i \in \mathcal{I}^j) P_j)$$

$$B_{\mathcal{R},\mathcal{P}} = \sum_j \frac{N_j^T}{N} \pi(WTA_i \leq P_j | i \in \mathcal{I}^j) \tau(WTA_i \leq P_j | i \in \mathcal{I}^j)$$

, where $\pi(WTA_i \leq P_j | i \in \mathcal{I}^j)$ denotes the share of group j whose WTA is lower than the incentive payment P_j and $\tau(WTA_i \leq P_j | i \in \mathcal{I}^j)$ denotes their average treatment effect. The designer seeks to maximize the total benefit of the intervention, subject to a budget constraint on the total cost.

5.1.2 Estimation

Our experiment allows us to estimate the costs and benefits of a broad range of recruitment targeting and participation incentive schemes. In particular, we can use the joint distribution of WTA_i and τ_i, along with other measures we collect, to estimate $\tau(WTA_i < P_j | i \in \mathcal{I}^j)$ and $\pi(WTA_i < P_j | i \in \mathcal{I}^j)$ up to $P_j = R\$104$, the participation offer in the experiment. We use the data from our recruitment on social media to estimate c. Based on our recruitment costs for the experiment, we assume $c = R\$18$.[29] We assume the total population size to be 94 million, the estimated number of adult Meta users in Brazil that can be reached through advertisements. In the simulations, we use estimates of τ_i from the midterm endline survey, capturing effects of the intervention within four weeks.

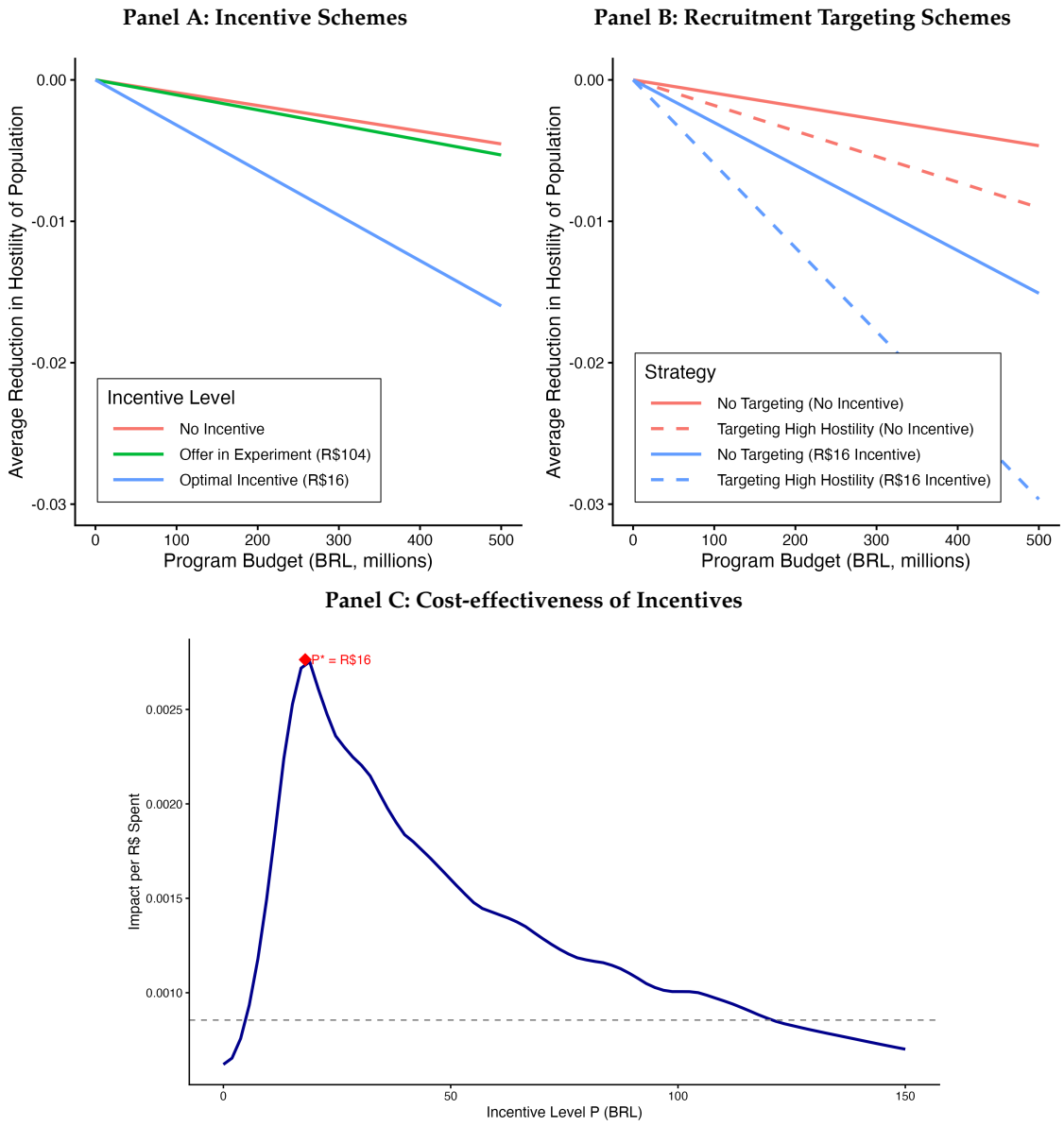
5.1.3 Results

In this section, we present the results of a simulation exercise to identify the optimal incentive scheme. Panel A of Figure 16 plots the average reduction in the population's hostility achieved for a given level of budget, under three flat fee incentive schemes. First, we plot a volunteer scheme

[29] This is the advertising cost for a participant who opened our screening survey, consented to the study, and completed the screening survey.

in red, offering no monetary incentives, as currently implemented by our partner organizations. Second, we plot a flat incentive of R$104 in green, which equals the incentive we offered in our experiment. Finally, we plot the optimal flat fee incentive scheme in blue, maximizing impact per dollar spent. Panel B of Figure 16 plots the average reduction in hostility as a function of the flat incentive payment, assuming budget levels of 500 million BRL or lower.

FIGURE 16: Simulations of Incentive and Recruitment Targeting schemes



Notes: This figure plots results of simulations of various incentive schemes (Panel A and C) and recruitment targeting schemes (Panel B). Estimates are based on treatment effects and WTA values derived from the experiment. The average reduction in hostility of the population is measured in standard deviations.

The figures demonstrate that moderate incentives outperform the volunteer scheme at any

budget level. This is the case for two reasons. First, volunteers have a lower average treatment effect compared to participants with higher levels of WTA. More importantly, only a small fraction of the population is willing to participate in the intervention without a financial incentive. As a result, a substantial share of the project budget is spent on advertising to individuals who choose not to participate voluntarily. On the other hand, if the offered monetary incentive is too large, the direct cost of paying participants outweighs the benefit of attracting more participants with greater treatment effects. As shown in Panel C of Figure 16, given our estimates, volunteer schemes outperform incentivized interventions offering flat payments greater than R$120.

Next, we turn to the results of a simulation exercise that compares various targeting schemes. Panel B of Figure 16 compares the effectiveness of two targeting schemes: first, without any targeting, and second, targeting individuals with high levels of hostility, a group we have found to have significantly larger treatment effects relative to low-hostility partisans. We compare both targeting schemes for two incentive schemes: the volunteer incentive scheme (R$0) and an incentive of R$16.

This exercise demonstrates that, regardless of the participation incentive, the organizer can substantially improve the effectiveness of the intervention by targeting recruitment based on baseline hostility. As shown in our experiment, high-hostility participants have substantially higher treatment effects. Although high-hostility individuals require higher participation payments, on average, reduced participation among this population is offset by their greater treatment effects.

5.2 Alternative Policy

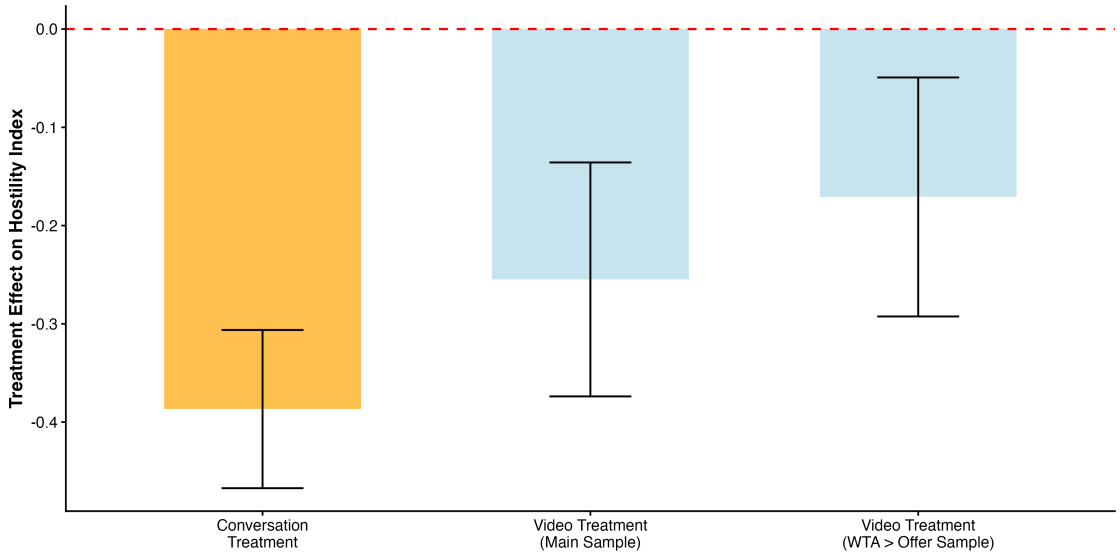
Our findings highlight that participants with high potential treatment effects are costly to recruit, making it difficult to roll out contact interventions in this population at scale.

We thus designed an alternative, cost-effective policy intervention with the non-profit organization *More in Common*. We developed a video campaign that distills several key components of the effective but costly contact intervention. It provides testimonials of participants before and after the conversation, alongside information on the actual experience of participants in the experiment, as well as actual overlap in political and personal characteristics of partisans in Brazil. We designed the intervention based on the model of similar interventions studied in recent work on polarization in the US (Voelkel et al. 2024).

The treatment effect of the video intervention on outgroup avoidance is, thus, indicative of the causal role of misspecified beliefs in individuals' decisions to avoid cross-partisan political conversations across the full sample. On the other hand, the treatment allows us to speak to the direct effect of a cost-efficient, scalable information intervention on outgroup hostility and avoidance. We administered this alternative policy as a separate randomized treatment arm in our main research sample. In addition, we randomized the intervention among the sample of about 20% of participants with WTA values exceeding our offer for the conversation.

The video intervention leads to significant reductions in hostility, even among participants

FIGURE 17: Treatment Effects of Video Treatment on Hostility in Short-term Endline



with WTA values exceeding the offer in the experiment. As shown in Figure 17, the video intervention leads to a reduction in hostility of 0.25 standard deviations, which is significant at the 1% level. The effect among the sample of participants with WTA values exceeding the financial incentive offered for completing the conversation treatment is significant yet smaller than for the main sample.

We find that the video leads to comparable average treatment effects on both beliefs and interest in future conversations. Figure A10 compares the impact of the video and the conversation on unincentivized measures of beliefs about conversation outcomes (Panel A) and interest in cross-partisan conversations in day-to-day interactions (Panel B).

These findings suggest that, at least in the short run, the cost-effective video intervention may be a viable alternative to direct exposure through the conversation treatment. In a future version of the paper, we will present results from the long-term endline surveys to track decay of the effects separately for the video and the conversation intervention.

6 Conclusion

This paper studies the determinants and effects of cross-partisan interactions on political polarization through a field experiment with over 5,000 Brazilians. We document a strong association between political segregation and political polarization. Those most polarized are least likely to encounter the outgroup in their daily lives and avoid political discussions with the other side. As a result, the participants most in need of a policy intervention are also least interested in participating in interventions that facilitate cross-partisan interactions.

In our experiment, we can provide the first evidence on the effects of conversations on highly

polarized individuals who require large financial incentives to engage in cross-partisan conversations. We find that exposing partisans to political conversations with the outgroup significantly reduces partisan hostility. These reductions are of a similar size to the increases of partisan hostility seen in Brazil and the US in the last 10 years. The effects of the conversations outweigh the effects of an alternative, lower-cost informational video.

Importantly, we uncover a strong degree of heterogeneity in the impact of the intervention. Treatment effects are strongest among the most polarized individuals who are also most avoidant of cross-partisan contact. Moreover, the conversations are more effective when matching participants with more ideologically distant partners.

This raises a fundamental challenge for existing programs that rely on voluntary participation: those who benefit most from the intervention and make it effective are least likely to sign up for it. Future research should address the question of how those highly polarized individuals with low levels of demand can be recruited for interventions. Moreover, it's important to understand whether similar findings hold for contact interventions in different settings that require voluntary enrollment.

References

Ahler, D. J. and Sood, G. (2018). The Parties in Our Heads: Misperceptions about Party Composition and Their Consequences. *The Journal of Politics*, 80(3):964–981.

Allcott, H., Braghieri, L., Eichmeyer, S., and Gentzkow, M. (2020). The Welfare Effects of Social Media. *American Economic Review*, 110(3):629–676.

Allport, G. (1954). *The Nature of Prejudice*. Garden City, NJ Anchor.

Autor, D., Dorn, D., Hanson, G., and Majlesi, K. (2020). Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure. *American Economic Review*, 110(10):3139–3183.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., and Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221.

Bailey, M., Johnston, D. M., Koenen, M., Kuchler, T., Russel, D., and Stroebel, J. (2022). The Social Integration of International Migrants: Evidence from the Networks of Syrians in Germany. *NBER Working Paper No 29925*.

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago press.

Bertrand, M. and Duflo, E. (2017). Field Experiments on Discrimination. In Banerjee, A. V. and Duflo, E., editors, *Handbook of Economic Field Experiments*, volume 1 of *Handbook of Field Experiments*, pages 309–393. North-Holland.

Blattner, A. and Koenen, M. (2023). Does Contact Reduce Affective Polarization? Field Evidence from Germany. *Working Paper*.

Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., and Eccles, J. (2006). Empathy or Antipathy? The Impact of Diversity. *American Economic Review*, 96(5):1890–1905.

Boxell, L., Gentzkow, M., and Shapiro, J. M. (2017). Greater Internet Use Is not Associated with Faster Growth in Political Polarization among US Demographic Groups. *Proceedings of the National Academy of Sciences*, 114(40):10612–10617.

Braghieri, L., Schwardmann, P., and Tripodi, E. (2025). Talking across the Aisle. *Working Paper*.

Broockman, D. and Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282).

Brown, J. R., Cantoni, E., Enos, R., Pons, V., and Sartre, E. (2025). Sources and Extent of Rising Partisan Segregation in the U.S. - Evidence from 143 Million Voters. *NBER Working Paper*, No 33422.

Bursztyn, L., Chaney, T., Hassan, T., and Rao, A. (2021). The immigrant next door: Exposure, prejudice, and altruism. *NBER Working Paper No. 28488*.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Chetty, R., Jackson, M. O., Kuchler, T., and Stroebel, J. (2022). Social Capital II: Determinants of Social Connectedness. *NBER Working Paper No. 30314*.

Currarini, S., Jackson, M. O., and Pin, P. (2009). An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica*, 77(4):1003–1045.

Cutler, D. M. and Glaeser, E. L. (1997). Are Ghettos Good or Bad? *The Quarterly Journal of Economics*, 112(3):827–872.

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

Field, E., Levinson, M., Pande, R., and Visaria, S. (2008). Segregation, Rent Control, and Riots: The Economics of Religious Conflict in an Indian City. *American Economic Review*, 98(2):505–510.

Finseraas, H., Hanson, T., Johnsen, Å. A., Kotsadam, A., and Torsvik, G. (2019). Trust, ethnic diversity, and personal contact: A field experiment. *Journal of Public Economics*, 173:72–84.

Heuser, S. and Stötzer, L. S. (2022). The Effects of Face-To-Face Conversations on Polarization: Evidence from a Quasi-Experiment. *Working Paper*.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science*, 22:129–46.

Jha, S. (2013). Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia. *American Political Science Review*, 107(4):806–832.

Kalla, J. L. and Broockman, D. E. (2022). Voter Outreach Campaigns Can Reduce Affective Polarization among Implementing Political Activists: Evidence from Inside Three Campaigns. *American Political Science Review*, 116(4):1516–1522.

Levendusky, M. and Stecula, D. (2021). *We Need to Talk - How Cross-Party Dialogue Reduced Affective Polarization*. Elements in Experimental Political Science. Cambridge University Press.

Levendusky, M. S. (2018). Americans, Not Partisans: Can Priming American National Identity Reduce Affective Polarization? *The Journal of Politics*, 80(1):59–70.

Levy, R. (2020). Social Media, News Consumption, and Polarization: Evidence from a Field Experiment. *Forthcoming, American Economic Review*.

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

Lowe, M. (2025). Has Intergroup Contact Delivered? *Annual Review of Economics*, 17:321–344.

Martin, G. J. and Yurukoglu, A. (2017). Bias in Cable News: Persuasion and Polarization. *American Economic Review*, 107(9):40.

Nunes, F. and Traumann, T. (2023). *Biografia do Abismo*. HarperCollins Brasil, 1st edition.

Paluck, E. L., Green, S. A., and Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, 3(02):129–158.

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

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

Schelling, T. C. (1969). Models of Segregation. *The American Economic Review*, 59(No 2):488–493.

Voelkel, J. G., Stagnaro, M. N., Chu, J. Y., Pink, S. L., Mernyk, J. S., Redekopp, C., Ghezae, I., Cashman, M., Adjodah, D., Allen, L. G., Allis, L. V., Baleria, G., Ballantyne, N., Van Bavel, J. J., Blunden, H., Braley, A., Bryan, C. J., Celniker, J. B., Cikara, M., Clapper, M. V., Clayton, K., Collins, H., DeFilippis, E., Dieffenbach, M., Doell, K. C., Dorison, C., Duong, M., Felsman, P., Fiorella, M., Francis, D., Franz, M., Gallardo, R. A., Gifford, S., Goya-Tocchetto, D., Gray, K., Green, J., Greene, J., Güngör, M., Hall, M., Hecht, C. A., Javeed, A., Jost, J. T., Kay, A. C., Kay, N. R., Keating, B., Kelly, J. M., Kirk, J. R. G., Kopell, M., Kteily, N., Kubin, E., Lees, J., Lenz, G., Levendusky, M., Littman, R., Luo, K., Lyles, A., Lyons, B., Marsh, W., Martherus, J., Maurer, L. A., Mehl, C., Minson, J., Moore, M., Moore-Berg, S. L., Pasek, M. H., Pentland, A., Puryear, C., Rahnama, H., Rathje, S., Rosato, J., Saar-Tsechansky, M., Almeida Santos, L., Seifert, C. M., Shariff, A., Simonsson, O., Spitz Siddiqi, S., Stone, D. F., Strand, P., Tomz, M., Yeager, D. S., Yoeli, E., Zaki, J., Druckman, J. N., Rand, D. G., and Willer, R. (2024). Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity. *Science*, 386(6719).

Appendix Figures and Tables

A1. Sample Characteristics

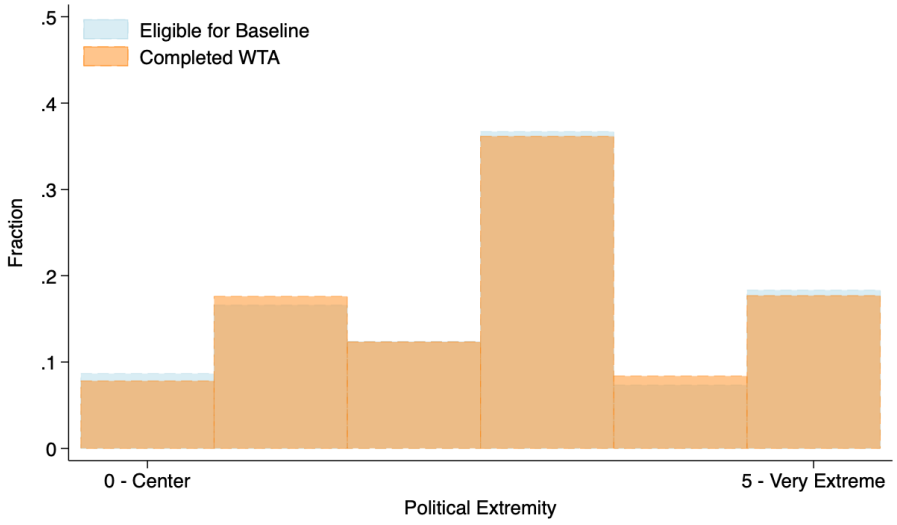
TABLE A1: Sample Statistics from Baseline to Endline

	Full Sample					Eligible for Conversation			
	Started Baseline	Finished Baseline	Scheduled & Randomized	Completed Endline 1	Completed Endline 2	Conversation Treatment	Control or Video Treat	Balance Test	Brazilian adult population
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Demographics									
Female	0.63	0.61	0.60	0.60	0.61	0.57	0.56	0.01	0.51
	(0.48)	(0.49)	(0.49)	(0.49)	(0.49)	(0.50)	(0.50)	(0.67)	
34 years or younger	0.35	0.40	0.38	0.38	0.38	0.37	0.37	0.00	0.35
	(0.48)	(0.49)	(0.49)	(0.48)	(0.49)	(0.48)	(0.48)	(1.00)	
White	0.56	0.57	0.57	0.58	0.58	0.57	0.55	0.02	0.43
	(0.50)	(0.49)	(0.49)	(0.49)	(0.49)	(0.50)	(0.50)	(0.39)	
Protestant / Evangelical	0.27	0.25	0.26	0.26	0.26	0.26	0.27	-0.01	0.27
	(0.44)	(0.43)	(0.44)	(0.44)	(0.44)	(0.44)	(0.44)	(0.62)	
High School Degree	0.43	0.40	0.36	0.35	0.35	0.35	0.37	-0.02	0.31
	(0.50)	(0.49)	(0.48)	(0.48)	(0.48)	(0.48)	(0.48)	(0.37)	
University Degree	0.52	0.58	0.62	0.63	0.63	0.63	0.61	0.02	0.25
	(0.50)	(0.49)	(0.49)	(0.48)	(0.48)	(0.48)	(0.49)	(0.37)	
Political Characteristics									
Voted Lula	0.43	0.48	0.47	0.48	0.47	0.45	0.47	-0.02	0.39
	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.39)	
Voted Bolsonaro	0.39	0.34	0.37	0.36	0.36	0.40	0.36	0.04	0.37
	(0.49)	(0.48)	(0.48)	(0.48)	(0.48)	(0.49)	(0.48)	(0.08)	
Left	0.46	0.52	0.51	0.51	0.51	0.47	0.51	-0.04	0.29
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.09)	
Center	0.09	0.08	0.08	0.07	0.07	0.08	0.09	-0.01	0.22
	(0.28)	(0.27)	(0.26)	(0.26)	(0.26)	(0.27)	(0.28)	(0.43)	
Right	0.45	0.40	0.42	0.41	0.42	0.45	0.41	0.04	0.49
	(0.50)	(0.49)	(0.49)	(0.49)	(0.49)	(0.50)	(0.49)	(0.08)	
Ideologically extreme	0.26	0.27	0.27	0.27	0.27	0.27	0.26	0.01	
	(0.44)	(0.44)	(0.45)	(0.44)	(0.45)	(0.45)	(0.44)	(0.63)	
Above Median Hostility	0.50	0.49	0.49	0.49	0.48	0.45	0.46	-0.01	
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.67)	
Above Median Aversion	0.50	0.49	0.49	0.49	0.48	0.41	0.36	0.05	
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)	(0.48)	(0.03)	
Attention / Comprehension									
No WTA	0.53	0.59	0.62	0.63	0.63	0.60	0.59	0.01	
Comprehension Errors	(0.50)	(0.49)	(0.49)	(0.48)	(0.48)	(0.49)	(0.49)	(0.66)	
Passed First Two	0.83	0.92	0.92	0.93	0.92	0.92	0.92	0.00	
Attention Checks	(0.37)	(0.27)	(0.26)	(0.26)	(0.26)	(0.27)	(0.27)	(1.00)	
Passed Technology	0.84	0.95	1.00	1.00	1.00	1.00	1.00	0.00	
Check	(0.36)	(0.22)	(0.03)	(0.04)	(0.04)	(0.05)	(0.03)	(1.00)	
N	13841	5283	2539	2403	2227	848	1019	1867	

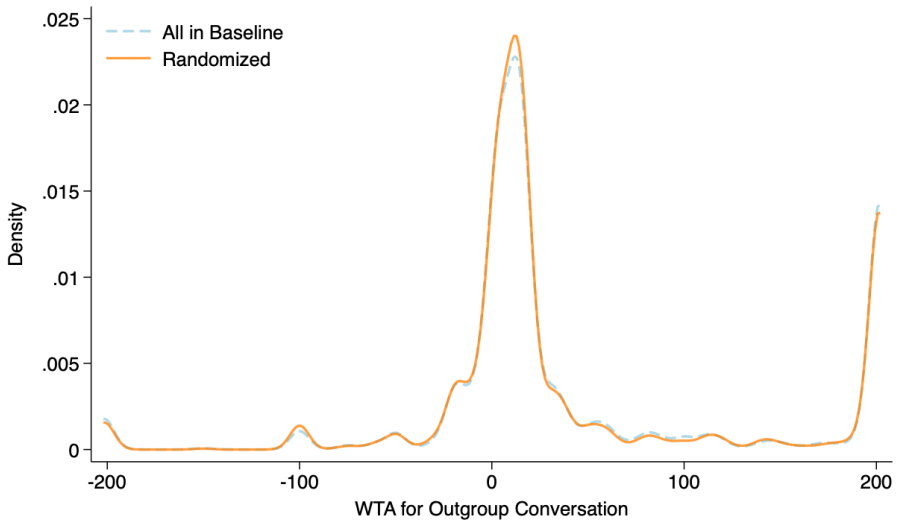
Notes: This table shows means, standard deviations (in parentheses), and sample sizes for our sample at various stages. Column 3 includes participants who successfully entered the platform on the scheduled date and time, and were subsequently randomized. Column 8 shows the difference in means in columns 6 and 7, and the p-value of a t-test between the two samples (in parentheses). Statistics for variables collected in the baseline survey are calculated based on the subsample of participants who reached the respective section of the survey.

FIGURE A1: Selective Attrition from Experiment

Panel A: Recruitment to WTA Elicitation



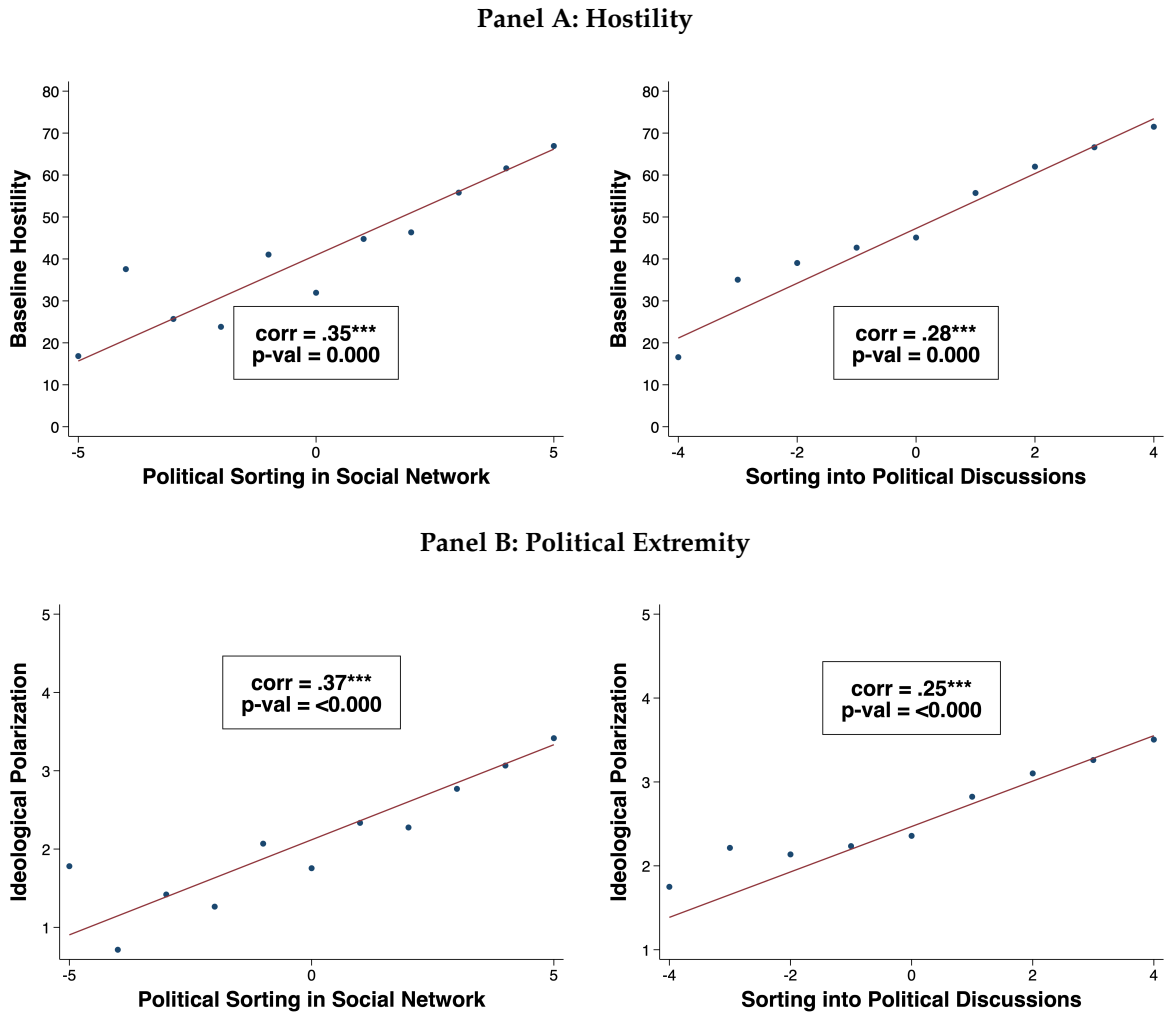
Panel B: Baseline to Randomization



Notes: This figure plots tests for selective attrition from the experiment. Panel A plots political extremity for participants who completed the screening and were eligible for the baseline and those who started the baseline and completed the willingness-to-accept (WTA) elicitation. Panel B plots the WTA for all participants who completed the elicitation in the baseline (blue line) and those who are randomized (orange line).

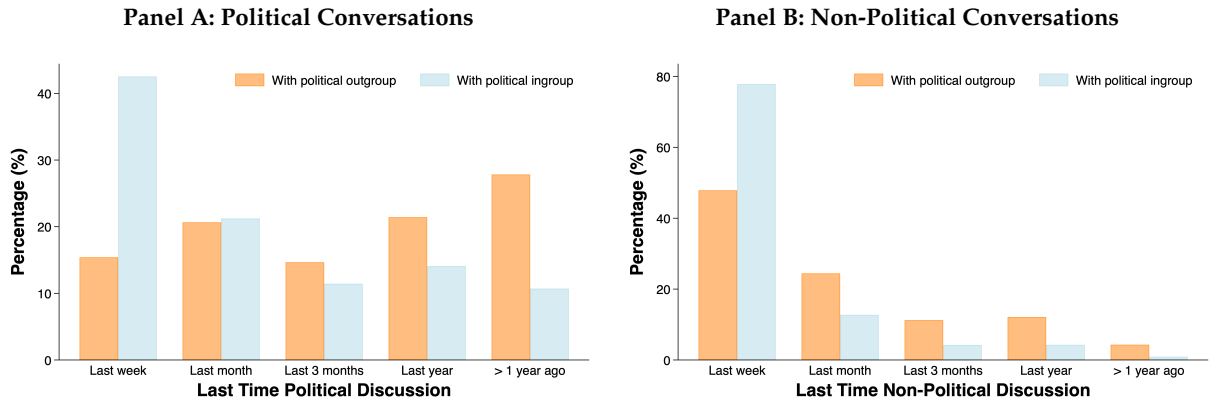
A2. Descriptive Statistics

FIGURE A2: Segregation, Hostility, and Political Extremity



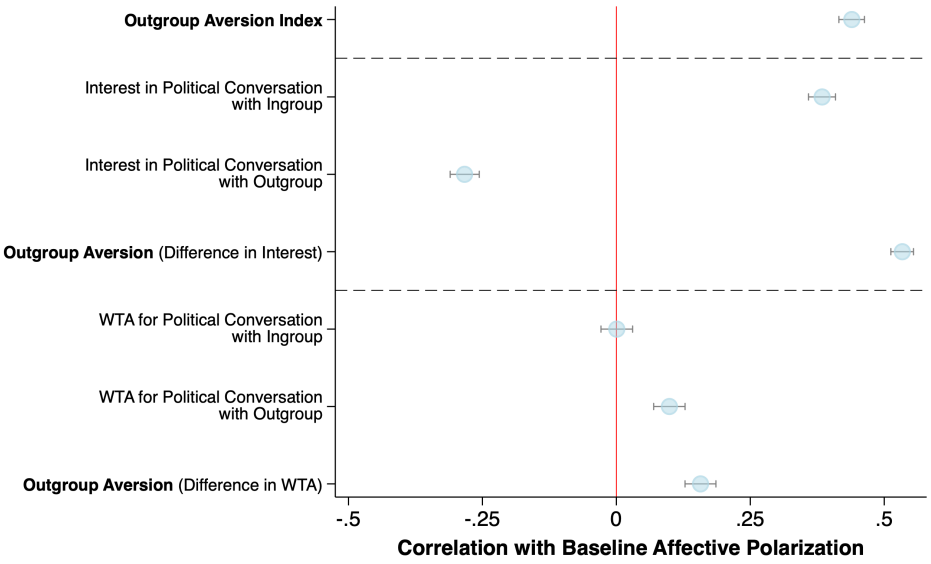
Notes: This figure shows binned scatter plots of outgroup hostility and measures of political segregation (Panel A) and of political extremity and measures of political segregation (Panel B). Political sorting in social networks is defined as the difference between the number of ingroup and outgroup friends. Sorting into political discussions is defined as the difference between the time of the last political conversation with an outgroup and an ingroup member. Hostility is defined as the difference in feelings between ingroup and outgroup members. Ideological extremity is defined as the self-reported distance from the center on a left-right scale.

FIGURE A3: Political Sorting into Conversations



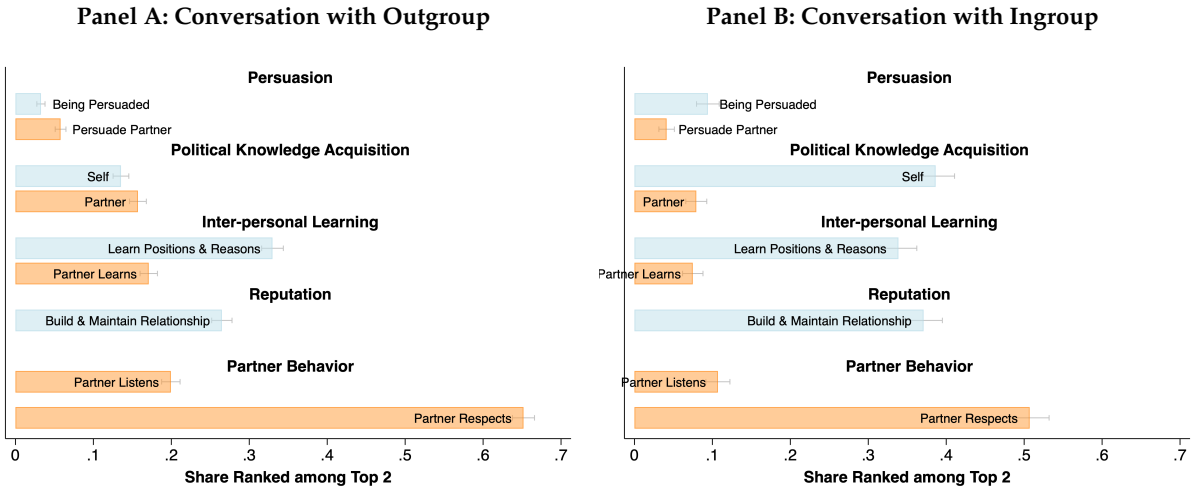
Notes: This figure plots the time of the last political conversation with an outgroup and an ingroup member (Panel A) and of non-political conversations (Panel B).

FIGURE A4: Outgroup Avoidance and Hostility



Notes: This figure plots the correlation between various measures of outgroup avoidance and baseline levels of outgroup hostility. The outgroup avoidance index is a standardized index, averaging over outgroup avoidance in day-to-day interactions and the differential WTA by political conversations with the outgroup and the ingroup. Outgroup hostility is measured as the difference in feelings between ingroup and outgroup members.

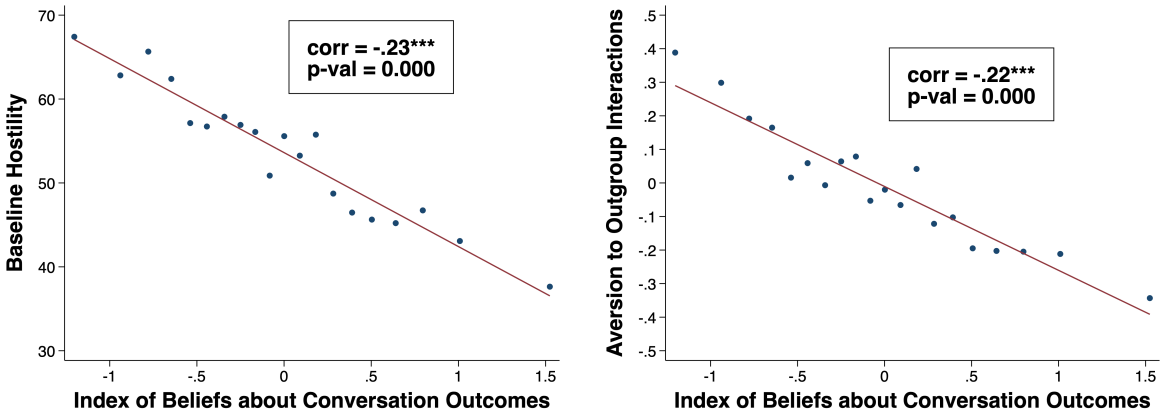
FIGURE A5: Preferences for Outcomes of Political Conversations



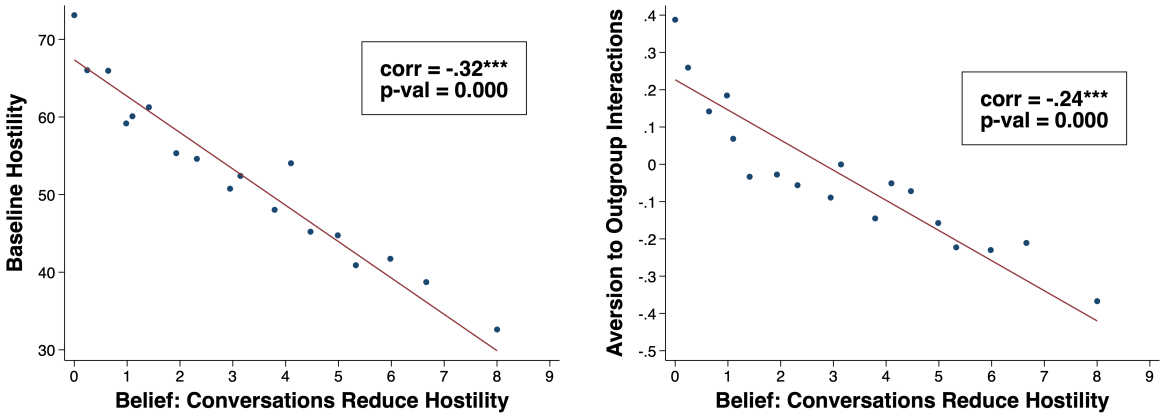
Notes: This figure plots participants' preferences over conversation outcomes. Participants were asked to rank nine conversation outcomes, which were identified in open-ended questions during piloting. The figure plots the share of participants ranking a given conversation outcome highest or second-highest, separately for conversations with the outgroup (Panel A) and conversations with the ingroup (Panel B).

FIGURE A6: Beliefs, Hostility, and Avoidance of Interactions

Panel A: Beliefs about Conversation Outcomes



Panel B: Beliefs about Conversation Impact

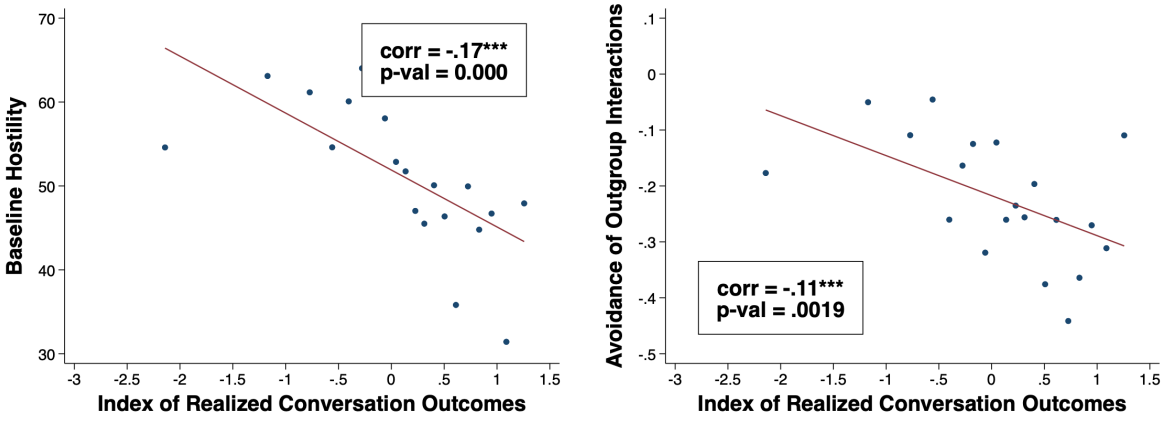


Notes: This figure shows binned scatter plots of outgroup hostility and outgroup avoidance against predictions of conversation outcomes in the experiment (Panel A) and predictions of conversation impact on hostility (Panel B). Predictions of conversation outcomes is a standardized index of the predictions about conversations with the outgroup shown in Figure 6. Predictions of conversation impact on hostility is measured as participants' prediction of the average response of other participants to the question whether they feel more positively towards the outgroup after a conversation.

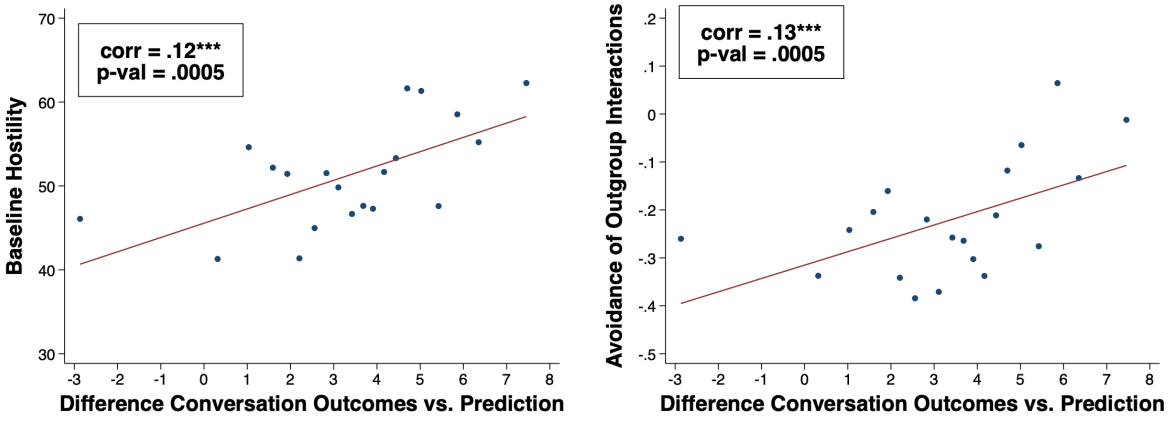
A3. Impact of Conversation Treatment

FIGURE A7: Conversation Outcomes, Hostility, and Avoidance of Interactions

Panel A: Conversation Outcomes

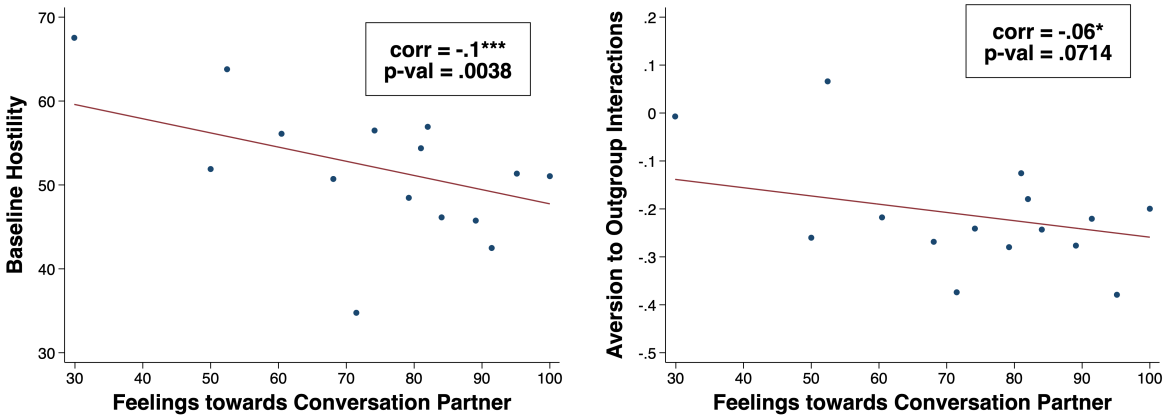


Panel B: Difference between Individual Outcomes and Predictions



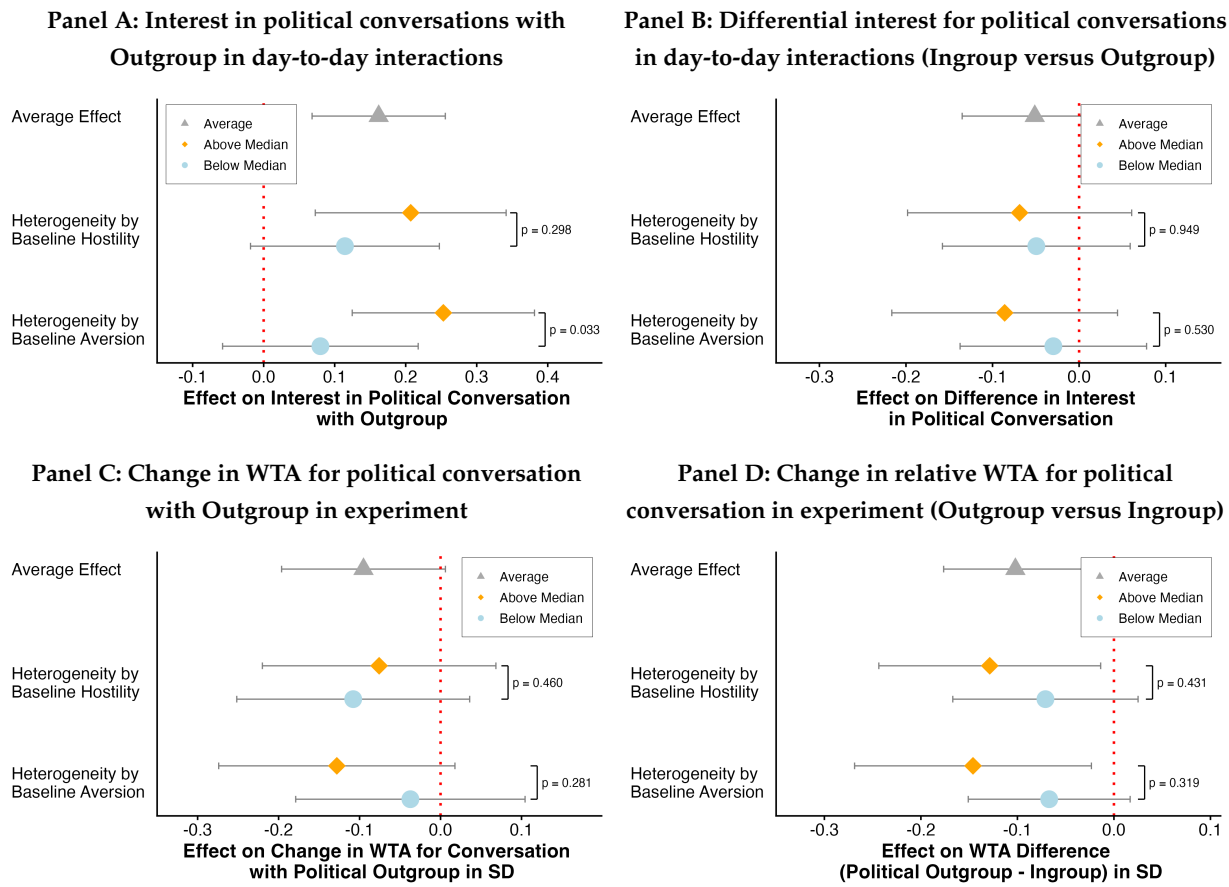
Notes: This figure shows binned scatter plots of outgroup hostility and outgroup avoidance against actual conversation outcomes in the experiment (Panel A) and the difference between actual outcomes and predictions (Panel B). Actual outcomes are measured as the average of the outcomes of conversations with the outgroup shown in Figure 6. The difference between outcomes and predictions is the average of the differences between actual outcomes and predictions across the outcomes shown in Figure 6.

FIGURE A8: Feelings towards Conversation Partner, Hostility, and Avoidance of Interactions



Notes: This figure shows binned scatter plots of outgroup hostility and outgroup avoidance against feeling ratings towards the conversation partner after the conversation.

FIGURE A9: Treatment Effects on Avoidance of Interactions with Outgroup



A4. Impact of Video Treatment

FIGURE A10: Treatment Effects of Video Treatment on Beliefs and Interest in Cross-partisan Conversations