

# Does Contact Reduce Affective Polarization? Field Evidence from Germany

## Working Paper

Adrian Blattner\*

Martin Koenen

July 2023

### Abstract

We analyze whether and how exposure to political opponents can impact attitudes towards political opponents (*affective* polarization) and extremity of political opinions (*ideological* polarization). We present findings from a quasi-experiment in Germany that matched 15,000 participants for a virtual one-on-one conversation with a stranger. Leveraging staggered treatment assignment, we find significant reductions in affective polarization among treated participants in both incentivized economic interactions and survey outcomes. The reductions are concentrated among participants who are more polarized and less interested in conversations at baseline. In contrast, we do not find corresponding effects on *ideological* polarization suggesting that exposure increases tolerance but not support for opposing positions. In ongoing work, we are extending the analysis to a series of field experiments in Brazil and the U.S. to study factors that drive demand for contact and mechanisms explaining under which conditions contact leads to durable reductions in animosity.

---

\*Blattner: Stanford University (email: ablatt@stanford.edu); Koenen: Harvard University (email: martin\_koenen@g.harvard.edu). We thank My Country Talks for sharing their data and in particular Sara Cooper and Hanna Israel for their dedication and trust in our project. Blattner thanks Pascaleine Dupas and Muriel Niederle for their invaluable guidance. We also thank Doug Bernheim, Nicholas Bloom, Matt Gentzkow, Caroline Hoxby, Stefanie Stantcheva, Noam Yuchtman, and seminar and conference audiences at Stanford, Harvard, the ESA North America Meeting 2021, the American Association for Public Opinion (AAPOR) conference, and the CAGE Summer School for Political Economy 2022 for helpful comments. We thank the Stanford Center on Philanthropy and Civil Society (PACS), Stanford Institute for Research in the Social Sciences (IRISS), Stanford Europe Center, Stanford Institute for Economic Policy Research (SIEPR), Harvard Institute for Quantitative Social Science (IQSS), Harvard Lab for Economic Applications and Policy (LEAP), and the Harvard Graduate School of Arts and Sciences (GSAS) for financial support for this project.

# 1 Introduction

Animosity between political groups has increasingly become a focus of debate in the media and policy discussions. In contrast to mere disagreement about policy positions across party lines, commonly referred to as *ideological polarization*, *affective polarization* describes the degree to which political opponents dislike and distrust each other (Iyengar et al. 2019). High degrees of affective polarization pose a potential threat to the functioning of democratic institutions and social cohesion more broadly (Boxell et al. 2022; Finkel et al. 2020).<sup>1</sup> The consequences of affective polarization go beyond political spheres, shifting individual preferences in economic interactions, such as bargaining and the selection of job applicants and workplaces (Colonnelli et al. 2022; Gift and Gift 2015; McConnell et al. 2018; Michelitch 2015).<sup>2</sup>

Affective polarization is tightly linked to homophily in social interactions. In the U.S., social networks are increasingly shaped by partisan sorting, influencing friendships, the selection of dating partners, and the marriage market (Huber and Malhotra 2016; Iyengar et al. 2018). It remains an open question whether social segregation across party lines can be attributed to rising affective polarization and whether it may in turn exacerbate partisan animus further (Iyengar et al. 2019). For instance, increased homophily of social networks may lead to misperceptions about political opponents that have been shown to influence affective polarization (Ahler and Sood 2018).

In this paper, we ask whether contact with members of a political outgroup affects levels of affective and ideological polarization. While observational data suggest a correlation between the degree of exposure to political opponents on the one hand and affective polarization on the other hand, this relationship may be driven by other unobservable factors and the direction of causality is unclear: More polarized individuals may seek less contact, but lower degrees of contact may simultaneously cause individuals to be more polarized.

To uncover the causal effect of contact on affective polarization, we partner with the non-profit organization *My Country Talks* (MCT). MCT collaborates with national media outlets in various countries to recruit and match participants for a conversation with a stranger with different political positions. As of 2023, MCT has facilitated conversations for more than 200,000 participants in 33 countries. In this paper, we study the 2021 round of *Germany Talks* (GT) — MCT’s largest national initiative — which recruited about 15,000 participants in Germany from May to September 2021 for an unguided virtual conversation with a stranger. We collect baseline and endline surveys for a subset of participants to study characteristics of the conversations and assess the effects of contact on participants’ attitudes towards political opponents (*affective polarization*). We also consider the effects of conversations on participants’ political positions (*ideological polarization*), party support, and voting intentions.

Using our sample of around 4,400 respondents who completed both the registration for GT and our baseline survey, we confirm a strong relationship between baseline levels of contact and affective polarization: Respondents who have close contacts (family members, friends, or co-workers) in parties they do not support (outgroup parties) tend to have substantially warmer feelings towards voters of these parties and exhibit lower levels of affective polarization.

About one week after participants registered for GT, they are matched to a partner who disagrees on several policy domains and asked to confirm the assigned partner. If both partners confirm the assigned match, participants receive each other’s email addresses and then independently schedule a time for a conversation. We find that in about half of the matched pairs, one or both participants do not accept each other. Close to 90% of the respondents who had a conversation with their assigned partner report that their conversation lasted for more than one hour, and 33% report conversations lasting longer than two

---

<sup>1</sup>Recent work finds that experimentally increasing affective polarization in the short run does not affect support for democratic norms (Broockman et al. 2022).

<sup>2</sup>A potential consequence is significant assortative matching in labor markets that may outweigh common sources of discrimination, such as gender and race in some contexts (Colonnelli et al. 2022).

hours.

To assess whether the conversations had an impact on the participants' degree of polarization, we exploit the staggered nature of the treatment. At the time at which participants complete our endline survey, some respondents already had a conversation with their partners while others are scheduled but did not meet their partner yet. This variation stems from two sources. First, the organizers matched participants who signed up for GT in a given week on a date and time unknown to participants. To harness this feature for identification, we vary the timing of the endline survey invite: we group individuals who completed the baseline survey (and hence signed up for GT) shortly before and shortly after the weekly matching round was conducted and invited both types of individuals to complete our endline survey at the same time.<sup>3</sup> The second source of variation stems from the scheduling process: Since participants need to schedule meetings independently, the timing of the conversation may vary depending on scheduling conflicts between the two partners.

We combine both sources of variation for our identification strategy and estimate the effect of conversations by comparing individuals who already had a conversation to those who have scheduled one, but are yet to have the conversation. This strategy relies on the assumption that variation in scheduled meeting times is uncorrelated with potential outcomes among those participants who already scheduled a meeting. To provide support for this assumption, we conduct extensive balance tests based on a range of demographic characteristics and baseline polarization measures and demonstrate that our control and treatment groups are indeed comparable based on observable characteristics.

Our first set of results focuses on the effect of conversations on affective polarization. Our main proxies for affective polarization are two incentivized economic interactions. In these interactions, participants are asked to allocate a fixed budget of 100€ between themselves and a real counterpart whose party identity we vary at random. For one half of the sample, the economic interaction is a one-shot game (dictator game), while for the other half the amount shared by the respondent is tripled by the researcher and their counterpart then decides what share of the tripled amount to send back to the respondent (trust game).<sup>4</sup>

We find that the conversations lead to significant reductions in affective polarization, measured within two weeks after exposure. Pooling the results from the dictator and trust games, we estimate a treatment effect of -0.40 standard deviations (SD) relative to a base level of outgroup discrimination towards supporters of the two parties the respondent ranks least favorably of 0.52 SD in the control group. This corresponds to a 77% reduction in affective polarization as a result of the conversation. The effect is comparable across games and robust to the inclusion of a wide range of demographic controls and survey measures of baseline affective polarization levels. In line with these results, we also find meaningful reductions in affective polarization ranging from 0.11 to 0.22 SD based on feeling thermometer questions, an unincentivized survey measure that is commonly used in the economics and political science literature (Boxell et al. 2022; Iyengar et al. 2019). In both the incentivized and unincentivized outcomes, the effects are driven by a combination of a reduction in outgroup discrimination and a reduction of ingroup favoritism.

Our second set of results focuses on the impact of the GT conversations on ideological polarization and political engagement. We do not find evidence for an effect of conversations on either of these outcomes. We estimate treatment effects that are insignificant and small in magnitude. We confirm this null result in an incentivized donation decision about a politically salient and divisive issue (climate change). While there are large differences in respondents' propensity to donate that correlate strongly with party affiliation, the conversations did not affect the likelihood to donate. Finally, we find no evidence that

---

<sup>3</sup>This procedure leads to the vast majority of respondents receiving the invite to the endline survey between one and two weeks after the baseline survey.

<sup>4</sup>Respondents are informed that their actions matter as the actions of two randomly selected respondents were implemented.

conversations affected participants' stated intentions to vote in the upcoming election, party choice, or the relative ranking of parties.

Taken together, we interpret these results as evidence that while the kind of contact with individuals with different political views generated by the MCT initiative can lead to meaningful reductions in affective polarization in the short run, the same treatment does not affect participants' political views. This underscores the fact that affective and ideological polarization are two distinct concepts, and policies aimed at addressing polarization need to take this dichotomy into account.

While the experimental design does not directly allow us to test for the mechanisms driving the observed effects in a causal manner, we provide descriptive evidence on several self-reported measures of conversation content and outcomes. First, we find that conversations are not limited to the political realm: virtually all participants (99%) indicate that they talked about personal topics with their partner as well. Moreover, we find evidence that baseline disagreement in a given political domain, which is conditionally randomly assigned in the matching process, is strongly predictive of coverage of that political topic during the conversation.

We also elicit several self-reported measures of conversation outcomes. Only a small share of respondents (6%) report believing that they strongly convinced their partner about their own political position. This finding matches our observed null effect on respondents' political positions and is evidence against the hypothesis that effects are driven by a convergence in political opinions as a result of the conversation. In contrast, we find some support for the importance of learning about the motives for opposing policy positions and the existence of at least some shared policy goals. A vast majority of respondents (73%) state that they were able to understand the arguments and reasons for the political views of their conversation partner, despite disagreeing with the positions themselves. Moreover, more than half of respondents identified a strong overlap in positions with their partner (57%). We also find suggestive evidence for the importance of the personal relationship built during the conversation. 71% of respondents state to have identified at least some personal similarities with their partner and the average thermometer rating for their conversation partner (74.6) is close to the average rating for voters of respondents' most preferred party (77.9).

What is the external relevance of our results? Our study focuses on a population that actively sought contact with an individual who holds different political opinions. Moreover, by design participation was only possible for consumers of participating media outlets who encountered and then engaged with survey questions on the respective website. Among this sample, we estimate treatment effects for participants who follow through with the full sign-up procedure of Germany Talks and who further respond to our baseline and endline survey. We provide two pieces of evidence suggesting that the effects may at least partially translate to a more general target population. First, using heterogeneity in baseline characteristics within our sample, we find that the effects are concentrated among respondents who are more polarized and who are less interested in outgroup-contact ex-ante. Moreover, we show that our main results are robust to several approaches of re-weighting our sample so that it matches the German population along various observable dimensions such as gender, education status, and voting behavior.

An important follow-up question is whether the effects we measure in the context of GT would replicate in a sample of participants who did not actively seek contact. We focus on this angle in ongoing work in collaboration with the organizers, in which we recruit representative samples and test the role of several recruiting strategies.

This paper relates to several strands of literature. First, it contributes to a vast literature in economics and political science testing the contact hypothesis, first formulated by Allport (1954). In line with its premise, studies have documented significant effects of exposure to individuals with a different ethnicity (Bursztyn et al. 2021; Finseraas et al. 2019), immigrant background (Bailey et al. 2022), religion (Jha 2013), or socio-economic status (Rao 2019), individuals from a different caste (Lowe 2021), or transgender people (Broockman and Kalla 2016) on attitudes towards members of the respective outgroup. Within this

domain, the growing literature on the effect of cross-partisan contact on affective polarization has also gained increased attention in political science.

While to the best of our knowledge, to date no systematic evaluation of efforts to facilitate cross-party contact exists (Iyengar et al. 2019), some existing work has recruited subjects for conversations, randomizing both exposure to conversations and conversation topics based on prompts or priming articles. Santoro and Broockman (2022) find that 10-minute conversations between a Republican and a Democrat reduce outparty animus in the short-run but only if the conversation topic is apolitical. In contrast to their findings, in our setting, we observe a treatment effect of conversations even among participants discussing divisive political topics at great length. Levendusky and Stecula (2021) combine a cross-partisan conversation treatment with texts that highlight partisan similarities and find significant reductions in affective polarization as a result of the combined treatment. Fishkin et al. (2021) study an intervention that exposes participants to moderated group discussions. In contrast to the existing studies, we study cross-partisan interactions in a field environment with little to no guidance of the conversations. Moreover, we extend the analysis of the treatment effect of the conversations to a wider range of outcomes that include incentivized economic measures.

The most closely related work to our study is a working paper by Heuser and Stötzer (2022) who study the 2018 round of Germany Talks and estimate that conversations among participants with *higher* levels of baseline policy disagreement reduce the likelihood that respondents self-report that people with opposing political views are associated with negative stereotypes.<sup>5</sup> Moreover, they estimate increases in the extremity of stated policy positions for conversation among participants with *lower* levels of baseline policy disagreement. We are able to push the analysis further in a number of ways: we draw on an identification strategy that relies on significantly weaker assumptions,<sup>6</sup> have significantly lower levels of attrition,<sup>7</sup> include both unincentivized and incentivized outcomes of party-by-party levels of affective polarization and policy domain-specific ideological polarization that allow for direct comparisons of our results to similar interventions, and have a rich set of baseline covariates measured before treatment assignment.<sup>8</sup>

Second, this paper aims to complement a literature in political economy and political science documenting and assessing factors contributing to the rise of ideological and affective polarization (Ahler and Sood 2018; Autor et al. 2020; Bail et al. 2018; Boxell et al. 2017, 2020; Iyengar et al. 2019; Levendusky 2018; Levy 2020; Martin and Yurukoglu 2017). In contrast to the majority of existing studies focusing on identifying factors that have *aggravated* affective polarization in the past, the objective of our study is to assess the potential effects of an initiative that explicitly aims at *reducing* polarization. Moreover, we generate evidence on the role of network homophily and lack of exposure to political opponents, a question that is still at the center of debate in the literature.

Section 2 describes the institutional background and structure of Germany Talks. Section 3 presents the evaluation design, the identification strategy, and outcome measures. Section 4 presents our results. Section 5 presents evidence on mechanisms and heterogeneity results. Section 6 discusses the implications of our findings and future work.

---

<sup>5</sup>Respondents were asked to picture a person that gave very different answers to the sign-up questions and state their agreement with several negative stereotypes, including whether they have different moral values, low cognitive abilities, or are poorly informed.

<sup>6</sup>Our identification strategy relies on the assumption that variation in scheduled meeting times is uncorrelated with potential outcomes among those participants who already scheduled a meeting. Heuser and Stötzer (2022)'s approach relies on the assumption that participants who were not accepted by their partners are comparable to those who were, conditional on a set of covariates. However, the authors do not observe the full set of covariates (in particular the open-ended sign-up responses) necessary for this approach.

<sup>7</sup>Around 75% of our respondents who completed the baseline sample start the endline survey, compared to less than 50% in Heuser and Stötzer (2022)'s study.

<sup>8</sup>The baseline survey in Heuser and Stötzer (2022) was filled out by most subjects after treatment status was assigned.

## 2 Germany Talks

The goal of this paper is to study whether contact with individuals with different political views has an impact on affective and ideological polarization. In observational data, it is difficult to simultaneously observe variation in contact with political opponents and have reliable outcome data available since affective and ideological polarization are typically measured using survey questions. In this paper, we thus study a unique setting that both generates individual-level variation in exposure to political opponents and allows us to track the effect of contact on relevant outcomes.

We evaluate Germany Talks, an initiative founded in 2017 by the German news outlet *ZEIT ONLINE*<sup>9</sup> that annually matches strangers for one-on-one conversations based on their political opinions. GT's objective is to generate "*discussions between people with completely different views*" and thereby establish "*a new form of political debate*".<sup>10</sup> Over the past years, the initiative has grown — it has been replicated in 33 countries (including the U.S.) with more than 200,000 participants — and is now officially organized by the My Country Talks (MCT) initiative.

We partnered with MCT to evaluate its 2021 round in Germany. GT has typically attracted the largest numbers of participants who are recruited by a diverse set of media outlets: the 2021 round we study was co-organized by numerous prominent media outlets,<sup>11</sup> covering a large share of the political spectrum and affiliates of all major political parties. In the year of our study, GT recruited participants over a period of five months, leading up to the federal elections held in September 2021.

### 2.1 Institutional Background

To put our study and its findings in context, it is important to highlight several key features of Germany's political system. Germany has a multi-party system with six parties represented in the parliament at the time of GT 2021. Since WWII, all German governments were formed through coalitions led by one of the two parties in the center-left (Social Democratic Party, SPD) and center-right (Christian Democratic Union, CDU/CSU). Since the 1990s, the main coalition partners were the Green Party (Greens) and the Free Democratic Party (FDP). In addition to these four parties, the far left (The LEFT) and the far right (AfD) also hold seats in the parliament.

In recent decades, the German parliament has become increasingly fragmented with the vote share of the top two parties dropping from 77% in 2002 to 50% in 2021. In the same period, parties on the far-right and far-left have gained traction among voters, indicating that polarization in positions *across* parties is a relevant feature of the German context.<sup>12</sup> Recent comparative work argues that *affective* polarization levels in Germany are similar to those in other Western democracies (Gidron et al. 2020; Wagner 2021). Animosity between supporters of the two historically largest parties in Germany, as well as other measures of overall polarization, have somewhat decreased in the last decades (Boxell et al. 2022).

### 2.2 Recruitment

GT 2021 recruited more than 15,000 participants online via the websites of participating German media outlets. Panel A of Figure 1 shows two front-page articles about GT 2021 published on the websites of

<sup>9</sup>ZEIT ONLINE is the online version of the newspaper *Die ZEIT*, one of the best-selling weekly newspapers in Germany, and has more than 10 million unique users annually. See <https://www.iqdigital.de/Portfolio/Digital/ZEIT-ONLINE>.

<sup>10</sup>See mission statement at <https://www.mycountrytalks.org>.

<sup>11</sup>The participating media outlets are Frankfurter Allgemeine Zeitung (FAZ), Freie Presse, Handelsblatt, Norddeutscher Rundfunk (NDR), Stern and ZEIT ONLINE.

<sup>12</sup>The right-wing populist party AfD was founded in 2013 and increased its vote share in the federal elections from 5% in 2013 to 13% in 2017 (10% in 2021). At the time of GT 2021, it was the fourth-largest party. On the left end of the political spectrum, The LEFT has increased its vote share from 2.4% in 1990 to 9% in 2017 (5% in 2021).

two major German newspapers in May 2021, illustrating the salience of the initiative.

In addition to these front-page newspaper articles launched at the beginning of the initiative, opportunities to sign-up for GT were embedded in a large number of online articles from participating media outlets from May until September 2021. As part of an online article, readers encountered survey questions asking them about their agreement with a policy position. If a reader engaged with the survey question, they were presented with seven additional survey questions. The survey questions covered issues from the current public debate, such as COVID-19, climate change, economic inequality, and gender equality.

If a reader completed all eight questions, they were informed about the opportunity to be matched with another person who has responded differently to the eight policy positions. GT then collected responses to four additional open-ended questions and basic demographics (gender, age, postal code).<sup>13</sup>

To provide some intuition on the sign-up questions, Table A.1 shows the distribution of all sign-up questions used in GT 2021.<sup>14</sup> While the proportions of participants agreeing or disagreeing with the various statements vary, the large majority of questions achieves a split that is between 30% and 70% indicating that there is a significant degree of disagreement across statements. In addition, Table A.1 also highlights that responses to the sign-up questions correlate with the political parties a given participant supports.

Panel B of Figure 1 maps the geo-coded location of the 15,000 participants who completed the registration process. The map highlights that participants are recruited from all states in Germany, with a higher concentration of participants in cities and areas with larger populations.

## 2.3 Matching and Confirmation Process

The goal of GT is to match a registered participant with a partner who holds different political views. To achieve this goal, the organizers use a greedy algorithm that takes the responses to the binary policy questions from the sign-up process as input and aims to maximize the number of differences in those questions. In 2021, GT conducted the matching repeatedly over the course of a five-month period between May and September 2021. The algorithm is run at the discretion of the organizers, roughly once a week and includes all individuals who signed up in a given week or could not be matched in a previous week.<sup>15</sup> We illustrate and describe this process in greater detail in Appendix Section A.1.

After two participants are matched, both receive an email informing them about the successful match. The email contains several pieces of information about the assigned partner: their age, gender, first name, text responses to open-ended questions, and to which sign-up policy questions they responded differently. Both partners then have the option to confirm the partner through a link at the end of the email. If both partners confirm, they receive another email with the contact details of each other.

## 2.4 Scheduling and Conversations

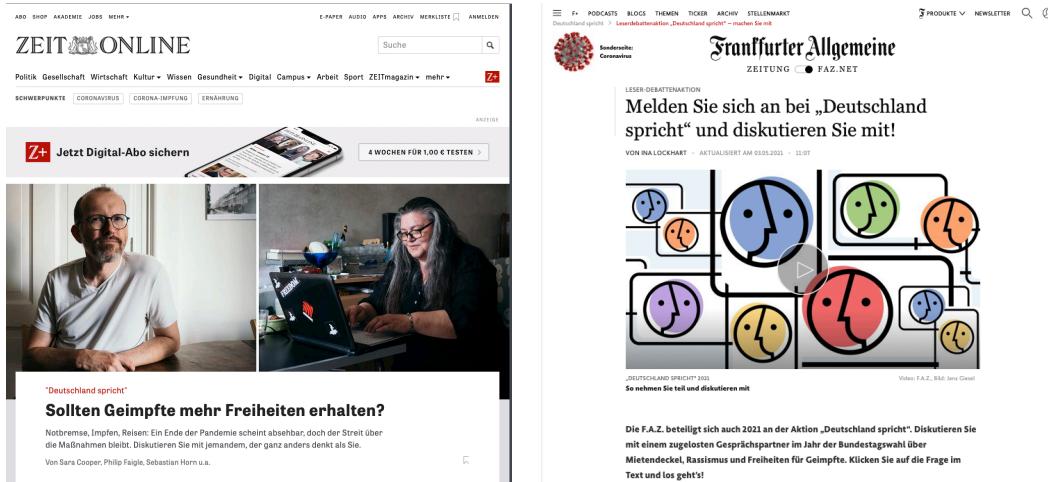
After having received the email address of their assigned match, participants are asked to reach out to each other to schedule a time to meet and have a conversation. GT refrains from setting a specific date or time so that participants can find the time that works best for them. More importantly, the conversations

<sup>13</sup>The open-ended questions asked what the participant is currently worried about, what they're hoping for, what their current occupation is, and what they dislike.

<sup>14</sup>The data is comprised of individuals who also participate in our surveys (the sample selection is discussed in more detail in Section 3).

<sup>15</sup>From this participant pool, the algorithm then generates random subsets of the sample for easier processing. Within a given match sample, it picks a random participant and picks a match that maximizes the level of disagreement, measured by the number of diverging responses on the binary sign-up questions. The algorithm then iterates over all participants and match samples. It leaves a subset of participants (5-10%) unmatched if it cannot find a partner with a distance in sign-up questions of at least three.

### (A) Newspaper Coverage



### (B) Map of Participants

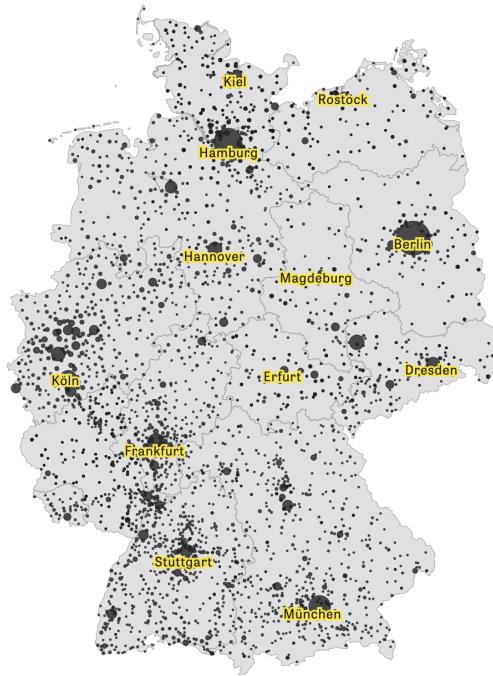


FIGURE 1: Germany Talks 2021

*Notes:* Sources: ZEIT ONLINE (2021), ZEIT ONLINE (2020)

are not structured or explicitly guided: while GT informs participants of each other's responses to the sign-up question, they do not make any suggestions on conversation topics or even the length of the conversations. As a result, GT merely facilitates the occurrence of the conversation but does not control its structure or content. So far, the organizers of GT had only limited insight into the topics and outcomes of the conversations. We characterize the conversations in more detail for those individuals who participate in our surveys in Section 5. Finally, in contrast to prior rounds of MCT, GT was held fully remotely in

2021 due to COVID-19 and participants were asked to meet their match online.

### 3 Evaluation Design

#### 3.1 Overview

Figure 2 summarizes the design of the evaluation. Registration for Germany Talks was open from May to September 2021. Germany Talks ran the matching process approximately once a week, including all participants who had signed up since the previous matching round. There were 20 matching rounds prior to the federal elections in September 2021. The date of the matching was not known by participants before or during the registration process. This process coupled with the matching algorithm described above resulted in about 95% of participants being matched each round.

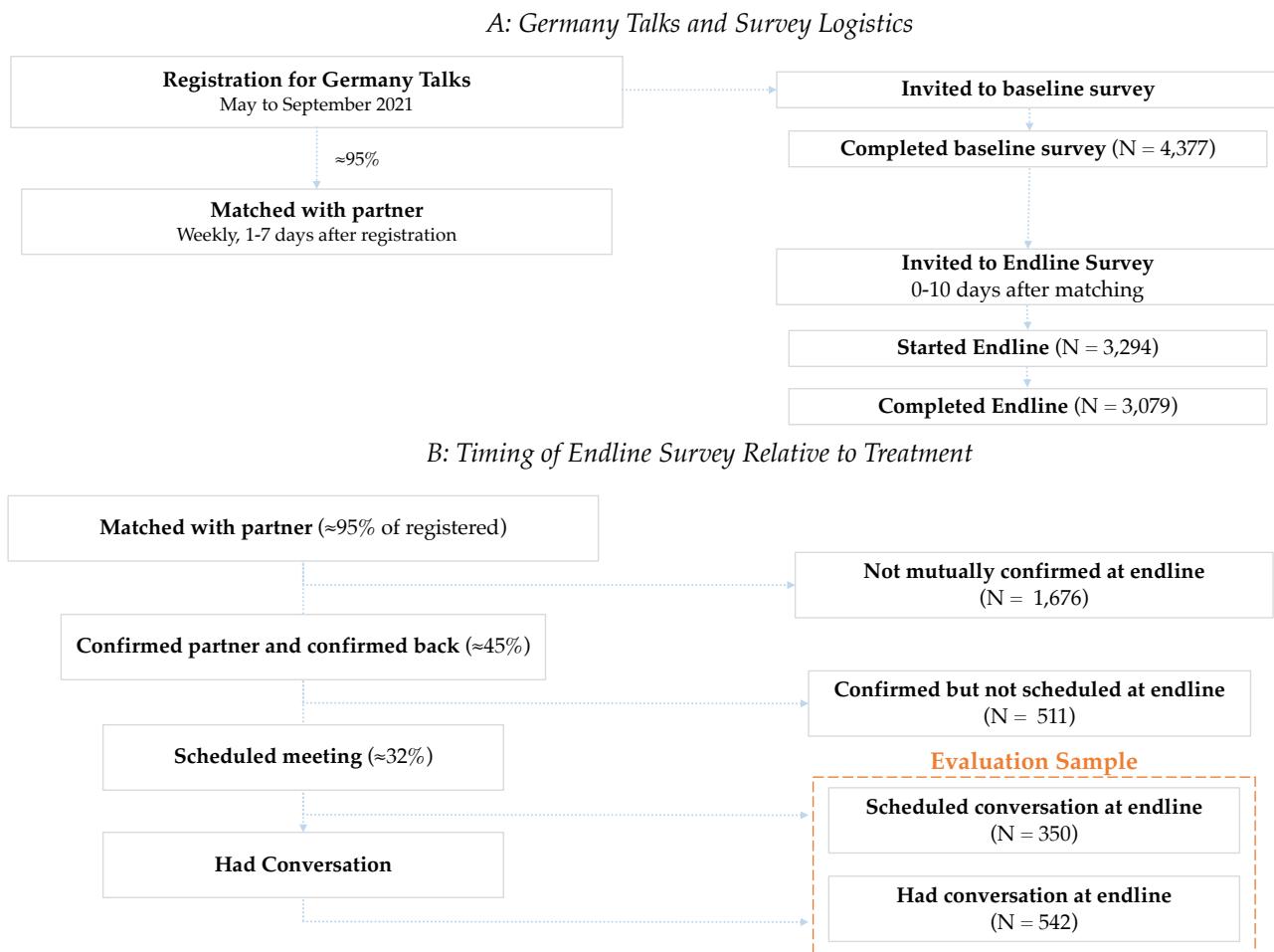


FIGURE 2: Evaluation Design

Upon completion of the registration process, all participants of Germany Talks were automatically invited to participate in our study. Importantly, our study was advertised as a study about political opinion in Germany during the COVID-19 pandemic. In addition, all questions about Germany Talks and the GT conversations were displayed at the end of the surveys to avoid revealing the study's focus on the effects of GT. Completion of our surveys was incentivized with a gift card lottery amounting to around €5,000 in total. About 4,400 participants of Germany Talks completed our baseline survey and

passed basic attention and quality checks.<sup>16</sup> We invited participants to our endline survey within 10 days after their respective matching date. About 3,100 participants of Germany Talks completed both the baseline and endline survey.

To address potential concerns over selective attrition, Appendix Table A.2 presents summary statistics for individuals who only respond to our baseline sample (column 1) and contrasts those to the respective values among the sample of those responding to both of our surveys (column 2). Column 3 presents the difference between the two and column 4 shows p-value corresponding to t-tests testing for differences between the two groups. We find that respondents who do not complete the endline survey tend to be less educated on average, earn lower incomes, are less likely to be working, and have somewhat lower interest in politics. Yet, importantly we observe virtually no differences with respect to various measures capturing baseline ideological and affective polarization (Panel B).<sup>17</sup>

### 3.2 Sample

Table 1 shows baseline demographics of GT participants, GT participants who completed our baseline and endline surveys, and statistics for the German population. Comparing columns 1 and 4, we find that GT participants have a similar average age and likelihood to live in East Germany, relative to the full German adult population. However, GT participants are predominantly male (73% vs 50%), relative to the full population.

Comparing our survey sample (column 2), for which we collected a larger set of covariates, to the German population (column 4), we find that our sample is representative along various key dimensions: The average age (50.2 years vs. 51.9 years), marriage rates (51% vs. 50%), and employment rates (77% vs. 78%) are comparable. However, participants in GT are substantially more educated than the full German population: around two-thirds of participants have a college degree in our sample compared to just one third of the German population. In addition, GT participants are more likely to have high annual incomes exceeding 60,000 € (30% relative to 22% of Germany's population) and to be strongly interested in politics (35% relative to 17% in nationally representative data).

Importantly, our sample achieves a broad representation of political opinion in Germany. Our sample covers voters of the full political spectrum from the far-left (The LEFT) to the far-right (AfD). The proportion of individuals who voted for one of the two extreme parties is somewhat lower compared to the full population (15% vs. 22%).

To address the fact that our sample is selected along various dimensions, we assess the robustness of our primary findings to various ways of re-weighting the sample. This exercise re-weights observations so that it more closely matches a representative sample of Germany's population along *observable* characteristics. Moreover, in Section 6, we further discuss how to tackle potential concerns over selection along *unobservable* characteristics in ongoing work.

In addition to the comparisons with the German population, we assess how our sample compares to the full set of participants of GT. Yet, this analysis is limited by the fact that the organizers collect only limited information on participants' backgrounds. We find that our sample is comparable to the full sample of GT participants: for the full sample (described in column 2 of Table 1, the difference in average age lies within 1.5 years (50.2 vs. 48.9), and women are slightly over-represented in our sample (33% vs. 27%). Finally, the proportion of individuals residing in West Germany is similar (80% vs. 78%).

Panel C of Table 1 displays summary statistics regarding the matches based on admin data by GT: in around two thirds of the matches, the assigned partners have the same gender, and about one third of

---

<sup>16</sup>These checks included a question asking participants to enter a fixed number to indicate that they are paying attention. We also dropped participants with incomplete responses to survey questions used as main outcome measures and covariates.

<sup>17</sup>Since we observe treatment take-up only for respondents who completed the endline survey, we cannot conduct an attrition test for differential treatment take-up.

TABLE 1: Sample Composition

	Registered for Germany Talks		Germany	
	Completed Baseline and Endline Surveys			
	(1)	(2)		
<i>Panel A: Demographic Characteristics</i>				
Female	0.27	0.33	0.50	
Age	48.90	50.21	51.89	
East Germany	0.22	0.20	0.19	
College degree		0.66	0.37	
Immigrant		0.13	0.27	
Income > 60,000		0.30	0.22	
Working		0.77	0.78	
Married		0.51	0.50	
Very strong pol. interest		0.35	0.17	
General trust		0.44	0.46	
Voted center left in 2017		0.40	0.30	
Voted center right in 2017		0.31	0.44	
Voted far-left/right in 2017		0.15	0.22	
Plan to vote in 2021		0.92	0.80	
<i>Panel B: Baseline Affective Polarization Levels</i>				
Party Feeling Thermometer Min		7.29	5.27	
Party Feeling Thermometer Max		79.94	83.05	
Party Feeling Thermometer Range		72.65	77.78	
<i>Panel C: Match Characteristics</i>				
Same gender	0.65	0.54		
Similar age (+/- 10 years)	0.34	0.34		
Same state	0.11	0.10		
Signed up through same media outlet	0.38	0.36		
Different responses in sign-up questions	6.80	6.08		
N	14456	3079	2196	
			83m	

Notes. This table reports characteristics of Germany Talks participants (1), Germany Talks participants who completed the baseline and endline survey (2), the subset of respondents we could match with admin data from Germany Talks (3), and the German population (4). Admin data is not available for respondents who asked Germany Talks to delete their data before the data transfer was conducted or who used a different e-mail address for the surveys than for Germany Talks. Statistics for Germany are calculated based on data obtained from the German Statistical Office, the Micro Census, the German Longitudinal Election Study (GLES), and the World Values Survey. East Germany includes Berlin. Immigrant refers to respondents who indicated to have an immigration background (at least one parent or respondent not born in Germany). Party thermometer values for Germany are from GLES and ask for feelings towards parties, while the party thermometer in our survey sample asked for feelings towards party supporters.

all matches consist of partners of similar age (within ten years) and partners who signed up through the same participating media outlet. Finally, partners on average gave different answers to about seven of the sign-up questions, which corresponds to more than three quarters of all sign-up questions. This indicates that the policy views of the majority of matched pairs differed substantially.

For a subset of our sample (column 3), we were able to merge admin data from GT, allowing us to observe characteristics of the assigned partner for about two thirds of our survey sample.<sup>18</sup> Relative to the full sample of GT participants, the matched pairs in our sample are somewhat less likely to have the same gender (54%) and have somewhat lower levels of disagreement in the policy positions (6.08 questions).

<sup>18</sup>About one third of our survey respondents could not be matched with the GT admin data for two reasons: they either used a different email address in our survey than to register for GT or they asked GT to delete their personal information before we obtained the admin data from GT.

### 3.3 Outcome Measures

We consider two groups of outcome measures. Our primary outcomes are survey-based and incentivized measures of animosity towards voters of other parties (affective polarization). The secondary outcomes are survey-based and incentivized measures for polarization in political positions (ideological polarization). In addition, we measure respondents' political engagement, voting behavior, and several beliefs related to polarization.

#### A. Survey Measures of Affective Polarization

Our main outcome of interest is the participants' levels of affective polarization. In the political science and economics literature, attitudes towards political opponents are usually captured using unincentivized survey questions. The predominantly used measure is a feeling thermometer, which asks survey respondents to rank their feelings towards supporters of a party on a scale from 0 to 100, where values 0-49 indicate cold feelings, values 51-100 warm feelings, and a value of 50 neutral feelings.<sup>19</sup> In our analysis, we replicate the feeling thermometer for all six parties in the German context.<sup>20</sup> Panel B of Table 1 shows that the average thermometer for respondents' most preferred party is 80, compared to just 7 for their least preferred party. This corresponds to a significant range of 73 and is somewhat smaller compared to representative surveys for the German population (78).<sup>21</sup>

In addition to the feeling thermometer for party supporters, we construct two additional measures of affective polarization to address two shortcomings: First, to capture animosity towards supporters of policy *positions* rather than parties, we include thermometer questions for five policy issue areas. These ask respondents how they feel towards supporters of certain political *positions* irrespective of their personal opinion of the issue at hand.<sup>22</sup> Second, to address issues related to image concerns and experimenter demand effects due to the unincentivized nature of the above survey measures, we use an incentivized economic interaction as our main outcome: modified versions of the dictator game and the trust game.

#### B. Dictator and Trust Game

In the endline survey, we randomize each respondent to play either the dictator or trust game in the endline survey. In both games, the respondent plays the game as Player 1 and is allocated a budget of 100€. The respondent is asked to split the budget between themselves and another person (Player 2). We incentivize decisions by implementing the actions of a randomly chosen respondent in each game, which is known by the respondents. Subjects are also informed that Player 2 is not their conversation partner from GT but rather a respondent in a similar online survey who will not receive any additional information about Player 1 other than the amount shared with them.<sup>23</sup>

Importantly, in order to measure attitudes towards political opponents, we inform the survey respondent about several randomized characteristics of Player 2. With equal probability, Player 2 leans towards one of the three parties that are most strongly represented in our sample (Greens, CDU/CSU, and FDP).

<sup>19</sup>Other commonly used survey outcomes include questions about personality traits associated with the outgroup (e.g. patriotic, honest, selfish), and attitudes towards being close friends with members of the outgroup, or having children marry someone from the outgroup (Druckman and Levendusky 2019; Iyengar et al. 2012; Iyengar and Westwood 2015).

<sup>20</sup>The feeling thermometer was visualized and respondents were asked to state their response on a slider.

<sup>21</sup>The survey questions for the full German population ask for feelings towards parties rather than party supporters. The range is larger than comparable measures from the most recent wave of the Annual National Election Study (ANES) for the U.S. (about 70 for the ingroup versus about 20 for the outgroup).

<sup>22</sup>We include a positive and a negative statement for each issue area and use the standardized difference between the two as a polarization metric.

<sup>23</sup>It's important to highlight that this removes additional considerations of Player 1 about Player 2's reaction to Player 1's characteristics. Most importantly, it eliminates any concerns about Player 1's privacy.

Further, Player 2 is randomized to belong to one of three age groups (30-39, 40-49, and 50-59) and live in either East or West Germany. Finally, Player 2 is fixed as male. These demographic traits are merely introduced for the purpose of distracting from the fact that the focus of our study lies on attitudes towards people preferring different parties.

The features described so far are common to both games, but there is one important difference between the two: while the dictator game is a single-shot game, i.e. the game ends after Player 1 has decided how much to share with Player 2, the trust game has two stages. Specifically, in the trust game, once Player 1 has decided how to split the fixed amount of money between themselves and Player 2, the amount shared with Player 2 is then tripled by the experimenter and Player 2 is asked what share of the received amount to send back to Player 1. Appendix Figure A.3 shows the instructions and an example of a randomized profile.

Differences in the amount shared with Player 2 in the dictator game depending on their group affiliation are generally interpreted as a taste for discrimination towards that group by the respondent (Fershtman and Gneezy 2001). In contrast, the trust game aims to proxy for cooperation between two players since it allows for significant gains from cooperation, even if the subgame perfect equilibrium suggests that no money should be sent by either player (Fershtman and Gneezy 2001).

Baseline statistics from these two games uncover significant discrimination and differences in trust levels. Table A.3 highlights that in the dictator game, respondents share on average 10€ more if randomized to a Player 2 who leans towards a party they rank first, second, third, or fourth, relative to those randomized to a player 2 who leans towards a party they rank fifth or sixth. This difference is statistically significant at all conventional levels. Similarly, in the trust game, respondents share 9€ more with Player 2 in that case. These results indicate that we see evidence for discrimination in both games.

### *C. Secondary Outcomes: Ideological Polarization and Political Engagement*

Our main measure of ideological polarization is based on a battery of policy statements. Respondents are asked to state their agreement on a 5-point Likert scale with policy positions in the following areas: rent control, racism, gender, environment, and COVID-19. We then use the standardized distance between responses and the average response in the sample as a measure of the extremity of positions, i.e. ideological polarization. We then also construct a standardized composite index that averages over the measures of ideological polarization in each policy area.

In addition to these survey measures, we also include an incentivized outcome to capture changes in policy views. Each respondent is asked if they wish to authorize a 500€ donation to Fridays-for-Future, a youth-led movement demanding drastic political changes to fight climate change. Respondents were informed that one respondent will be drawn randomly and their decision will be implemented. We emphasized that the researchers will pay for the donation if the randomly drawn respondent decided to authorize the donation.

Finally, we elicited respondents' intentions to vote in the 2021 federal elections in both the baseline and endline survey. All respondents who stated at least a weak intention to vote were also surveyed about their most likely party choice in the election. We also re-elicit the ranking of all parties in the endline to test for potential changes relative to the baseline survey.

## **3.4 Identification Strategy**

Our study attempts to estimate the causal effect of a conversation on affective and ideological polarization. The ideal design to identify causal treatment effects in our setting would be to randomly assign conversations to a subset of our sample. However, GT's goal is to match all participants to generate the maximum number of meetings possible. We thus rely on a staggered treatment assignment introduced

by the timing of the matching rounds and the scheduling behavior of participants to identify treatment effects.

Panel B of Figure 2 summarizes the key steps for treatment assignment in our evaluation. At the time of the endline survey, around 18% of respondents had a conversation with a partner. This number is explained by several factors. First, each week, around 5% of participants are unmatched if the matching algorithm is unable to generate a match with a sufficient distance in policy positions. Second, after pairs are matched, each participant receives an email with information about their assigned partner and has the option to decide whether or not to confirm the assigned match. This stage leaves around  $\approx 46\%$  of respondents in mutually confirmed pairs who received the contact information of their assigned partner. Finally, among matched and confirmed pairs, some pairs may have scheduling issues or fail to contact each other. These steps combined introduce selection into treatment, conditional on both the partner assigned and individual characteristics.

Table 2 quantifies the role of several baseline covariates for the selection into treatment. In column 1, we show balance tests for the sample of participants in pairs in which both partners confirmed each other. Panel B highlights that are significantly younger, more likely to be female, and more educated. Moreover, they exhibit a higher baseline interest in the conversation and lower levels of ideological polarization. In column 2, we contrast this finding against balance tests for the decision to schedule a conversation among those pairs in which both partners confirmed each other. We still find significant selection based on age, education, and baseline levels of ideological polarization.

To overcome concerns about selection into treatment, we argue that for participants who already scheduled a meeting, exposure to treatment is randomly staggered over time. This relies on the assumption that the timing of the conversation is uncorrelated with potential outcomes, conditional on having accepted (column 1) and scheduled a meeting with the assigned partner (column 2).

Variation in meeting timing, relative to our endline survey stems from two sources. First, participants may have scheduling conflicts leading them to choose a meeting date that varies by several weeks. Second, we varied the timing of the endline survey relative to the matching date. Recall that the matching is conducted weekly, drawing on all participants that have signed up until the time the matching is executed. This procedure implies that participants signing up shortly after a given matching date will receive a potential match and thus have the chance to receive treatment an entire week later, relative to a participant who signed up shortly before a given matching date.<sup>24</sup> In order to exploit this feature of the institutional design, we sent invites to the endline survey to subjects who just missed a given matching round (i.e. registered up to 72 hours after a matching round) at the same time as the invites to subjects who were still included in that same matching round. We visualize this process in Figure A.2.

To substantiate our assumption that the timing of meetings conditional on having scheduled is close to random, columns (3)-(4) of Table 2 report balance tests, conditioning on participants who scheduled a meeting. The results support the main assumption that participants who scheduled a meeting but did not meet yet and those who already met are comparable based on observable characteristics. Across a range of baseline covariates, there appear virtually no statistically significant differences between sign-ups before and after the matching deadline. A joint test for significance yields a F-statistic of about 0.5.

---

<sup>24</sup>Importantly, participants are neither informed about the date on which the matching is conducted, nor do they know about this process more generally.

TABLE 2: Selection into Conversations and Balance Test

	Selection into Conversations		Balance Test	
	Both Partners Confirmed (1)	Scheduled if Both Confirmed (2)	Met Partner if Scheduled (3)	p-value (4)
<i>Panel A: Descriptive Statistics</i>				
Mean	0.46	0.68	0.61	
SD	(0.50)	(0.47)	(0.49)	
Observations	3079	1403	892	
<i>Panel B: Correlations with Participant Characteristics</i>				
Female	0.09*** (0.02)	-0.03 (0.03)	-0.02 (0.03)	0.468
Age	-3.21*** (0.59)	2.13** (0.91)	0.42 (1.11)	0.702
East Germany	-0.02 (0.01)	-0.02 (0.02)	-0.00 (0.03)	0.883
College degree	0.06*** (0.02)	0.07*** (0.03)	0.04 (0.03)	0.217
Immigrant	-0.01 (0.01)	0.02 (0.02)	-0.03 (0.02)	0.170
Income > 60,000	0.01 (0.02)	0.08*** (0.03)	0.02 (0.03)	0.510
Working	0.01 (0.02)	0.03 (0.03)	0.01 (0.03)	0.745
Married	-0.02 (0.02)	0.03 (0.02)	0.01 (0.03)	0.751
Left leaning	0.07*** (0.02)	0.01 (0.02)	-0.02 (0.03)	0.484
Very strong political interest	0.01 (0.02)	0.03 (0.03)	-0.02 (0.03)	0.541
General Trust	0.05** (0.02)	-0.02 (0.03)	0.03 (0.03)	0.455
Voted far-left/right in 2017	-0.02 (0.01)	-0.02 (0.02)	0.00 (0.02)	0.889
Voted center left in 2017	0.05*** (0.02)	0.02 (0.03)	0.00 (0.03)	0.937
Plan to vote in 2021	0.01 (0.01)	0.01 (0.02)	-0.00 (0.02)	0.930
Affective Polarization Index	0.01 (0.04)	-0.07 (0.06)	-0.11 (0.07)	0.131
Ideological Polarization Index	-0.04** (0.02)	-0.07** (0.03)	-0.05 (0.04)	0.161
Strong Interest in Conversation	0.08*** (0.02)	0.02 (0.03)	-0.03 (0.03)	0.431
Observations	3079	1403	892	
<i>Panel C: Effects of Match Characteristics</i>				
Numer of Policy Disagreements	0.06*** (0.02)	-0.01 (0.03)	0.03 (0.04)	0.435
Partner Same Gender	-0.06*** (0.02)	0.04 (0.03)	0.00 (0.04)	0.954
Partner Age Difference < 10	0.01 (0.02)	-0.01 (0.03)	0.03 (0.04)	0.403
Partner From Same State	-0.00 (0.01)	0.04** (0.02)	0.00 (0.03)	0.986
Observations	2196	1060	690	
Sample: all participants	Yes	No	No	
Sample: both partners confirmed	No	Yes	No	
Sample: meeting scheduled	No	No	Yes	

Notes. This table reports descriptive statistics and correlations for conversation take-up decisions. Panel A reports the mean and standard deviation for the variables indicated in the header. Panel B shows results from bivariate OLS regressions, regressing the respective participant baseline characteristic on the variable in the header and an intercept. Panel C shows the results from multivariate OLS regressions, regressing the variable in the header on the respondent's sign-up policy positions, the (conditionally on the sign-up positions) randomly assigned match characteristic, and an intercept. The sample in Panel C is restricted to participants for which admin data from Germany Talks is available. Standard errors are in parentheses in Panels B and C.

## 4 Results

In this section, we present estimates of the effect of conversations on participants' degree of polarization. We begin by presenting descriptive evidence on the role of contact for affective polarization and the conversation duration and topics in 4.1. We then present our main findings on the effect of the conversations on affective in Section 4.2 and on ideological polarization as well as voting behavior in Section 4.3.

### 4.1 Descriptive Evidence

Existing research suggests that affective polarization and the degree of contact with political outgroups are strongly linked. To motivate the hypothesis that contact with members of a political outgroup could affect partisan animosity in our context, Figure 3 plots a simple measure of baseline levels of affective polarization (the difference in respondents' thermometer feelings towards supporters of respondents' own party and supporters of all other parties) against the extent to which respondents have close contacts (friends, family members, or co-workers) in those out-parties.

Figure 3 underscores a strong association between exposure to supporters of other parties and affective polarization: those who do not have close contacts among outgroup parties exhibit more than twice as high levels of affective polarization compared to those who report having many close contacts in the outparties. While this finding is generally consistent with contact potentially playing an important role for affective polarization, the causality and direction of this relationship is not clear and motivates this study.

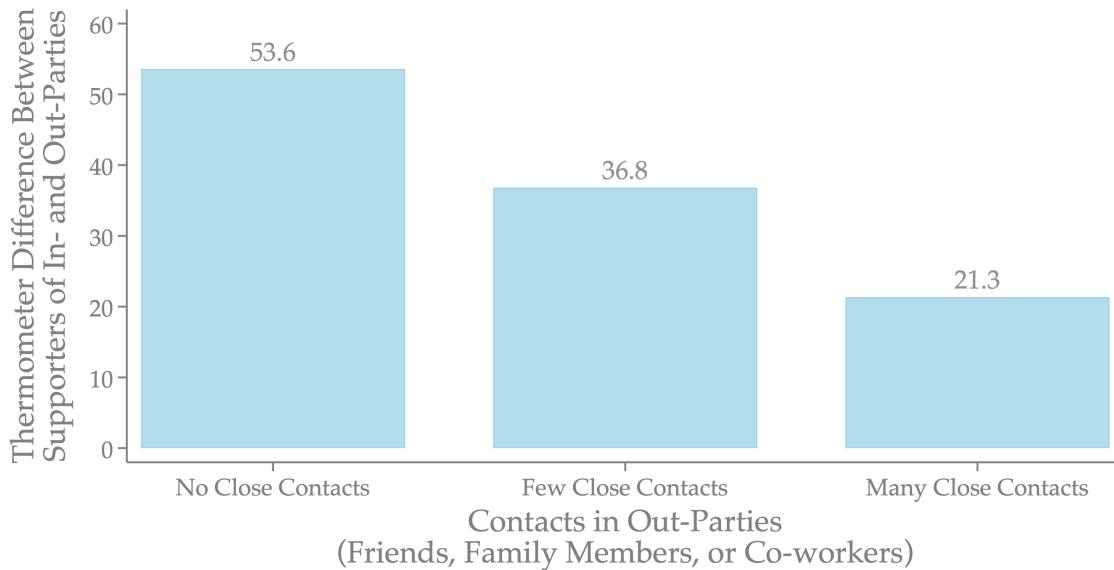


FIGURE 3: Affective Polarization and Contact

*Notes:* This figure plots a measure of affective polarization against the number of outgroup parties in which respondents report to have a close contact. The affective polarization measure is the difference in feeling thermometer ratings between the respondent's in-party (ranked first) and the average of the out-parties (parties that the respondent ranks second to sixth out of the six parties represented in the German parliament). Close contacts are defined as family members, friends, or colleagues. The sample consists of about 4,400 participants of Germany Talks who completed our baseline survey in 2021.

To assess to what extent contact with a political opponent has a causal impact on affective polarization, we study the exposure to political opponents induced by GT. In 2021, GT matched more than 14,000

participants for an unguided virtual conversation. Table 3 presents summary statistics based on respondents' self-reports of the duration and topics of the conversations. First, we find that conversations have significant lengths: 88% of respondents state that their conversation lasted more than one hour and one in three conversations (32%) lasted more than two hours. Second, we find that conversations are not limited to political topics: Virtually all respondents report that the conversations covered both political and personal topics, such as respondents' family and friends, jobs, or hobbies.

TABLE 3: Conversation Descriptive Statistics

	Conversation Duration			Conversation Topics	
	< 1 Hour (1)	1-2 Hours (2)	> 2 Hours (3)	Political Topics (4)	Personal Topics (5)
Mean	0.12 (0.33)	0.56 (0.50)	0.32 (0.47)	0.98 (0.12)	0.99 (0.10)
Observations	540	540	540	533	532

Notes. This table reports means for self-reported conversation duration and topics. We code the following conversation topics as personal: hobbies and interests, family and social environment, job, and current life situation. The sample consists of survey respondents who report to have had a conversation at the time of taking the endline survey. Standard deviations are in parentheses.

## 4.2 Effects on Affective Polarization

### A. Dictator and Trust Game

Figure 4 presents average amounts shared in the dictator game by treatment status and party affiliation of the other player. Recall that each respondent is randomized to play either one round of the dictator game or one round of the trust game in the endline survey. In both games, respondents were asked to allocate a budget of 100€ between themselves and another player (Player 2) whose party affiliation and demographic characteristics we varied at random.<sup>25</sup>

The three bars on the left of Figure 4 correspond to amounts shared by respondents who are scheduled but have not yet had a meeting through Germany Talks (our control group), while the three bars on the right show values for respondents who already had a conversation (our treatment group). Within treatment status, we group the amounts shared by the randomized party affiliation of Player 2 into top parties (Player 2 leans towards a party that the respondent ranked first or second), center parties (ranked third or forth), and bottom parties (ranked fifth or sixth).<sup>26</sup>

In the control group, we observe clear evidence for discrimination based on the party affiliation of Player 2 in the dictator game. Respondents who were randomly matched with another player who supports one of the respondent's top-ranked parties on average share 53€, compared to 51€ for parties they rank third or fourth. This difference is substantially larger relative to respondents matched with someone supporting a party they ranked fifth or sixth (41€). The differences between the amount given to the top-

<sup>25</sup>Respondents observe Player 2's party affiliation (randomly drawn from the three most common parties in our sample), age group (randomly drawn from three groups), location (randomly drawn from East or West Germany), and gender (fixed as male). See details in Section 3.3

<sup>26</sup>For instance, if a respondent ranked the Green Party first and was randomly assigned to a Player 2 who also supported the Green Party, the amount given would be captured in the first (or fourth) bar in the plot corresponding to the amount given to top-ranked parties.

ranked parties and the bottom-ranked parties are highly statistically significant with a p-value of 0.001 level, while the difference between the center parties and the bottom is significant at the 5% level.

This pattern is summarized by the gray line, indicating a negative slope of -5.8 (standard error = 1.95): those randomly matched with a player in the dictator game who supports a party the participant ranks favorably share substantially higher amounts compared to those matched with another player whose party is ranked less favorably.

In contrast, the corresponding series for the treatment group displays a different picture: while participants continue to share somewhat higher amounts when randomly assigned to another player from a top-ranked party, the differences are much smaller. On average, individuals give 50€ to counterparts supporting their top-ranked parties, 46€ to those supporting their center-ranked parties, and 45€ when paired with someone supporting one of the least favorably ranked parties. While the difference between the top- and bottom-ranked parties is significant at the 10% level, we fail to reject a t-test between the center- and bottom-ranked parties.

In summary, the relationship between amounts shared and party ranking is muted noticeably relative to the control group. This finding is confirmed by a change in the slope to a value of -3.13, which is about a 45% reduction, compared to the control group.

In Appendix Figures A.4 and A.5, we present analogous analyses from the trust game and both games pooled. The results portray a qualitatively similar picture: amounts shared vary less with the party affiliation of player 2 for respondents in the treatment group than they do for individuals in the control group, although we cannot reject a test for the equality of the slopes.

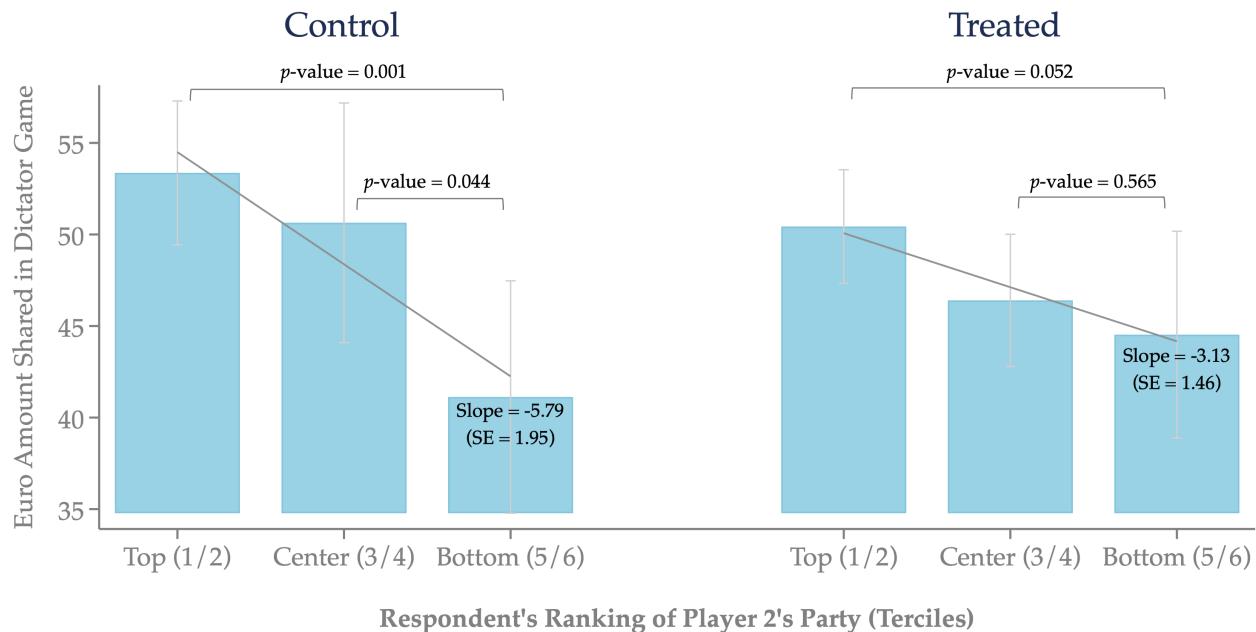


FIGURE 4: Treatment Effects on Affective Polarization: Dictator Game

*Notes:* This figure plots the Euro amounts shared in the dictator game among control (left panel) and treated subjects (right panel), separately by tertiles of the respondent's ranking of the party of the randomly assigned other player. Top refers to parties the respondent ranks first or second, Center to parties the respondent ranks third or fourth, and Bottom to parties the respondent ranks fifth or sixth. Confidence Intervals are at the 95% level.

To assess the effects of the treatment more formally, Table 4 reports estimates from regression models

that pool observations from both the treatment and control group. We estimate

$$Y_i = \alpha + \sum_{p=1}^j \beta_p I(\text{RankPartyPlayer2} = p) + \sum_{p=1}^j \delta_p I(\text{RankPartyPlayer2} = p) \times \text{Treated}_i + \tau \text{Treated}_i + \mathbf{X}_i + \epsilon_i \quad (1)$$

where  $Y_i$  represents the standardized amount shared in the respective game. The set of dummies  $I$  indicates whether the respondent was randomly assigned to a Player 2 who is affiliated with a party that is on rank  $p$  of the respondent's ranking of the six parties represented in the German parliament. The omitted category is made up of respondents who were randomly matched with a Player 2 who is affiliated with a party from the bottom tercile of the respondent's party ranking (fifth or sixth).<sup>27</sup>  $\text{Treated}_i$  is an indicator taking on the value 1 if a given individual is in the treatment group, and 0 otherwise. Finally,  $\mathbf{X}_i$  a vector of individual-level controls including demographics, baseline interest in contact, baseline survey measures of affective polarization, and demographic characteristics of Player 2.<sup>28</sup>

In this model, we're interested in the coefficient on the interaction term  $\delta_p$ , capturing changes in discrimination towards the ingroup as a result of the treatment. In the models reported in Panel A of Table 4, we further reduce the dimensionality by grouping the parties in a way analogous to Figure 4: the first two parties are referred to as top-ranked parties, the third and fourth-ranked parties are considered center parties, and the remaining two parties are bottom parties (the reference category).

Columns 1 and 2 of Table 4 show results from models that combine observations from both the dictator and trust game, first before and then after the inclusion of baseline controls. To combine the results from the two games we standardize all amounts given within game before running the regressions to account for the fact that the distributions differ substantially by game type.

The first row of Table 4 reports estimates for  $\tau$ , indicating that those in the treatment group did not share systematically higher (or lower) amounts with Player 2 than those in the control group: all coefficient estimates are statistically indistinguishable from zero. The second and third row report coefficients for  $\beta$ , a measure of ingroup bias. The estimates indicate that respondents matched to a Player 2 supporting a top-ranked (center-ranked) party share a 0.49 (0.40) SD higher amount on average.

Rows four and five report estimates for  $\delta$ , our main coefficient of interest. The estimates suggest that discrimination is reduced significantly among treated respondents, as the difference in the amount shared with top or center-ranked parties is reduced by -0.26 SD and -0.36 SD, respectively, compared with the omitted group of bottom-ranked parties. This corresponds to a reduction in affective polarization levels of about two thirds relative to the control group. This result is robust to the inclusion of a wide set of demographics and baseline polarization controls, reported in column 2. Columns 3 to 6 report results separately for the dictator and trust game, demonstrating qualitatively similar results across both games.

It is important to note that these coefficient estimates are larger than the coefficient for the treatment indicator  $\tau$ , reported in the first row. This suggests that conversations led to a reduction in discrimination against supporters of other parties rather than increased levels of altruism or general generosity.

In Panel B, we present results from an alternative specification, where we group the top- and center-ranked party in a single dummy as ingroup. We find that the treatment reduces ingroup bias by 0.3 SD, relative to a base of 0.44 SD. In contrast, the estimates presented in Panel C indicate that we do not find an effect of the conversations on discrimination between the top-ranked and all other parties, consistent with the patterns shown in Panel A indicating that decreases in ingroup favoritism are largest for center-ranked

---

<sup>27</sup>About 5% of participants are matched with a player from a party they rank sixth. We thus pool parties ranked fifth and sixth to increase power.

<sup>28</sup>We control for the respondent's age, gender, immigrant status, income, education, whether they live in East or West Germany, their ranking of the six major parties, the standardized thermometer feeling towards members of all six parties, positioning on a left-right scale, the number of baseline contacts in other parties, frequency of conversations about politics, interest in conversations, voting intentions, and a survey measure of general trust.

TABLE 4: Treatment Effects on Affective Polarization: Dictator and Trust Game

	Games Pooled		Dictator Game		Trust Game	
	No Baseline Controls (1)	Baseline Controls (2)	No Baseline Controls (3)	Baseline Controls (4)	No Baseline Controls (5)	Baseline Controls (6)
<i>Panel A: Parties Ranked at Top and Center Split</i>						
Treated ( $\tau$ )	0.20 (0.14)	0.19 (0.14)	0.13 (0.19)	0.15 (0.19)	0.26 (0.22)	0.29 (0.22)
<i>Ingroup Bias (<math>\beta</math>)</i>						
Top Parties	0.48*** (0.13)	0.50*** (0.13)	0.54*** (0.18)	0.51*** (0.18)	0.39* (0.21)	0.47** (0.21)
Center Parties	0.39*** (0.14)	0.39*** (0.14)	0.41** (0.21)	0.35* (0.20)	0.37* (0.21)	0.43** (0.20)
<i>Treated X Ingroup Bias (<math>\delta</math>)</i>						
Top Parties	-0.26 (0.17)	-0.26 (0.16)	-0.28 (0.22)	-0.30 (0.22)	-0.24 (0.26)	-0.24 (0.25)
Center Parties	-0.36** (0.18)	-0.36** (0.17)	-0.32 (0.25)	-0.30 (0.24)	-0.39 (0.26)	-0.49* (0.26)
Baseline controls		Yes		Yes		Yes
Game Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	892	892	460	460	432	432
<i>Panel B: Parties Ranked at Top and Center Pooled</i>						
Treated ( $\tau$ )	0.20 (0.14)	0.19 (0.14)	0.13 (0.19)	0.15 (0.19)	0.26 (0.22)	0.29 (0.22)
Ingroup Bias ( $\beta$ )	0.43*** (0.13)	0.44*** (0.12)	0.47*** (0.17)	0.42** (0.17)	0.38** (0.20)	0.45** (0.19)
Treated X Ingroup Bias ( $\delta$ )	-0.30* (0.16)	-0.30* (0.15)	-0.29 (0.21)	-0.29 (0.21)	-0.31 (0.24)	-0.35 (0.24)
Baseline controls		Yes		Yes		Yes
Game Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	892	892	460	460	432	432
<i>Panel C: Parties Ranked at Top Pooled</i>						
Treated ( $\tau$ )	-0.03 (0.09)	-0.04 (0.09)	-0.05 (0.13)	-0.02 (0.13)	-0.02 (0.12)	-0.05 (0.12)
Ingroup Bias ( $\beta$ )	0.21** (0.10)	0.24** (0.10)	0.28** (0.14)	0.29* (0.15)	0.12 (0.14)	0.15 (0.14)
Treated X Ingroup Bias ( $\delta$ )	-0.02 (0.12)	-0.02 (0.12)	-0.09 (0.17)	-0.12 (0.17)	0.04 (0.18)	0.11 (0.18)
Baseline controls		Yes		Yes		Yes
Game Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	892	892	460	460	432	432

Notes. This table reports results from regressions of the standardized amount shared in the game indicated in the header on dummies indicating whether the randomized player 2 belongs to the respondents party ingroup (ranked in top or medium tercile), treatment status, and the interaction of the two. The omitted category are respondents randomized to a player 2 who is affiliated with a party the respondent ranks 5th or 6th (the bottom tercile). Panel A reports the results of models with separate estimates for parties ranked 1st and 2nd (Top Parties) and parties ranked 3rd or 4th (Center Parties). Panel B reports the results of regressions pooling the Top and Center Parties as ingroup. Panel C reports the results of regressions pooling the Top Parties as ingroup. The sample is limited to participants who met their partner (treated) or already scheduled a meeting with their partner (control). All regressions include controls for the other randomized characteristics of player 2 and baseline party rankings. Columns (2), (4), and (6) add demographic controls and baseline affective polarization controls. Robust standard errors in parentheses.

parties.

To address concerns that our results may not be representative of a more general population given

the selected nature of our sample, in Appendix Table A.5, we show that the above results are robust to re-weighting our sample to match the German population based on several observable characteristics.<sup>29</sup>

Taken together, the above results suggest that conversations as part of GT led to a significant reduction in discrimination against supporters of other parties in incentivized economic interactions. Under our identifying assumption of exogenous timing of treatment conditional on having scheduled a meeting with the assigned partner (discussed in detail in Section 3.4), we can interpret these findings as causal.

### *B. Survey-based Measures of Affective Polarization*

In order to assess the robustness of our findings and compare our results to similar interventions discussed in the literature, we next study effects on affective polarization using feeling thermometer questions that are commonly used in the political science literature (Iyengar et al. 2012).

In both the baseline and endline survey, respondents were asked to rate their feelings for supporters of each of the six major parties on a scale ranging from 0 (very cold) to 100 (very warm). Based on the responses to these questions, we construct four within-respondent measures of affective polarization to capture effects on several parts of the distribution: first, the difference between the thermometer rating of supporters of the party the respondent ranks first and the rating of supporters of the party the respondent ranks last. Second, the difference between supporters of the party ranked first and the average rating of all other parties represented in our sample. Finally, the difference between the top-ranked parties (first or second) and the bottom-ranked parties (fifth and sixth), as well as between the top- and center-ranked parties and the bottom-ranked parties.

To estimate treatment effects on these measures, we regress the respective outcome on an indicator for treatment status as well as the same set of control variables as in equation 1:

$$Y_i = \alpha + \tau Treated_i + \delta + \mathbf{X}_i + \epsilon_i \quad (2)$$

where the coefficient of interest is  $\tau$ .

Figure 5 presents regression results assessing the effects of conversations on both relative and absolute changes in feeling thermometer ratings for supporters of the six main parties. The dark blue squares correspond to specifications without baseline controls while our preferred estimates in the light blue diamonds include controls for demographic characteristics and the baseline levels of affective polarization, thus representing changes relative to the baseline level of the outcomes.

The second row of Figure 5 indicates that the difference in the thermometer rating of supporters of the top-ranked parties and supporters of the center- and bottom-ranked parties is reduced by about 0.23 SD for those who had a conversation as part of GT. The treatment effect is only slightly affected by the inclusion of baseline control variables, providing support for our identification strategy.

Similarly, the third row indicates that the difference between the top-ranked parties and the bottom-ranked parties is reduced by 0.19 SD. We find reductions of somewhat smaller magnitudes when comparing the effect on differences in the ratings for supporters of parties ranked first versus those ranked sixth or all other parties, respectively. The average effect across these four measures (reported in the first row) is -0.15 SD.

The bottom panel of 5 plots treatment effects on the standardized absolute ratings of party supporters by the respondent's ranking of that party. This exercise highlights that the treatment effects on affective polarization are driven by both a reduction in the rating of the party ranked favorably (-0.11 SD for the

---

<sup>29</sup>We re-weight the sample to match the German adult population along based on age bins, gender and migration status (interacted). We repeat the same exercise this time re-weighting the sample along age by gender by education status bins. Finally, we re-weight the sample to account for differences in voting behavior. The choice of these dimensions and their interaction stems from the results suggesting selection into the sample based on these characteristics discussed in Table 1 and data availability in terms of the distribution of the German population along observable characteristics.

first-ranked party and slightly smaller effects for second- and third-ranked parties) and increases in the ratings of the parties ranked fifth (0.18 SD) and sixth (0.08 SD).

This finding implies that conversations led to a compression of the distribution to the center, rather than just increased ratings of outgroup members. This conclusion is confirmed by a coefficient estimate for the average thermometer rating that is close to zero and is also consistent with the results shown above from the incentivized games.

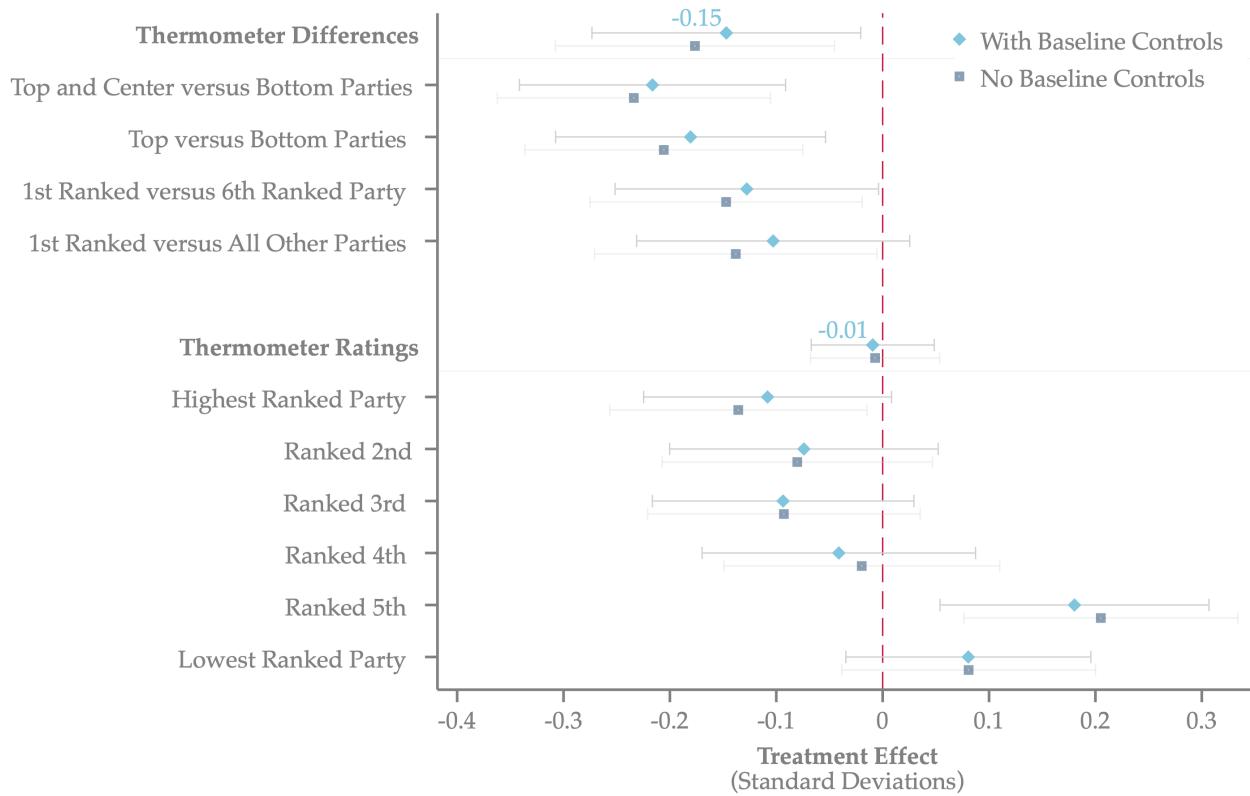


FIGURE 5: Treatment Effects on Thermometer Ratings of Party Supporters

*Notes:* This figure plots results from regression models, regressing differences in thermometer levels between parties on treatment status. The outcome measure is computed by standardizing the difference in the mean thermometer level (ranging from 0 to 100) across the parties indicated. The estimates for the overall thermometer differences and ratings are unweighted averages of the standardized measures below. In the control group, we include respondents who have scheduled a meeting but have not yet met a partner. The treatment group is comprised of respondents who have already met their partner. Coefficient estimates shown in dark blue squares include no controls while those in the light blue diamonds control for baseline demographics and baseline survey affective polarization measures. Confidence Intervals are at the 95% level.

In addition to affective polarization measures based on party affiliation, we also construct a measure of attitudes toward supporters of specific policy positions. Using the same thermometer scale, respondents were asked to rate their feelings towards a supporter of a given policy position. We included a supporting and an opposing statement for five different policy domains and take the difference between the two in each domain as a measure of individual affective polarization in the respective domain.

Figure A.6 presents regression results assessing the effects of conversations as part of GT on these issue-based measures of affective polarization, following the estimation in equation 2. As in Figure 5,

we present estimates from specifications with and without baseline controls. Aggregating the five policy domains covered, the estimates indicate that the conversations led to reductions in issue-based affective polarization of -0.12 SD. Disaggregating this finding into the five policy domains, we find that the average effect is predominantly driven by stronger effects on animosity towards supporters of other positions related to gender (-0.2 SD) and the free economy (-0.12 SD). In contrast, while the coefficient estimates still point towards reductions in affective polarization for positions towards climate change, Covid-19, and migration, these reductions are smaller in magnitude and not statistically significant.

### 4.3 Effects on Beliefs, Voting Intentions, and Ideological Polarization

#### *A. Beliefs and Voting Intentions*

Having observed economically meaningful and statistically significant effects for various measures of affective polarization, we next assess the effects of conversations on political opinions and ideological polarization. We begin by studying the treatment effects of the GT conversations on individuals' beliefs about polarization, their voting intentions, and party preference using the same empirical approach as in Figures 5 and A.6.

The top panel of Figure 6 shows the effect of conversations on two separate beliefs related to polarization.<sup>30</sup> First, we do not find an effect of conversations on respondents' beliefs whether Germany is politically divided.<sup>31</sup> Second, respondents in the treatment group are 0.12 SD more likely to agree that it's important to tolerate other positions to overcome the division of German society. Averaging over these two outcomes, we find a statistically significant treatment effect of 0.11 SD.<sup>32</sup>

The bottom panel of Figure 6 presents treatment effects on various measures of respondents' self-reported voting behavior and party support. In contrast to the previous results, we do not find any effects on these outcomes: whether GT participants already had a conversation or not does not change their intention to vote in the upcoming general election, their party ranking relative to the baseline, or which party they are going to vote for. These results are not only statistically insignificant but also small in magnitude.

#### *B. Ideological Polarization*

We next turn to studying the effects of conversations on ideological polarization. Our main measure of ideological polarization is the extremity of respondents' positions in six policy domains.<sup>33</sup> In the top row of Figure 6, we present regression results based on an index of ideological polarization that is constructed by averaging over the policy domain-specific measures of ideological polarization. We again employ the same estimation approach as in the previous figures. Our results indicate no significant effects of the conversations on ideological polarization. The coefficient on the treatment effect on the ideological polarization index is small (-0.01) and statistically insignificant.

Disaggregating this null-result into the six policy domains, we find inconsistent and insignificant results: while some coefficient estimates are negative (e.g. for banning domestic flights or rent control),

---

<sup>30</sup>Both questions were elicited on a five-point Likert scale and then standardized.

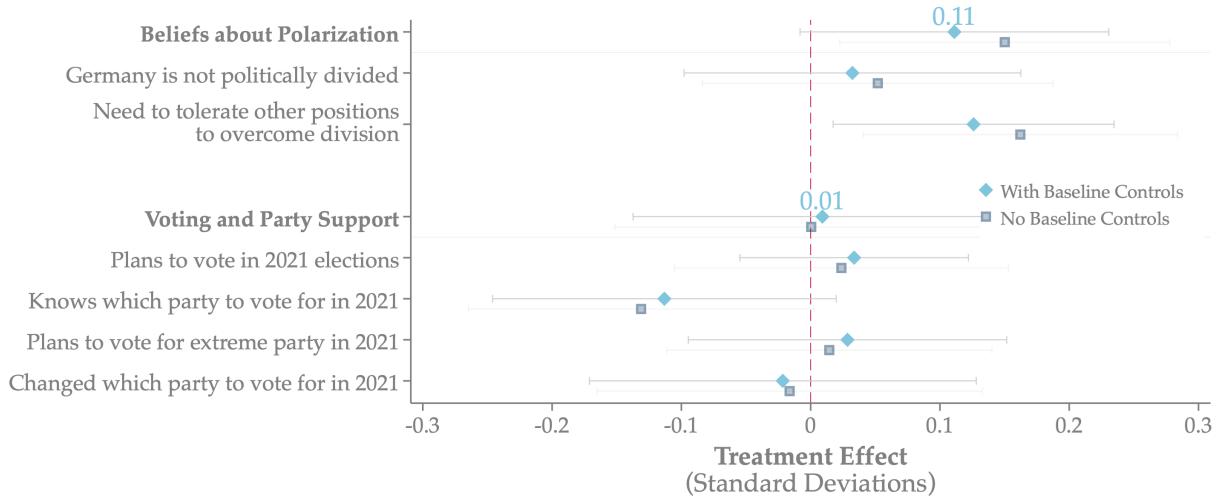
<sup>31</sup>The survey asked respondents whether they agree that Germany is politically divided. We reversed the order of the responses in the analysis for easier interpretation.

<sup>32</sup>Note that this value is not equal to the simple average as we re-standardize the index for ease of interpretation.

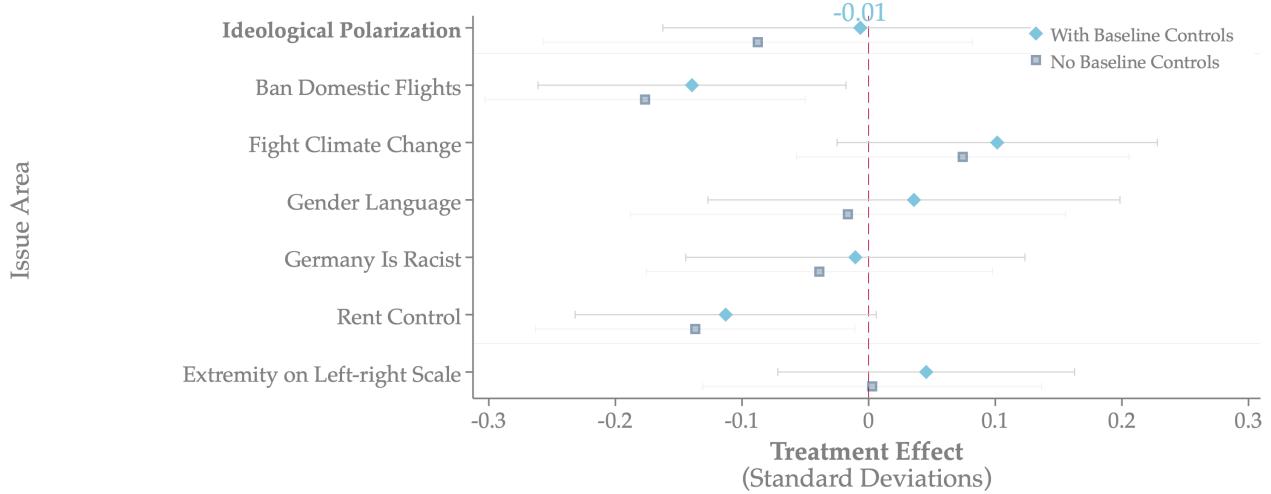
<sup>33</sup>We follow a similar approach to Allcott et al. (2020). For each issue area, we first standardize the responses to a five-point Likert question, then recenter, such that the average of the three left-leaning parties is negative, the average of the three right-leaning parties is positive, and the two averages sum up to 0. Finally, we multiply the average for left-leaning parties by -1, such that the resulting measure captures the alignment with the average position of the parties on the aligned side of the political spectrum.

FIGURE 6: Treatment Effects on Beliefs, Voting, and Ideological Polarization

(A) Beliefs, Voting, and Party Support



(B) Ideological Polarization Survey Outcomes



*Notes:* This figure plots results from regression models, regressing the standardized outcome indicated on the left on treatment status. Panel (A) focuses on beliefs about polarization, voting intentions, and party support. The beliefs about polarization were elicited using five-point Likert questions. Panel (B) shows results for survey-based measures of ideological polarization following a similar approach to (Allcott et al. 2020). For each issue area, we first standardize the responses to a five-point Likert question, then recenter, such that the average of the three left-leaning parties is negative, the average of the three right-leaning parties is positive, and the two averages sum up to 0. Finally, we multiply the average for left-leaning parties by -1, such that the resulting measure captures the alignment with the average position of the parties on the aligned side of the political spectrum. "Ideological Polarization" is an unweighted standardized average of the standardized measures below. In both Panel (A) and (B) the control group consists of respondents who have scheduled a meeting but have not yet met a partner. The treatment group are respondents who have already met their partner. Coefficient estimates shown in dark blue squares do not include controls while those in the light blue diamonds control for baseline demographics and baseline levels of the outcome measure.

Confidence Intervals are at the 95% level.

others have a positive sign. This result is confirmed by a small and insignificant effect on the extremity of the self-reported positioning on a left-right scale.

Finally, in addition to the survey-based measures reported in Figure 6, we also elicited respondents' political views in an incentivized decision. All respondents were asked if they wish to authorize a 500€ donation to Fridays-for-Future, a youth-led movement demanding drastic political changes to fight climate change.<sup>34</sup> In Figure A.7, we report the likelihood that survey respondents authorize the donation by treatment status and by the ranking of the Green Party in the baseline survey.<sup>35</sup>

The left three bars of Figure A.7 show that in the control group, donation decisions strongly correlate with support for the Green party. Among respondents who rank the Green party first or second, 86.1% decided to authorize the donation, compared to just 21.8% (3.0%) for those ranking the Green Party third or fourth (fifth or sixth). The right set of bars demonstrates that the treatment led to no change in respondents' propensity to authorize the donation to FFF: the proportion of those supporting a donation to FFF in all three groups in the control group is virtually identical to the corresponding groups in the treatment group.

Moreover, for both the treatment and control group, we find an identical slope of -41.5 corresponding to the extent to which respondents' willingness to donate drops as the ranking of the Green Party moves from top to center and then bottom. In corresponding regression results (not shown), we find that this null-result remains unchanged when we include the same rich set of control variables as in the regressions above.

## 5 Mechanisms and Heterogeneity

While the setting we study does not allow for a causal test of the mechanisms driving our result, this section presents descriptive evidence on the conversation topics in Section 5.1 and conversation outcomes in Section 5.1. Moreover, to assess the external relevance of our findings, we present heterogeneity results in Section 5.3.

### 5.1 Conversation Topics

The organizers of GT do not impose any structure on the conversations. Besides information about the assigned partner, GT does not provide a script or makes suggestions about how to approach the conversations. The topics of conversation can therefore be interpreted as endogenous choices by the participants, conditional on the information available. A natural first question is whether conversations resembled purely political debates or also focused on non-political topics. We elicited respondents' self-reports regarding the conversation topics covered and find that nearly all participants (> 98%) report talking about both political and personal topics, as shown in columns 4 and 5 of Table 3.

Panel A of Table 5 provides statistics on the extent to which specific political topics were covered and how match characteristics shape the topics discussed. Columns 1 to 6 indicate that there is significant variation across the political domains covered. With around 90% (50%) of respondents reporting that they talked about environmental issues at least some time (extensively), this topic appears to be the most commonly discussed political issue. Foreign policy and health issues related to COVID-19 are the topics discussed least frequently: only around 60% of respondents indicate it was covered at least some time in

<sup>34</sup>Respondents were informed that the researchers will pay for the donation if a randomly drawn respondent decided to authorize the donation. See Section 3.3 for more details.

<sup>35</sup>We focus on the Green Party because its policy positions are most closely linked to efforts to fight climate change and it is very well represented among survey respondents, indicating that climate change is likely a central issue of debate in many GT conversations. This is confirmed by more than 90% of respondents reporting to have discussed issues related to the environment in their conversation as shown in 5.

their conversation.

TABLE 5: Conversation Topics

	Racism and Migration (1)	Foreign Policy (2)	Inequality (3)	Covid-19 Health (4)	Covid-19 Social/Econ (5)	Environment (6)
Covered at least some time	0.84 (0.36)	0.61 (0.49)	0.77 (0.42)	0.63 (0.48)	0.86 (0.34)	0.92 (0.28)
Covered Extensively	0.51 (0.50)	0.24 (0.42)	0.31 (0.46)	0.20 (0.40)	0.45 (0.50)	0.54 (0.50)
Observations	534	534	534	534	534	534
<i>Panel B: Effects of Match Characteristics on Probability Topic Covered at Least Some Time</i>						
<i>Policy Disagreement</i>						
Germany is racist	0.41*** (0.07)	0.00 (0.06)	0.11* (0.06)	-0.00 (0.06)	-0.09 (0.07)	0.07 (0.07)
Sanctions for russia	0.02 (0.07)	0.16** (0.06)	0.01 (0.07)	0.07 (0.06)	-0.02 (0.07)	0.06 (0.07)
Support for rural areas	0.10 (0.07)	0.04 (0.06)	0.14** (0.06)	-0.05 (0.05)	0.04 (0.06)	0.06 (0.07)
Covid-19 measures	-0.13* (0.07)	-0.04 (0.06)	-0.09 (0.07)	0.14** (0.06)	0.02 (0.07)	-0.08 (0.07)
Rent control	0.14 (0.10)	-0.08 (0.09)	0.07 (0.10)	0.08 (0.09)	-0.02 (0.10)	0.13 (0.11)
Ban domestic flights	0.02 (0.09)	0.05 (0.07)	-0.10 (0.08)	-0.10 (0.07)	-0.07 (0.08)	0.07 (0.09)
<i>Partner Demographics</i>						
Same Gender	0.03 (0.05)	0.12*** (0.04)	-0.00 (0.05)	-0.10** (0.04)	-0.10* (0.05)	0.02 (0.05)
Age Difference < 10	-0.02 (0.06)	-0.00 (0.05)	-0.05 (0.05)	0.08* (0.05)	0.03 (0.05)	-0.11* (0.06)
Same State	0.02 (0.08)	0.03 (0.07)	-0.04 (0.08)	0.01 (0.07)	0.07 (0.08)	-0.02 (0.08)
Outcome Mean	0	0	0	0	0	0
Outcome SD	1	1	1	1	1	1
Policy Question Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	361	361	361	361	361	361

Notes. This table reports self-reported conversation topics (panel A) and results from regressions predicting self-reported conversation topics (panel B) among participants who met their partner (treated). In panel B, we regress a dummy indicating whether the conversation topic indicated in the header was at least covered some time on a set of dummies indicating whether the two assigned partner disagreed on a policy domain or shared demographic characteristic. Respondents' self-reported whether they covered a topic on the following scale: not at all, some time, extensively, most of the time. All regressions include controls for the sign-up responses of the participant. Policy disagreement refers to dummies indicating whether the participant provided a different binary response to the policy question than their partner. Standard deviation (panel A) and robust standard errors (panel B) in parentheses.

To better understand the sources of the variation in conversation topics across participant pairs, Panel B presents results from regressions predicting which conversation topics were covered by respondents. For this exercise, we draw on admin data from GT, which allows us to observe characteristics of the assigned partner for about two thirds of our sample.<sup>36</sup>

<sup>36</sup>Recall that a subset of the sample could not be matched because they used a different email address in our survey than to

For each conversation topic, we regress a dummy indicating whether the topic was covered on a set of dummies indicating on which policy positions the respondent and their partner disagreed on.<sup>37</sup> Moreover, we include dummies indicating whether the assigned partner shares the same gender, is within ten years of age, and is from the same state as the respondent. Given that the matching algorithm takes only the responses to the sign-up policy questions as input, the domain on which participants disagree as well as the demographic characteristics are conditionally randomly assigned. We therefore control for all policy question responses of the respondents, so that the coefficients can be interpreted as the causal effect of the various features on the topics covered in the conversations.

Column 1 of Panel B shows the results for racism and migration. The estimate in the first row indicates that respondents disagreeing with their partner regarding the statement whether Germany is a racist country are 41 percentage points (pp) more likely to report racism and migration as a conversation topic. We observe a similar pattern in columns 2, 3, and 4: Similarly, the estimate in the third row of column 3 in Panel B indicates that respondents disagreeing on sanctions for Russia are 16pp more likely to cover foreign policy in their conversation. Disagreement on the question whether the German government should provide more support for rural areas leads to a 14pp higher probability of the topic of inequality being covered. Similarly, respondents who were assigned to a partner they disagreed with on any of the statements regarding measures to stop the spread of COVID-19 are 14pp more likely to report having talked about health issues related to COVID-19 with their partner. The estimates in columns 5 and 6 show that there are no strong predictors of whether the social and economic consequences of COVID-19 and the environment were covered as conversation topics. In Appendix Table A.4, we present analogous estimates for personal conversation topics.

Taken together, these findings suggest that the conversations cannot be characterized as a purely political debate but rather a combination of personal and political conversation topics. Moreover, we find that within the political domain, a wide range of topics is discussed and that respondents focus on domains they disagree on at baseline. In contrast to recent online experiments, such as Santoro and Broockman (2022) we find treatment effects in a setting that combines non-political with divisive political conversation topics.

## 5.2 Conversation Outcomes

In addition to the conversation topics, we elicited self-reports regarding several outcomes of the conversation from each respondent. While we cannot present heterogeneous treatment effects based on these outcomes since they are likely the result of endogenous factors, we present several descriptive statistics to provide some evidence on potential mechanisms driving the observed effects of the conversation.

First, we assess to what extent the observed effects may be driven simply by a warm glow effect from the meeting with a stranger. Figure A.8 compares average responses to the feeling thermometer questions regarding party supporters with the average feeling of respondents towards their GT conversation partner. The values for the conversation partner were elicited in the endline survey, while we use baseline survey measures for the ranking of party supporters.

Figure A.8 highlights that participants on average report very high feelings towards their GT conversation partner: the average value associated with the GT partner is 74.6 (light blue bar). These overwhelmingly positive feelings towards their partner are striking given the large political differences between partners that were explicitly targeted and communicated to partners by the GT initiative. Moreover, the average values are almost as high as the average rating for voters of respondents' most preferred party

---

register for GT. Another subset of the sample asked GT to delete their personal information before we obtained the admin data from GT.

<sup>37</sup>Respondents could state one of four options: not at all, some time, extensively, most of the time.

(77.9).<sup>38</sup>

This result is confirmed in responses to a survey question asking participants to what extent they liked the meeting with their partner. Column 5 of Table 6 shows that more than 90% of participants indicate that they strongly or very strongly liked their meeting. This seems to not be fully driven by the personal similarities of the assigned partners. Only 27% of respondents report having found strong personal similarities (such as hobbies, or a similar job) with their partner.

Next, we assess the role of several additional potential outcomes of the conversation. For instance, the conversation may have led participants to convince the other side of their own position, correct misperceptions about the political positions of the other side, or understand the arguments of the other side for the position they hold. To assess the role of these channels, columns 1, 2, and 3 of Table 6 present descriptive statistics for several self-reported conversation outcomes.

TABLE 6: Conversation Outcomes

	Political Outcomes			Personal Outcomes	
	Convinced Partner of Own Positions (1)	Found Overlap with Positions of Partner (2)	Understand Argument for Position (3)	Found Personal Similarities with Partner (4)	Liked Meeting with Partner (5)
Not at all or Hardly	0.40 (0.49)	0.08 (0.27)	0.04 (0.19)	0.30 (0.46)	0.02 (0.13)
Somewhat	0.54 (0.50)	0.35 (0.48)	0.23 (0.42)	0.44 (0.50)	0.06 (0.24)
Strongly or Very Strongly	0.06 (0.24)	0.57 (0.50)	0.73 (0.44)	0.27 (0.44)	0.92 (0.27)
Observations	582	582	582	582	582

Notes. This table reports self-reported conversation outcomes. Self-reported conversation outcomes were surveyed on a five-point scale: Not at all, hardly, somewhat, strongly, very strongly. The sample is limited to participants who met their partner (treated). All regressions include controls for the sign-up responses of the participant. Standard deviation in parentheses.

We find that only very few respondents (6%) believed that they strongly convinced their partner of their own political views. While substantially more respondents (54%) report that they somewhat convinced their partner, this indicates that convincing the other side may not be the primary driver of the observed effects. In contrast, the majority of 57% (92%) of individuals found that they had strong (at least some) overlap in political views with their partner. In light of the fact that partners were matched explicitly with the goal of maximizing the differences in their political views, the observed degree of overlap is striking. Finally, almost three out of four participants (73%) report that while they disagreed with the political position of their partner, they were able to understand the arguments or rationale of their partner for the positions they hold.

These findings point to three potential channels: first, participants may have updated regarding the extremity of the other side.<sup>39</sup> Second, respondents were able to identify at least some overlap in political positions with political opponents, thereby reducing animosity towards them. Finally, while still disagreeing on the policy position, respondents were able to better understand the rationale of the other side, thereby justifying the other side's policy views.

<sup>38</sup>It is important to note that an exact comparison of these values is difficult given the fact that the GT partner is a single individual who participants have met in real life while the supporters of the various parties are more hypothetical or abstract in nature.

<sup>39</sup>An alternative explanation is that matches were not as extreme as participants anticipated.

### 5.3 Heterogeneity

The participants of GT are selected based on one particular feature: they voluntarily registered for the event to be matched with a political opponent. This raises the question whether our findings are generalizable for a population that does not voluntarily seek contact. While we cannot measure treatment effects for the general population in our sample, we can use the existing variation in the degree of polarization and demand for exposure within our sample.

In Table 7, we present heterogeneity results to assess differences in treatment effects for three separate characteristics: high versus low baseline affective polarization, high versus low baseline interest in a conversation with a political opponent, and high versus low agreement that it's important to tolerate other positions to overcome polarization in the German society. For each of these characteristics, we distinguish between those with high and low levels by defining those with above-median values as 'high' and those with below-median values as 'low'.<sup>40</sup> Note that we measure all respondent characteristics at baseline, so that those characteristics cannot be affected by treatment status.

We run regressions analogous to those shown in Panel B of Table 4, each time splitting the sample into different subsamples depending on the baseline characteristic under study. As in Figures 4 we focus on  $\delta$ , i.e. the coefficient on the interaction of treatment and ingroup bias as the main coefficient of interest as it corresponds to changes in the amount shared in the respective game among those who had a conversation through Germany Talks. Recall that negative numbers indicate reductions in discrimination of players affiliated with out-parties (relative to in-parties) and positive numbers correspond to increases in discrimination.

The estimates in columns 1 and 2 of Table 7 suggest that the treatment effect highly depends on an individual's baseline level of affective polarization: we find a significant treatment effect among those with above median affective polarization levels ( $\delta = -0.60$ ), indicating that the conversation through Germany Talks led to a reduction in discrimination of 0.6 SDs.<sup>41</sup> In contrast, for those with below median levels of initial affective polarization, the estimated coefficient is very close to zero and insignificant. This finding indicates that the effect of the conversation is entirely concentrated among those with high initial levels of affective polarization.

We find similar differences in treatment effects for the other heterogeneity dimensions. Columns 3 and 4 highlight that the treatment effects are large and highly significant ( $\delta = -0.64$ ) for those with lower self-reported initial interest in conversations with political opponents and are again closer to zero and insignificant for those who report a higher interest in exposure. Similarly, columns 5 and 6 of Table 7 highlight again a high degree of heterogeneity in effect size as a function of respondents' agreement with the statement that it is important to tolerate other political positions to overcome polarization in Germany: treatment effects are large and statistically significant among those who are less likely to agree with that statement ( $\delta = -0.65$ ) and are close to zero and insignificant for respondents with above median responses.

While the three variables underlying these heterogeneity splits are clearly correlated, the three different splits actually result in different cuts of the data. At the bottom of Table 7, we report average levels of baseline affective polarization for each split. The baseline affective polarization levels vary substantially, suggesting that the patterns observed are not merely driven by subsetting the data in the same way.

Taken together, Table 7 shows that the treatment effects are concentrated among respondents who have high baseline levels of affective polarization, who are less interested in the conversations, and generally tolerate others' views less. We interpret these findings as suggestive that initiatives that bring people from different ends of the political spectrum together may be effective not just for those who very actively

<sup>40</sup>Note that the number of observations varies between the two groups even if use the median to assign individuals to different groups because of the discrete nature of the underlying variables.

<sup>41</sup>The underlying affective polarization variable is an index with mean 0 and a standard deviation of 1.

TABLE 7: Treatment Effects by Respondent Baseline Characteristics

	Affective Polarization		Interest in Conversation		Belief: Important to Tolerate other Positions	
	Low (1)	High (2)	Low (3)	High (4)	Low (5)	High (6)
Treated ( $\tau$ )	-0.10 (0.20)	0.44** (0.20)	0.39* (0.23)	0.02 (0.17)	0.38* (0.20)	-0.05 (0.19)
Ingroup Bias ( $\beta$ )	0.31* (0.17)	0.59*** (0.17)	0.58** (0.23)	0.30** (0.14)	0.68*** (0.18)	0.25 (0.16)
Treated X Ingroup ( $\delta$ )	-0.03 (0.22)	-0.60*** (0.22)	-0.64** (0.26)	-0.08 (0.19)	-0.65*** (0.22)	0.07 (0.22)
Mean Aff. Polarization	-0.77	0.90	0.12	0.00	0.19	-0.12
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	452	439	361	530	483	408

Notes. This table reports heterogeneous treatment effects by respondents' baseline characteristics. All estimates are from regressions of the standardized amount shared in the dictator and trust game (pooling both games) on dummies indicating whether the randomized player 2 belongs to the respondents party ingroup (ranked 1st to 4th), treatment status, and the interaction of the two. The omitted category are respondents randomized to a player 2 who is affiliated with a party the respondent ranks 5th or 6th. Columns 1 and 2 report results separately for the sample of respondents who had a baseline affective polarization level below (column 1) or above the median (column 2). Columns 3 and 4 split the sample at the median baseline interest in conversations measured by a five point Likert survey question. Columns 5 and 6 split the sample at the median response to a survey question asking about the importance of tolerating other positions to overcome polarization. Mean affective polarization reports the average of a baseline affective polarization index (with mean 0 and standard deviation of 1) by sample split. The sample is limited to participants who met their partner (treated) or already scheduled a meeting with their partner (control). All regressions include controls for the other randomized characteristics of player 2, baseline party rankings, demographic controls and baseline affective polarization

seek contact with members of the outgroup, but especially for those with lower initial interest in conversations with outgroup members. In related work, we address this question by recruiting representative samples that overcome the selection into treatment in our context.

## 6 Discussion

In this paper, we provide evidence that even a short conversation with a political opponent can significantly reduce animosity toward supporters of other political parties. In our sample of participants of the initiative GT that facilitated conversations between individuals with different political opinions, the conversations led to reductions in both incentivized and unincentivized proxies for affective polarization. In contrast to the reductions in affective polarization, we find little evidence that participants in our intervention adopted less polarized policy positions. We confirm this result in an incentivized donation decision. In line with this finding, we find that conversations also did not have an impact on voting behavior. These results indicate that exposure to political opponents increases tolerance of the other side but not support for opposing positions.

We also provide suggestive evidence on the mechanisms driving these effects. Participants covered both political and personal topics in their conversations, suggesting that the effect of the conversation is neither purely driven by an exchange of political arguments, nor limited to the personal domain. Moreover, within the political domain, participants focused on policy areas of disagreement in their conversations. This finding contrasts results from recent online experiments that found treatment effects for

conversations for non-political conversation scripts but not for conversations about policy disagreements (Santoro and Broockman 2022). Future work needs to assess why in some settings conversations lead to reductions in affective polarization but not in others.

Our findings suggest that the effect of our intervention is potentially driven by a combination of personal affect towards the conversation partner, updating regarding the policy positions of the other side, as well as an increased understanding of the rationale of political opponents for the positions they hold. Participants liked their conversation partner as much as supporters of their own party. At the same time, more than 90% of respondents stated that they identified overlap in the policy positions with their partner and that they were able to understand the rationale for their partner's policy positions. Our setting did not allow us to study these mechanisms using treatment variation. In ongoing work, we conduct field experiments in Brazil and the U.S. that randomly vary the characteristics of conversation partners and features of the conversation to study these mechanisms in a causal manner.

It is important to emphasize that in the setting we study, participants voluntarily decided to be exposed to a political opponent. This feature needs to be considered when assessing the external relevance of our findings. Within our sample, we find that respondents with higher initial levels of affective polarization and lower levels of interest in the conversations are driving the effect of the conversations. This raises the question whether the effects we find persist or increase when estimated in a sample that did not seek contact with the other side. This is the focus of our ongoing work, for which we recruit a sample that did not self-select into contact and incentivize treatment take-up.

Finally, the question remains if an intervention like GT can be replicated in wider settings. MCT has demonstrated the potential of scaling up its events by recruiting more than 200,000 participants in over 30 countries. Several other organizations have been founded with similar goals to connect partisans for an exchange. An area for future work is the question how such initiatives can be successfully scaled up to the broad public, for instance using large-scale social media campaigns. Moreover, future work needs to address how the findings from this context can be translated to other environments where contact could be facilitated, such as workplaces, schools, or campuses. Finally, we lack an understanding of the factors that drive partisans' demand for contact with the other side. We hope to answer these questions in ongoing work in collaboration with MCT.

## 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*.
- 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.
- Boxell, L., Gentzkow, M., and Shapiro, J. M. (2020). Cross-Country Trends in Affective Polarization. *Working Paper*.
- Boxell, L., Gentzkow, M., and Shapiro, J. M. (2022). Cross-Country Trends in Affective Polarization. *Review of Economics and Statistics*.
- Broockman, D. and Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282).
- Broockman, D. E., Kalla, J. L., and Westwood, S. J. (2022). Does Affective Polarization Undermine Democratic Norms or Accountability? Maybe Not. *American Journal of Political Science*, page ajps.12719.
- Bursztyn, L., Chaney, T., Hassan, T., and Rao, A. (2021). The immigrant next door: Exposure, prejudice, and altruism. *NBER Working Paper No. 28488*.
- Colonnelly, E., Neto, V. P., and Teso, E. (2022). Politics at Work. *Working Paper*.
- Druckman, J. N. and Levendusky, M. S. (2019). What Do We Measure When We Measure Affective Polarization? *Public Opinion Quarterly*, 83(1):114–122.
- Fershtman, C. and Gneezy, U. (2001). Discrimination in a Segmented Society: An Experimental Approach. *The Quarterly Journal of Economics*, 116(1):351–377.
- Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., and Druckman, J. N. (2020). Political sectarianism in America. *Science*, 370(6516):533–536.
- 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.

- Fishkin, J., Siu, A., Diamond, L., and Bradburn, N. (2021). Is Deliberation an Antidote to Extreme Partisan Polarization? Reflections on “America in One Room”. *American Political Science Review*, 115(4):1464–1481.
- Gidron, N., Adams, J., and Horne, W. (2020). *American Affective Polarization in Comparative Perspective*. Cambridge University Press, first edition.
- Gift, K. and Gift, T. (2015). Does Politics Influence Hiring? Evidence from a Randomized Experiment. *Political Behavior*, 37(3).
- Heuser, S. and Stötzer, L. S. (2022). The Effects of Face-To-Face Conversations on Polarization: Evidence from a Quasi-Experiment. *Working Paper*.
- Huber, G. A. and Malhotra, N. (2016). Political Homophily in Social Relationships: Evidence from Online Dating Behavior. *Journal of Politics*, 79(1).
- Iyengar, S., Konitzer, T., and Tedin, K. (2018). The Home as a Political Fortress: Family Agreement in an Era of Polarization. *Journal of Politics*, 80(4).
- 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.
- Iyengar, S., Sood, G., and Lelkes, Y. (2012). Affect, Not Ideology. *Public Opinion Quarterly*, 76(3):405–431.
- Iyengar, S. and Westwood, S. J. (2015). Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, 59(3):690–707.
- Jha, S. (2013). Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia. *American Political Science Review*, 107(4):806–832.
- 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.
- Martin, G. J. and Yurukoglu, A. (2017). Bias in Cable News: Persuasion and Polarization. *American Economic Review*, 107(9):40.
- McConnell, C., Margalit, Y., Malhotra, N., and Levendusky, M. (2018). The Economic Consequences of Partisanship in a Polarized Era. *American Journal of Political Science*, 62(1):5–18.
- Michelitch, K. (2015). Does Electoral Competition Exacerbate Interethnic or Interpartisan Economic Discrimination? Evidence from a Field Experiment in Market Price Bargaining. *American Political Science Review*, 109(1):43–61.
- 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).
- Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, 69:102199.

# A Appendix

## A.1 Details on Matching Process

Participants are matched based on a greedy algorithm that aims to maximize the distance in responses to the binary policy questions of assigned matches. This process is visualized in Figure A.1. The algorithm considers all participants who signed up since the last matching took place (typically within one week). The algorithm then generates random subsets of the sample for easier processing. Within a given match sample, it picks a random participant and picks a match that maximizes the level of disagreement, measured by the number of diverging responses on the binary sign-up questions. This generates some randomness in match assignment as highlighted in the example on the right side of the figure. Depending on which participant the algorithm picks first at random, it is random whether participant 1 is matched to participant 3 or 4. The algorithm then iterates over all participants and match samples. It leaves a subset of participants (5-10%) unmatched if it cannot find a partner with a distance in sign-up questions of at least two.

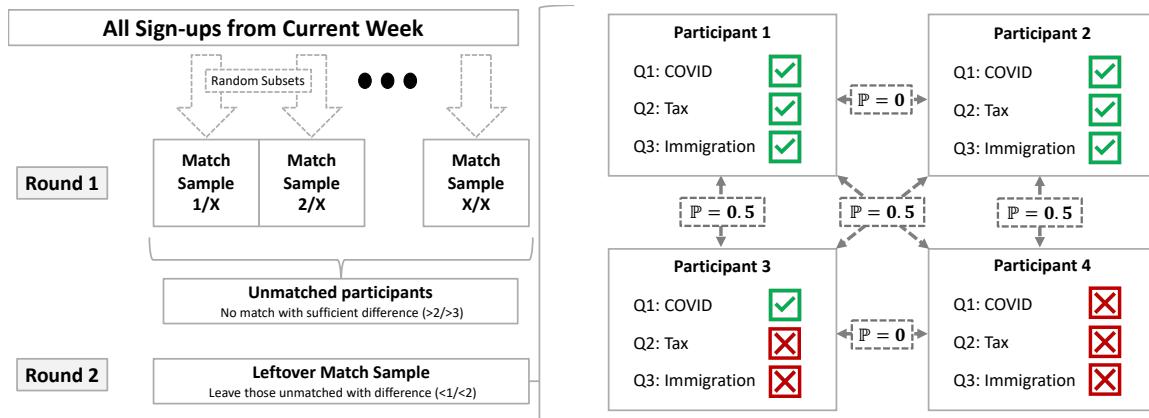


FIGURE A.1: Matching Procedure and Example Matching

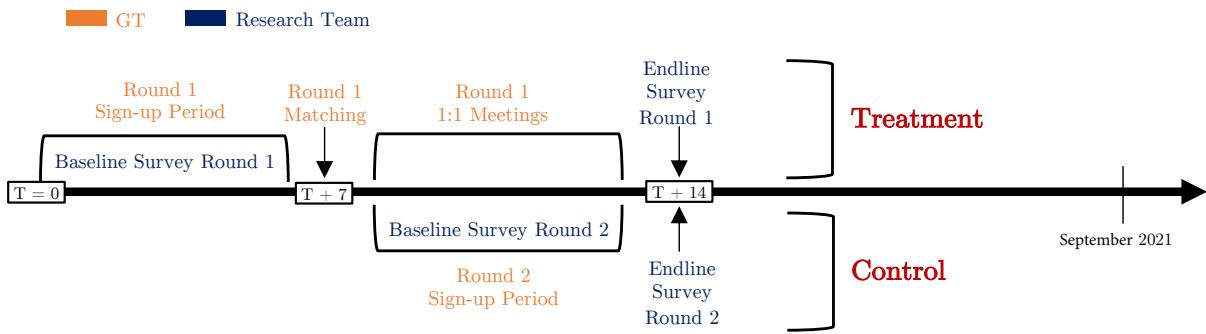


FIGURE A.2: Timing of Matching and Invites to Endline Survey

## A.2 Distribution of Sign-up Questions by Party Affiliation

TABLE A.1: Distribution of Sign-up Questions by Party Affiliation

Question	All (1)	Percentage of Respondents Answering "Yes"							N (8)
		Left-leaning Parties			Right-leaning Parties				
		The Left (LINKE) (2)	Greens (Grüne) (3)	Soc. Dem. (SPD) (4)	Conserv. (CDU / CSU) (5)	Liberals (FDP) (6)	AfD (7)		
Should children (>=12 years) be vaccinated?	63%	61%	69%	74%	66%	54%	19%	1419	
Should vaccinated people regain civic liberties?	74%	62%	74%	74%	81%	81%	48%	1279	
Is "COVID Emergency Brake" enough?	53%	47%	41%	50%	65%	66%	59%	1018	
Should there be a national vaccine mandate?	38%	40%	38%	52%	46%	30%	13%	494	
Should schools stay open even if infections go up?	72%	66%	62%	62%	93%	83%	100%	232	
Should "2G" COVID regulations be mandated everywhere?	65%	40%	67%	100%	60%	38%	50%	63	
Does Germany do too little for those living in rural areas?	31%	51%	37%	31%	17%	19%	35%	2828	
Should there be a national cap on rent?	51%	86%	69%	63%	25%	16%	36%	2828	
Does gender appropriate language help with inequality?	26%	34%	48%	26%	7%	4%	1%	1841	
Should flights within Germany be banned?	47%	73%	70%	48%	27%	16%	18%	2765	
Should gas be taxed more?	49%	58%	76%	46%	24%	28%	9%	1089	
Is Germany a racist country?	35%	60%	52%	38%	16%	13%	8%	2828	
Should Germany take in more people from Afghanistan?	65%	100%	90%	78%	42%	33%	18%	158	
Should the EU impose stricter sanctions on Russia?	52%	40%	68%	52%	47%	44%	17%	2828	
<u>Do the individual German states have too much power?</u>	39%	39%	45%	42%	39%	32%	26%	1793	

Notes. This table shows the distribution of responses to the sign-up questions. The sample consists of respondents to our baseline survey who could be matched to the admin data provided by GT. Column (1) presents the percentage of all respondents answering with yes. Columns (2) - Columns (7) show the percentage of respondents answering with yes by the party they rank first. Column (8) indicates the total number of respondents answering a given question.

### A.3 Attrition

TABLE A.2: Sample Attrition

	Baseline Survey Only (1)	Started Endline Survey (2)	Balance (3)	p-value (4)
<i>Panel A: Demographic Characteristics</i>				
Female	0.33	0.34	-0.00	0.88
Age	49.43	50.51	-1.08	0.06
College degree	0.58	0.66	-0.08	0.00
Immigrant	0.14	0.13	0.01	0.30
Income > 60,000	0.27	0.30	-0.04	0.03
Working	0.73	0.77	-0.04	0.01
East Germany	0.23	0.19	0.03	0.02
Married	0.46	0.51	-0.06	0.00
Very strong pol. interest	0.32	0.35	-0.02	0.19
General trust	0.45	0.44	0.02	0.34
Left-leaning	0.73	0.71	0.02	0.27
Voted center left in 2017	0.38	0.39	-0.01	0.41
Voted center right in 2017	0.29	0.31	-0.02	0.19
Voted far-left/right in 2017	0.16	0.15	0.02	0.21
Plan to vote in 2021	0.90	0.91	-0.01	0.17
<i>Panel B: Baseline Affective Polarization Levels</i>				
Party Thermometer Min	7.56	7.20	0.36	0.42
Party Thermometer Max	80.80	79.88	0.92	0.10
Party Thermometer Range	73.24	72.68	0.56	0.49
Ideological polarization	-0.01	-0.01	0.00	0.94
Observations	1083	3294	4377	

Notes. This table compares descriptive statistics for respondents who completed the baseline survey only (1) and those who completed both the baseline and started the endline survey (2). Column 3 reports differences between column 1 and 2, and column 4 p-values from a t-test.

## A.4 Dictator Game Details

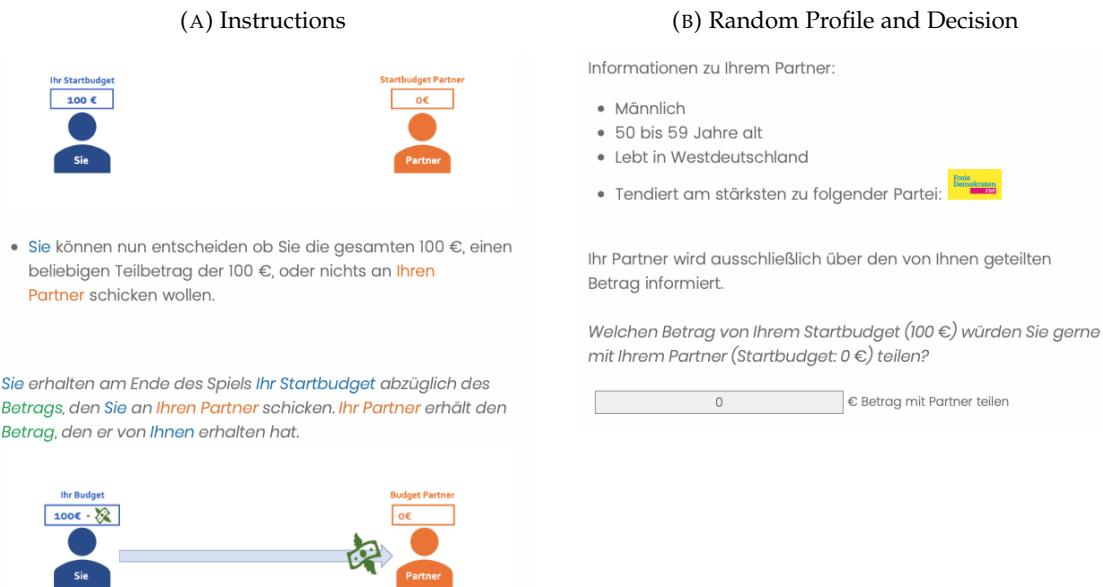


FIGURE A.3: Dictator Game Implementation

## A.5 Discrimination in Dictator and Trust Game in Control Group

TABLE A.3: Discrimination in Dictator and Trust Game in Control Group

	Dictator Game		Trust Game	
	Mean (1)	Regression Coefficient (2)	Mean (3)	Regression Coefficient (4)
Partner from Top Parties (1st/2nd)	0.41 (0.01)	7.85*** (1.29)	0.43 (0.01)	5.16*** (1.71)
Partner from Top or Center Parties (1st/2nd/3rd/4th)	0.77 (0.01)	10.14*** (1.49)	0.77 (0.01)	9.15*** (2.01)
Partner from Same Location (East/West Germany)	0.48 (0.01)	-2.37* (1.28)	0.51 (0.01)	1.42 (1.69)
Average Amount Shared		47.54		73.88
Observations	1225	1225	1264	1264

Notes. This table reports means and results from bivariate regressions of the amount shared in the dictator and trust game on characteristics of the second player in the respective game. The party affiliation (CDU/CSU, Greens, or FDP), the age group (30-39, 40-49, 50-59 years), and the location (East Germany, West Germany) were randomized. The sample is limited to participants who have not met a partner. Standard deviations in parentheses in columns 1 and 3, and standard errors in parentheses in columns 2 and 4.

## A.6 Predicting Personal Conversation Topics

TABLE A.4: Conversation Topics

	Family and Friends (1)	Job (2)	Hobbies (3)	Current Life Situation (4)
<i>Panel A: Descriptive Statistics</i>				
Covered at least some time	0.87 (0.34)	0.90 (0.31)	0.64 (0.48)	0.86 (0.35)
Covered Extensively	0.20 (0.40)	0.27 (0.45)	0.10 (0.29)	0.26 (0.44)
Observations	534	534	534	532
<i>Panel B: Effects of Match Characteristics on Probability Topic Covered at Least Some Time</i>				
<i>Policy Disagreement</i>				
Covid-19 measures	0.04 (0.06)	0.02 (0.07)	0.04 (0.04)	0.03 (0.07)
Germany is racist	-0.03 (0.06)	0.00 (0.06)	0.02 (0.04)	0.00 (0.06)
Sanctions for russia	-0.07 (0.06)	0.03 (0.07)	0.02 (0.04)	-0.11* (0.07)
Rent control	-0.06 (0.09)	-0.03 (0.10)	-0.16*** (0.06)	0.12 (0.10)
Support for rural areas	-0.04 (0.05)	0.04 (0.06)	-0.07* (0.04)	-0.08 (0.06)
Ban domestic flights	-0.02 (0.07)	0.02 (0.08)	0.02 (0.05)	-0.04 (0.08)
<i>Partner Demographics</i>				
Same Gender	0.09** (0.04)	0.01 (0.05)	0.03 (0.03)	-0.00 (0.05)
Age Difference < 10	0.14*** (0.04)	-0.01 (0.05)	0.00 (0.03)	0.02 (0.05)
Same State	-0.03 (0.07)	0.06 (0.08)	0.02 (0.05)	0.02 (0.08)
Outcome Mean	0	0	0	0
Outcome SD	1	1	1	1
Policy Question Controls	Yes	Yes	Yes	Yes
Observations	361	361	361	360

Notes. This table reports self-reported conversation topics (panel A) and results from regressions predicting self-reported conversation topics (panel B) among participants who met their partner (treated). In panel B, we regress a dummy indicating whether the conversation topic indicated in the header was at least covered some time on a set of dummies indicating whether the two assigned partner disagreed on a policy domain or shared demographic characteristic. Respondents' self-reported whether they covered a topic on the following scale: not at all, some time, extensively, most of the time. All regressions include controls for the sign-up responses of the participant. Policy disagreement refers to dummies indicating whether the participant provided a different binary response to the policy question than their partner. Standard deviation (panel A) and robust standard errors (panel B) in

## A.7 Additional Results for Trust Game and Games Pooled

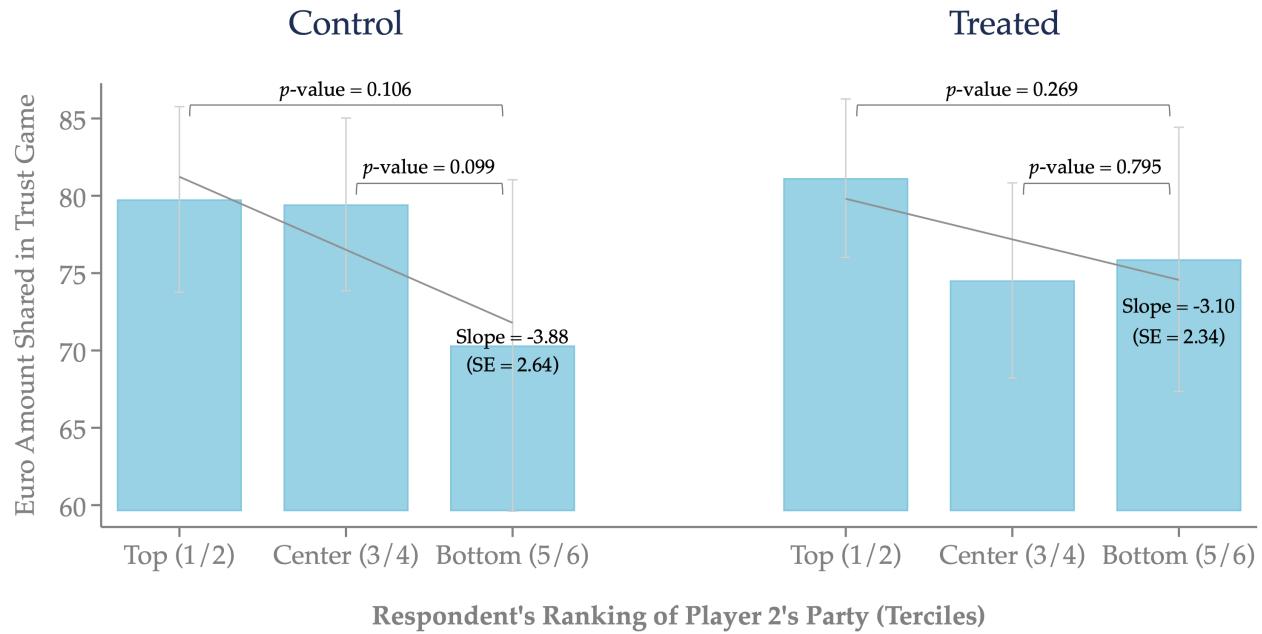


FIGURE A.4: Treatment Effects on Affective Polarization: Trust Game

*Notes:* This figure plots the Euro amounts shared in the trust game among control (left panel) and treated subjects (right panel), separately by terciles of the respondent's ranking of the party of the randomly assigned other player. Top refers to parties the respondent ranks first or second, Center to parties the respondent ranks third or fourth, and Bottom to parties the respondent ranks fifth or sixth. Confidence Intervals are at the 95% level.

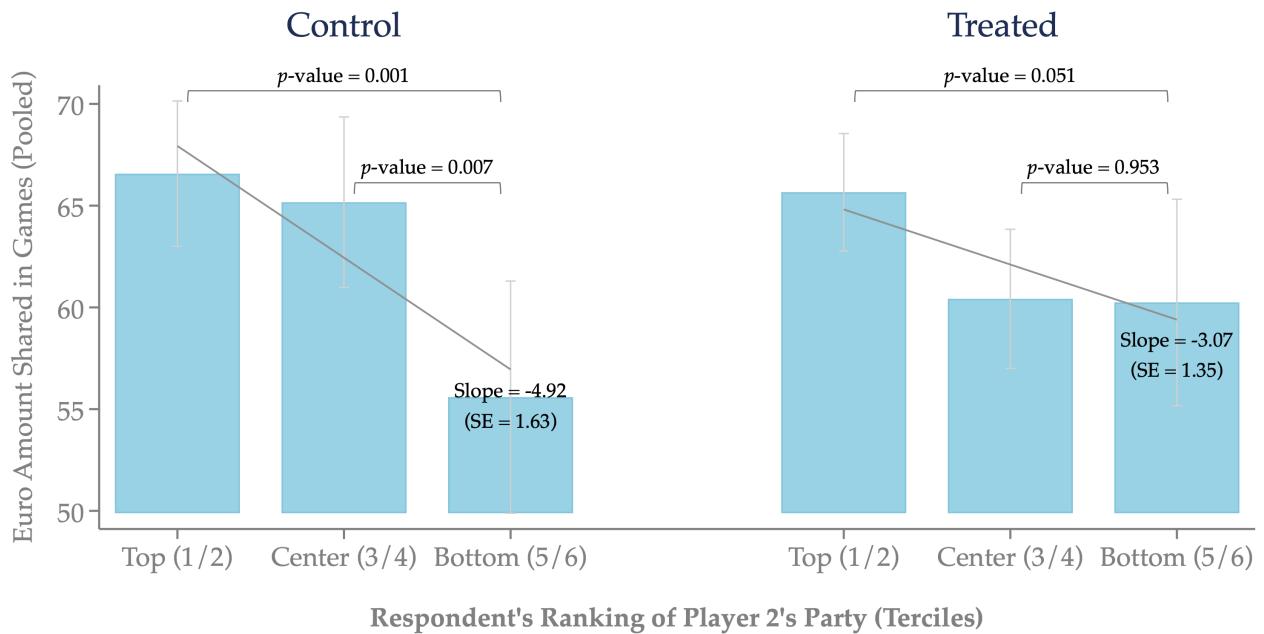


FIGURE A.5: Treatment Effects on Affective Polarization: Dictator and Trust Game Pooled

*Notes:* This figure plots the Euro amounts shared in the dictator and trust game among control (left panel) and treated subjects (right panel), separately by terciles of the respondent's ranking of the party of the randomly assigned other player. We pool the amounts shared in the two games by first subtracting the mean of each individual game, then creating a joint variable, and finally multiplying the value with the mean of the original amounts shared in the dictator and trust game. Top refers to parties the respondent ranks first or second, Center to parties the respondent ranks third or fourth, and Bottom to parties the respondent ranks fifth or sixth. Confidence Intervals are at the 95% level.

## A.8 Reweighting Sample Characteristics

TABLE A.5: Robustness: Game Results with Re-weighted Sample

	No Weights		Age X Gender X Migrant Status		Age X Gender X Education		Voting Behavior	
	No Baseline Controls (1)	Baseline Controls (2)	No Baseline Controls (3)	Baseline Controls (4)	No Baseline Controls (5)	Baseline Controls (6)	No Baseline Controls (6)	Baseline Controls (7)
Treated ( $\tau$ )	0.20 (0.14)	0.19 (0.14)	0.35** (0.18)	0.31* (0.18)	0.33* (0.19)	0.36** (0.17)	0.14 (0.18)	0.23 (0.17)
Ingroup Bias ( $\beta$ )	0.43*** (0.13)	0.44*** (0.12)	0.49*** (0.15)	0.51*** (0.15)	0.40* (0.22)	0.45** (0.20)	0.33** (0.14)	0.42*** (0.14)
Treated X Ingroup Bias ( $\delta$ )	-0.30* (0.16)	-0.30* (0.15)	-0.37* (0.20)	-0.36* (0.19)	-0.43* (0.23)	-0.47** (0.21)	-0.13 (0.20)	-0.24 (0.19)
Baseline controls	Yes		Yes		Yes		Yes	Yes
Game Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	892	892	892	892	892	892	855	855

Notes. This table reports results from regressions of the standardized amount shared in the one of the two games (pooled) on dummies indicating whether the randomized player 2 belongs to the respondents party ingroup (ranked 1st to 4th, top and medium tercile), treatment status, and the interaction of the two. The omitted category are respondents randomized to a player 2 who is affiliated with a party the respondent ranks 5th or 6th (the bottom tercile). Columns 1 and 2 present regressions results without any weights as in our baseline specifications. Subsequent columns are weighted in different ways to account for the selectivity of the sample along observable characteristics. Columns 3 and 4 weight respondents along age by gender by migration status buckets, columns 5 and 6 weight respondents by age by gender by education status bucket and columns 7 and 8 weight respondents by their voting behavior (i.e. whether they voted and which party they voted for) in the 2017 federal election. The sample is limited to participants who met their partner (treated) or already scheduled a meeting with their partner (control). All regressions include controls for the other randomized characteristics of player 2 and baseline party rankings. Columns (2), (4), (6) and (8) add demographic controls and baseline affective polarization controls. Robust standard errors in parentheses.

## A.9 Treatment Effects on Affective Polarization: Issue Thermometer

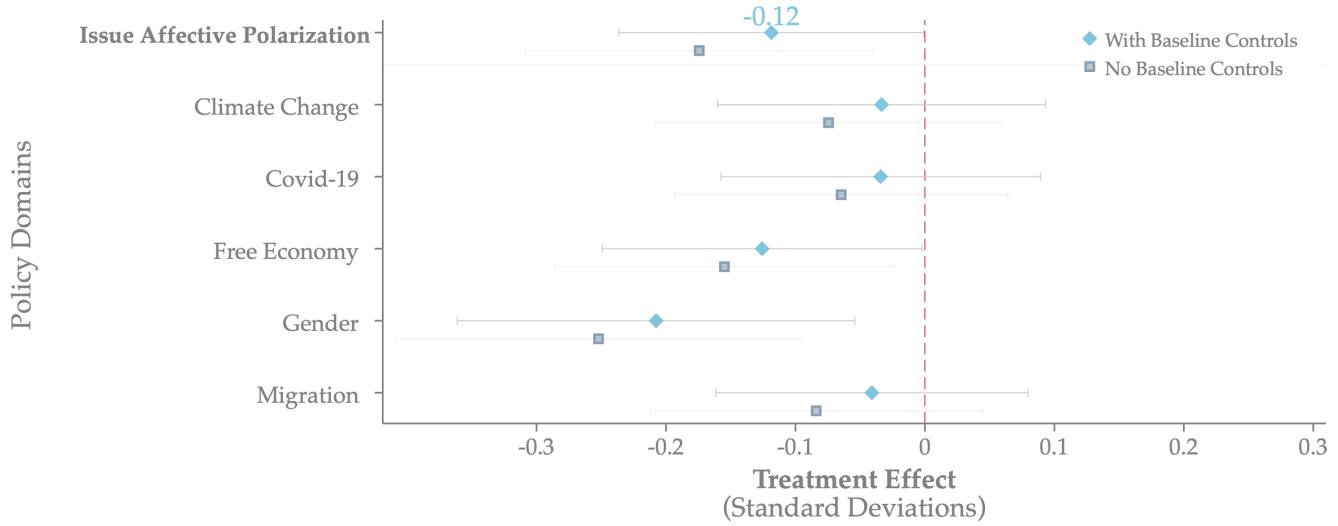


FIGURE A.6: Treatment Effects on Affective Polarization: Issue Thermometer

*Notes:* This figure plots results from regression models, regressing differences in thermometer levels between supporters and opponents of five policy issues on treatment status. The outcome measure is computed by standardizing the difference in the mean thermometer level (ranging from 0 to 100) between a supporter of a positive and a supporter of a negative statement of the five policy domains. The average effect is an unweighted average of the five standardized differences below. In the control group, we include respondents who have scheduled a meeting but have not yet met a partner. The treatment group is comprised of respondents who have already met their partner. Coefficient estimates shown in dark blue squares do not include controls while those in the light blue diamonds control for baseline demographics and baseline levels of the outcome measure.

Confidence Intervals are at the 95% level.

## A.10 Donation Results

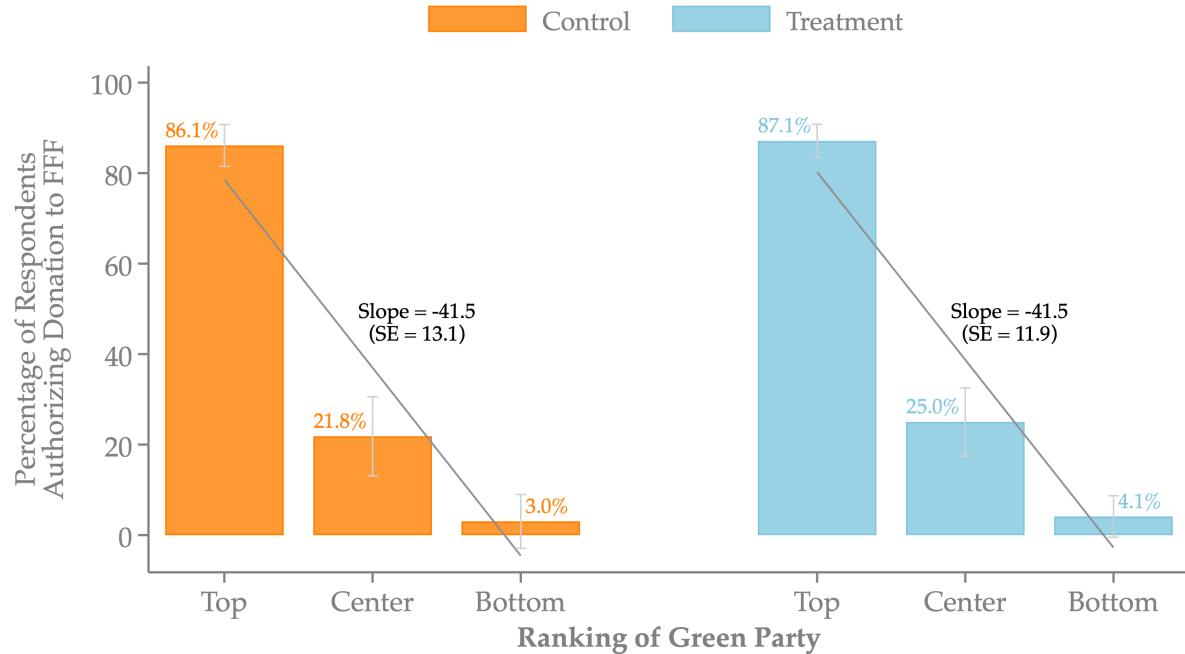


FIGURE A.7: Likelihood to Authorize Donation to Environmental Non-profit by Treatment Status

*Notes:* This figure presents the percentage of respondents who authorize a donation to Fridays for Future (FFF) by treatment status and by whether the respondent ranks the Green Party at the top (first or second), the center (third or fourth) or at the bottom (fifth or sixth) in the baseline survey. Bars on the left (in orange) present results for respondents in the control group, and bars on the right (in blue) show results for those in the treatment group. Confidence Intervals are at the 95% level.

### A.11 Feeling Thermometer Rating for GT Partner and Party Supporters

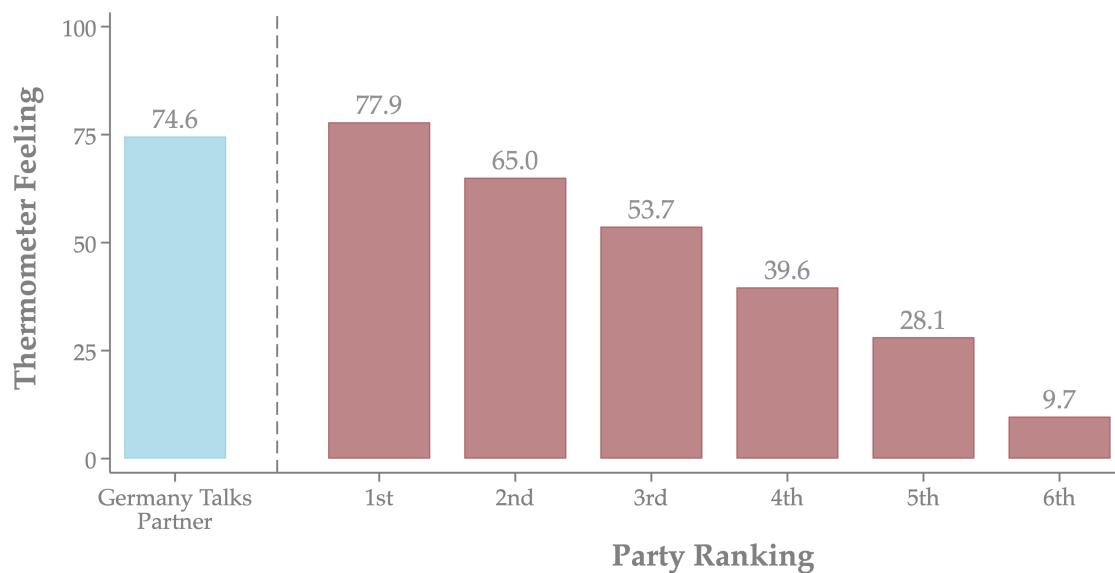


FIGURE A.8: Feeling Thermometer Rating GT Partner and Party Supporters

*Notes:* This figure plots the average feeling thermometer rating for the Germany Talks conversation partner against the average feeling thermometer ratings for party supporters, separately by the ranking of the corresponding party by each respondent.