The Consequences of Interparty Conversation on Outparty Affect and Stereotypes[*]

Erin Rossiter

Department of Political Science
Washington University in St. Louis

April 13, 2020

Abstract

Americans increasingly dislike members of the opposite political party and associate negative stereotypes with them such as close-minded, mean, and hypocritical. Yet, Americans interact, whether talking about politics or not, with opposing party members in their everyday lives. How do conversations across party lines impact the negative feelings and perceptions Americans hold for opposing party members? How might the consequences of conversations that touch on politics differ from those that do not? I assess the effect of interparty conversation on how partisans feel and think about the outparty using an experiment that manipulates whether a pair of opposing party members engage in conversation or not, and if so, whether they discuss a political or non-political topic. This experiment takes two novel approaches. First, I develop an algorithm to implement a blocked cluster design in settings where the researcher controls what clusters (e.g., conversations) form. Second, I develop a chat software so participants can have real-time, written conversations online. I find that conversation mitigates negative outparty affect and deters future use of negative stereotypes to describe the outparty. This positive effect holds for both non-political and political conversations. Surprisingly, I do not find evidence to suggest that talking politics is any less effective than avoiding overtly political topics for improving a partisan's negative view of the outparty. These results provide new evidence that interparty conversation, whether politically-charged or not, can work to undo the negative view of outparty members held by many Americans.

1 Introduction

Americans increasingly dislike members of the opposite political party and associate negative stereotypes with them such as close-minded, mean, and hypocritical. Yet, Americans interact, whether talking about politics or not, with opposing party members in their everyday lives. How do these conversations that cross party lines impact the negative feelings and perceptions Americans hold for opposing party members? When might conversation with an opposing party member reinforce, and when might it mitigate, the negative view of outparty members held by many Americans? How might the consequences of conversations that delve into political topics differ from those that do not?

Scholars often understand negativity toward the outparty through the framework of partisanship as a social identity. As a social identity, self-perceived membership in a political party leads Americans to categorize the world into the "ingroup" (their own party) or the "outgroup" (the opposing party) (Tajfel and Turner 1979), and this categorization impacts the way Americans *feel* and the way they *think* about the outparty. Negative feelings and perceptions about the outparty stem from a motivation to feel positively about ourselves and our ingroup, which is accomplished though direct comparison to the outgroup, or the outparty. As the country continues to divide along party lines, Americans' partisan identities grow stronger and outparty negativity intensifies (Mason 2018).

There are two plausible expectations concerning how a negative view of the outparty might change after conversing with an outparty member. One expectation, stemming from the increasing strength of Americans' partisan identities, is that exposure to outparty members will cue negative outparty feelings and perceptions, leading to an interaction that does not improve, or even worsens, an already negative view of the outparty. A second expectation stems from the literature on intergroup contact (e.g., Allport 1954), which finds that contact, such as conversation, tends to be an effective strategy for improving negative outgroup attitudes for other social groups, so the idea is that this strategy should apply to partisan groups, as well.

While it may seem as though social identity theory and theories of intergroup contact are in tension when it comes to explaining the effects of interparty conversations, I argue that these theories are actually more complementary than is currently appreciated in the political science

literature (e.g., Bond, Shulman and Gilbert 2018). Because social identity theory explains *why* we view outparty members negatively, it is a useful basis for understanding how these feelings and perceptions might *change* via contact. And by considering these theories in tandem, we can consider how the consequences of conversations that cross party lines may differ if partisans talk about family, friends, and hobbies, or if partisans talk about gun control, immigration, or other salient political issues.

Specifically, our social identities help us initially make sense of, and decrease uncertainty surrounding, any social interaction. So when conversing with an outparty member, we initially categorize them as such, and negative perceptions and stereotypes about the outparty fill in for what we do not yet know. Contact, however, can alter a partisan's representation of "outparty member." Conversations with outparty members that occur in our everyday lives are usually informal, impromptu, and lack any particular goal. These interactions allow opposing partisans to view each other as individual people rather than outparty prototypes, breaking down the understanding of the outgroup member by the negative terms which define the outparty (Brewer and Miller 1984). In doing so, I expect that outparty contact undermines the usefulness of relying on partisan identity to make sense of future social interactions and interferes with the use of a biased, oversimplified representation of the outparty, generalizing the benefits of contact to the outparty at large.

A positive social interaction is fairly easy to imagine when conversation avoids overtly political topics—talking about work, family, or hobbies. These conversations are likely to promote an understanding of opposing party members on a personal level, replacing partisanship as the most useful lens to view them through. However, it is harder to imagine conversation that delves into political topics having a similar effect. It is unclear if political conversation provides an environment for sharing information that allows opposing party members to view each other as anything other than "the outparty." Nonetheless, I expect that positive, personalized interactions can occur even when conversation drifts into political topics, such that political conversation across partly lines can improve negative outparty affect and stereotypes as well.

In this paper, I test these claims using an experiment in which I manipulate whether a pair of opposing party members converse with each other or not, and if so, whether they discuss an explicitly political topic or not. I am interested in two main outcomes: how partisans feel about the outparty and how partisans think about the outparty after conversation with an outparty member.

To rigorously assess my hypotheses, I make two methodological innovations. First, I develop an algorithm to implement a blocked cluster design in experimental settings where the researcher controls what clusters (e.g., conversation partnerships or groups) form, which is common in the political discussion and deliberation literature. Second, I designed a chat software called "Chatter" by which participants can have real-time, written conversations online. Chatter allows me to relatively easily emulate a real social experience without an existing academic laboratory. Taken together, the experimental design and chat software overcome a set of methodological and practical concerns to improve the experimental study of social interaction.

I find that interparty interaction mitigates negative outparty affect and deters future use of negative stereotypes to describe the outparty. This positive effect holds for both non-political and political conversations. Surprisingly, I do not find evidence that the effect of political conversation is any less than the effect of non-political conversation on improving negative outparty affect and perceptions. These results provide new evidence that interparty social interaction, regardless of whether the conversation is politically-charged or not, can work to undo the negative representation of outparty members held by many Americans. I conclude with a discussion of when we might expect partisans to engage in interparty interactions in their everyday lives, scope conditions that describe when interparty social interactions are likely to have these effects, and opportunities for future research.

2   How Americans feel and think about outparty members

The American political climate is characterized by a pronounced *non-ideological* division across party lines that is often said to take the general form of a "dislike and distrust" for members of the opposing party (see Iyengar et al. 2019). Research shows this trend manifests in many ways. There is a well-documented *affective* response toward outparty members—Republicans and Democrats increasingly report feeling negative toward members of the outparty (e.g., Iyengar, Sood and Lelkes 2012). Additionally, there is a *cognitive* response toward outparty members—partisans hold a negative, over-generalized representation of the outparty. This outparty representation includes negative trait stereotypes (Iyengar, Sood and Lelkes 2012), an overestimation of the extent to which outparty members belong to groups stereotypically associated with the outparty (Ahler and

Sood 2018), and an overestimation of the extremity of opposing party members' political views (Levendusky and Malhotra 2015).

One explanation for this general trend is a partisan-ideological sorting—conservatives increasingly identify as Republican and liberals increasingly identify as Democrat (Levendusky 2009; Mason 2015). Not only have ideological and partisan identities aligned, but race, religion, and more have sorted along the same partisan divide (Mason 2018). As identities that cut across party lines have decreased, the strength of Americans' partisan identities has increased, which has affective and cognitive consequences. Stronger partisans react with stronger emotion to perceived party threats, regardless of their ideological positions (Mason 2015, 2018). And, as clearer social distinctions are made between the parties and as Americans hold stronger partisan identities, it becomes easier and more tempting to make (potentially inaccurate) generalizations about the outparty (Westfall et al. 2015).

Talking across party lines has repeatedly been cited as a solution to America's deep, bitter partisan divide. Not only do media (e.g., Grumet 2019) and politicians (e.g., Fang 2017) offer this advice, but nonprofits spend a great deal of money promoting this philosophy.[1] But because the majority of political science research on the consequences of interparty conversation focuses on outcomes such as the sharing of political information (e.g., Berelson et al. 1954; Carlson 2019) and political participation (e.g., Huckfeldt and Sprague 1991; McClurg 2003; Mutz 2006; Sinclair 2012), the consequences of talking with the political opposition on how Americans feel about the outparty remains less clear. How one's view of the outparty changes in reaction to interparty conversation surely has downstream consequences on other political outcomes of social interaction, such as if information was distorted, if participation was hampered, and more. Therefore, it is important to also shed light on the immediate social psychological outcomes of conversation, such as how we feel and think about outparty members (e.g., Mutz 2002).

Additionally, it is important to study the consequences of interparty conversation on the feelings and perceptions of the outparty because we can derive two plausible expectations about this process from the literature. Consider the view of American partisanship offered by Mason (2018), who calls partisanship, now aligned with many other identities, a "mega-identity" which heightens feelings of

---

[1]For example, Better Angles (https://www.better-angels.org/), Village Square (https://tlh.villagesquare.us/), American Public Square (https://americanpublicsquare.org/), Listen First Project (http://www.listenfirstproject.org/), and many more.

anger, competition with the outparty, and a need to "win" not just in terms of political interests but in terms of protecting their partisan "team." On the one hand, it follows that these feelings of anger, competition, and threat might surface at the prospect of conversation with an opposing party member. Thus, the drive to maintain a win for one's partisan team could lead to an interaction that fails to improve, or even worsens, one's negative view of the outparty. On the other hand, conversation with outparty members could offer an opportunity to improve negative outparty attitudes as a form of intergroup contact, which is considered "one of psychology's most effective strategies for improving intergroup relations" (pg. 5 Dovidio, Gaertner and Kawakami 2003), and multiple meta-analyses have since shown that contact has the tendency to improve negative outgroup evaluations (Paluck, Green and Green 2019; Pettigrew and Tropp 2006). In sum, the political science literature lacks a clear understanding of how conversing with an opposing party member alters, if at all, the way partisans feel and think about the outparty.

3 Consequences of interparty conversation on outparty affect and stereotypes

In this section, I propose a resolution to the implicit tension in the political science literature between social identity theory and theories of intergroup contact. Specifically, I outline how social identity theory is a useful basis for understanding the consequences of contact with an outparty member, and I derive three expectations: that (1) non-political conversation and (2) political conversation with an outparty member improves a partisans' negative view of the outparty at large, but that (3) non-political conversation will be more effective than political conversation at doing so.

3.1 Partisanship as a social identity

Social identity theory (e.g., Tajfel and Turner 1979) and the related self-categorization theory (e.g., Turner et al. 1987) present one framework for understanding intergroup, including *interparty*, biases and prejudices (Green, Palmquist and Schickler 2004; Greene 1999; Iyengar, Sood and Lelkes 2012). According to this framework, individuals associate with groups as a cognitive tool to understand their place in a complex, social world. As a consequence of forming a social identity, an individual's sense of self becomes bound to the group, so maintaining a positive sense of self is tied to maintaining

a positive view of the ingroup. Because the ingroup is understood in comparison to the outgroup, social identity theory hypothesizes that individuals are motivated to positively differentiate the ingroup from the outgroup——I like "us" more than "them." Research supports this hypothesis, showing that partisans express explicit and implicit favoritism, or bias, for members of the inparty even in non-political settings (e.g., Iyengar and Westwood 2015).

When partisanship is salient in a social interaction, individuals initially categorize themselves and others as inparty or outparty members. There are two main consequences of this. First, when an individual self-categorizes, they comprehend and act within the interaction in accordance with how they see themselves as "a Republican" or "a Democrat" (Turner et al. 1987). Second, when a partisan categorizes an outparty member as such, they depersonalize the outparty member, thus viewing the outparty member as an oversimplified prototype of the broader group (Hogg and Reid 2006; Tajfel 1981). Affective and cognitive responses toward the outparty *as a whole* are applied to the outparty member. As discussed above, partisans hold a negative representation of the outparty that takes the form of a negative affect and negative trait stereotypes. This negative representation of the outparty will be used to form expectations about the outparty member's beliefs and behavior at the start of a social interaction.

3.2   Conversations as contact

In light of what the social identity perspective says about an individual's initial understanding of social interaction with an outparty member, how might interparty conversation unfold? Can contact alter the initial, negative representation of an outgroup member? I expect that contact can help partisans move beyond a prototypical understanding of the outparty member when personal, individuating information is shared (e.g., Brewer and Miller 1988, 1984; Ensari et al. 2012; Miller 2002). When such information is shared, partisanship can shift from being the most useful, or even the only, dimension shaping an understanding of the outparty member (and one's self). Instead of categorizing an outparty member as such, social interaction allows the outparty member to be understood better as an individual person, breaking down the initially relied upon inparty/outparty categories.

However, the benefits of contact are limited if they extend to only the outgroup member involved in the immediate social interaction. Can contact alter the negative representation of the outgroup as a

whole? While individuals initially understand a social interaction by categorizing—"us" verses "them," social interaction undermines this categorization by fostering interpersonal comparisons. Future interactions with outparty members then benefit from the personalized, rather than prototypical, representations of the outparty formed from the interparty contact (Brewer and Miller 1988, 1984). Moreover, evidence suggests that personalized contact improves biased outgroup affect and perceptions by interfering with the inclination to make sense of future social interactions based on those ingroup/outgroup categories (e.g., Bettencourt et al. 1992; Miller, Brewer and Edwards 1985). Therefore, I expect that interparty conversation will undermine the usefulness of partisan identity as a basis for making sense of future social interactions, and as a result, contact will improve a biased, oversimplified representation of the outparty.

However, different types of conversations may lead to variation in how effective contact is at fostering a personal, rather than prototypical, understanding of the outparty member. Consider *what* partisans talk about. It is relatively easy to imagine non-political conversations——talking about family or hobbies—personalizing the outparty member. Even though conversation about family, hobbies, or even your pets at home can cue partisanship (Hetherington and Weiler 2018), non-political conversations encourage partisans to understand each other as individuals, beyond (potentially incorrect) stereotyped views, group associations, and traits (e.g., Ahler and Sood 2018; Iyengar, Sood and Lelkes 2012; Levendusky and Malhotra 2015). Therefore, I expect non-political conversations can improve generalized outgroup affect and decrease the reliance on negative outgroup stereotypes in future interactions.

> **Non-political contact hypothesis**: Non-political conversation as contact with an outparty member, relative to no contact at all, improves (1) negative outparty affect and (2) negative outparty stereotypes.

It is less clear if interparty conversation can improve the negative view of the outparty even when conversation is about overtly political topics. However, I expect that in the context of political conversation, information can be shared that personalizes the outgroup member. While the outparty member begins as just one undifferentiated member of the outgroup, this understanding can be replaced with an understanding of them as an individual person, on both non-political and political dimensions. Therefore, I expect political conversations can also improve generalized outgroup affect and decrease the reliance on negative outgroup stereotypes in future interactions.

Political contact hypothesis: Political conversation as contact with an outparty member, relative to no contact at all, improves (1) negative outparty affect and (2) negative outparty stereotypes.

While I expect that both non-political and political conversation can facilitate the benefits of contact generalizing to the outparty as a whole, there are important differences between a non-political and political conversations to consider. Partisan group membership is inherently more salient when talking about politics, heightening the initial motivation to secure a win for one's partisan team. So when talking about politics, it may be more difficult to move beyond an understanding of the social interaction based on an inparty/outparty categorization. Therefore, I expect non-political conversation will be more effective than political conversation for improving negative outparty affect and stereotypes.

Topic hypothesis: Non-political conversation improves (1) negative outparty affect and (2) negative outparty stereotypes more than political conversation.

In summary, I expect that social interaction across party lines tends to improve a partisans' negative view of the outparty at large. I expect that non-political and political conversations are both effective settings to personalize the outparty member, break down inparty/outparty categories, and improve negativity toward the outparty. However, because partisan identities become more salient when conversation turns to politics, I expect non-political conversation will be more effective than political conversation at improving how partisans feel and think about the outparty.
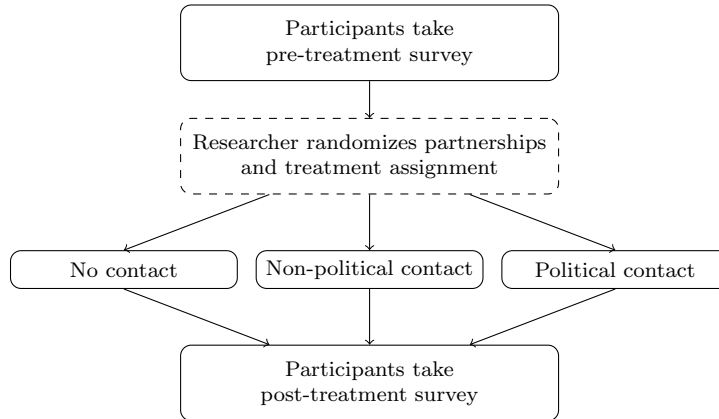
4 Experiment

To assess the consequences of interparty conversation on outparty affect and stereotypes, I conducted an experiment on Amazon Mechanical Turk (MTurk) involving conversation across party lines.[2] The experiment required four steps outlined in Figure 1. First, a set of potential participants took a pre-treatment survey to gather relevant pre-treatment covariates. At the conclusion of the survey, participants were asked if they would be willing to return for a follow up task involving an "online chat with another Worker or writing a short essay."

Second, I used the pre-treatment survey responses of participants willing to return for the follow-up task to randomize participants into partnerships, each containing one Republican and

---

[2]See Appendix F for participant compensation details.

Figure 1: Experimental stages

Note: Visualization of participants' and researcher's roles throughout the four stages of the experiment.

one Democrat. Then, I randomly assigned conversation partnerships to one of three experimental conditions: no contact with partner (instead write an individual short essay), (2) contact with partner and discuss a non-political topic, or (3) contact with partner and discuss a political topic. Participants selected for the experiment were invited via email through MTurk to complete the follow-up task. Having participants return for the experiment at all, let alone at the same time, presented a difficult coordination task. To minimize attrition between the pre-treatment survey and returning for the experiment, participants took the pre-treatment survey 10-30 minutes before the experiment. With the remaining 10 minutes, I randomized participants into partnerships, assigned partnerships to experimental conditions, emailed chosen participants 5 minutes before the experiment was live, and emailed chosen participants once more when experiment was live.

Third, participants selected for the experiment returned to complete the follow-up task where they spent a minimum of eight minutes writing a short essay or conversing with their assigned partner. Fourth, after completing their assigned task, participants proceeded to a post-treatment survey to assess how their feelings and perceptions about the outparty may have changed.

In what follows, I discuss several of these steps' details—the three experimental conditions, the measurement of outcome variables, how partnerships and treatment were randomly assigned via a blocked cluster experimental design, and finally, how conversation occurred via an online chat app.

4.1 Experimental conditions

The short essay and conversation prompts are shown in Table 1. Specifically, for partners assigned to have **no contact** with their outparty partner, each individual wrote separately about the meaning of life. For those assigned to the **non-political contact** condition, participants talked with their outparty partner about the meaning of life. I selected this topic because previous research has investigated how to induce close relationships, and thus foster personalized contact, in a laboratory experiment setting, finding that participants grow closer during a short interaction when communicating about "deep" (i.e., What is the meaning of life?) rather than "shallow" questions (e.g., What is your name? Where are your from?) (Sedikides et al. 1999; Tu, Shaw and Fishbach 2015). For those assigned to the **political contact** condition, participants conversed with their outparty partner about gun control. I selected this topic because it is a political issue salient to the average American so most participants are likely to have opinions they can converse about for a few minutes. Importantly, since intergroup contact implies outgroup membership is known during the contact situation, participants are told the partisanship of their conversation partner.

4.2 Outcome measures

I have two outcomes of interest. The first uses the standard 101-point feeling thermometer, where larger values indicate more favorable or "warm" feelings toward that person or group (e.g., Iyengar, Sood and Lelkes 2012).[3] Respondents rate both "Republicans across the country" and "Democrats across the country" in pre- and post-treatment surveys. My first outcome of interest is how contact can alter generalized *outparty affect*, which I define as the difference between pre- and post-treatment outparty feeling thermometer ratings.[4]

The second set of outcomes assess how contact can alter *perceptions of the outparty*, which I measure by asking participants to rate, using a five point Likert scale, how well several traits describe members of each political party (e.g., Levendusky 2018). Trait ratings were asked in the

---

[3] Feeling thermometer question wording is, "Please rate the following groups using the following thermometer. Ratings between 50 and 100 degrees mean that you feel favorable and warm toward the group. Ratings between 0 and 50 degrees mean that you don't feel favorable toward and don't care too much for that group. You would rate a group at the 50 degree mark if you don't feel particularly warm or cold toward the group."

[4] When using feeling thermometers and trait rating, research shows it is important to specify the group you want participants to rate beyond "the Democratic party" or "Democrats," for example (Druckman and Levendusky 2019). I ask respondents to rate "Democrats across the country" to target not only Democratic voters, but the entire outgroup that identifies as a Democrat.

Table 1: Instructions for experimental conditions

No contact	**Non-political contact**	**Political contact**
Please share your thoughts on the meaning of life. A conversation partner will not be joining you. Rather, we ask that you write about the meaning of life independently by sending messages in the chat box below.	We've randomly assigned you a partner that belongs to or leans toward the [Republican/Democratic] party. Please have a conversation with them about the meaning of life.	We've randomly assigned you a partner that belongs to or leans toward the [Republican/Democratic] party. Please have a conversation with them about gun control.
For example, survey research shows that many people mention family as the most important sources of meaning in their life. Survey research also shows that other people mention career, money, faith, friends, and hobbies as the most important source of meaning in their life.	Specifically, we are interested in you sharing what you think makes life meaningful and learning your conversation partner's thoughts as someone that might hold different values and beliefs.	Specifically, we are interested in you sharing your opinion on gun control and learning your conversation partner's opinion as someone that might hold different values and beliefs.
What do you think?	For example, survey research shows that many people mention family as the most important sources of meaning in their life. Survey research also shows that other people mention career, money, faith, friends, and hobbies as the most important source of meaning in their life.	For example, survey research shows that some people believe it is more important to protect the right of Americans to own guns than control gun ownership, while others believe the opposite. Survey research also shows that some people believe making it harder to own guns would result in fewer mass shootings, while others believe this would make no difference.
	What do you think?	What do you think?

Note: Instructions for three experimental conditions. References to survey research included in the prompts come from recent Pew Research Center polls (Pew Research Center 2017, 2018).

post-treatment survey only.

Respondents also indicate their partisan identification in the pre-treatment survey.[5] Participants who choose Independent or Other are asked toward which party they lean. Due to the evidence that "leaners" behave similarly as partisans (Greene 1999; Iyengar and Westwood 2015; Pew Research Center 2019), I collapse Independents who lean toward one party into that party. Using partisan identification, I identify each individual's outparty in order to construct the outcome measures of outparty affect and perceptions.

---

[5]Partisan identification question wording is, "Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what?" A follow up question asks strength of partisan identity, "Would you call yourself a strong [Republican/Democrat] or not a very strong [Republican/Democrat]?" Finally, participants indicating Independent or Other are asked, "Do you think of yourself as closer to the Republican or Democratic party?"

4.3 Experimental design

Lab experiments[6] involving social interaction amongst participants, like this one, are common across the political discussion and deliberation literature (e.g., Druckman and Nelson 2003; Karpowitz, Mendelberg and Shaker 2012; Klar 2014); however, several methodological and practical concerns arise with this type of experiment. Not only does social interaction complicate a researcher's design and subsequent data analysis, but small sample sizes, imbalance across experimental conditions, and more have implications for efficiency of estimation and the power of hypothesis tests. And as a practical matter, experimental studies involving participant interaction are resource-intensive, often prohibitively so, largely requiring an academic lab and existing subject pool.[7]

To rigorously test the hypotheses derived in Section 3, I sought to address several of these methodological and practical concerns through two specific approaches to the experiment. First, I implemented a blocked cluster experimental design to improve efficiency of my estimation, among other advantages. Second, I developed a chat software to more easily allow for participant social interaction. In what follows, I briefly discuss each of these approaches in turn.
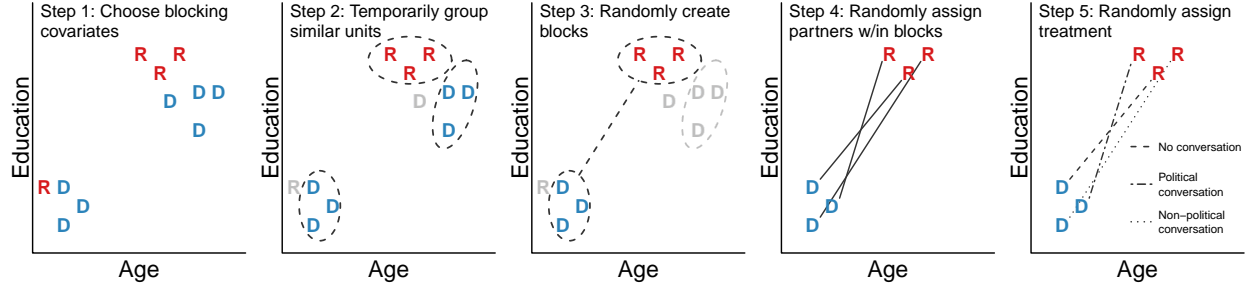
4.3.1 Blocked cluster design

For this experiment, I chose a blocked cluster design because (1) randomly assigning treatment at the *cluster*-level (here, conversation-level) is appropriate due to inherent interference between participants within a social interaction, and (2) randomly assigning treatment within *blocks* of clusters improves efficiency in estimation (e.g., Moore 2012). However, a blocked cluster design is typically used for field experiments which feature pre-existing clusters, such as cities or classrooms (e.g., Imai et al. 2009). To implement this design for a lab experiment, the researcher must somehow assign units to clusters (i.e., individuals to groups or partnerships). While guidance and tools exist for blocking (e.g., Moore 2012) and blocking with pre-existing clusters (e.g., Imai et al. 2009), it is less clear how to simultaneously block and cluster units. Therefore, I created an algorithm to construct a blocked, *randomized* cluster design.[8] Importantly, this design allows me to ensure

---

[6]"Lab experiment," as used here, refers to a context in which researchers have a high degree of control, particularly in the randomization of participants into groups and the randomization of treatment and control assignment. Lab experiments, by this definition, are not necessarily conducted in an academic laboratory.

[7]Appendix A discusses these methodological and practical concerns in more detail.

[8]It is important to note that the proposed blocked, randomized cluster experimental design is certainly more complicated than one featuring completely randomized assignment of clusters and of the treatment. If the costs, in

Figure 2: Blocked, randomized cluster experimental design algorithm

Note: Visualization of the algorithm for constructing a blocked cluster design when the researcher controls the construction of the clusters.

random assignment of what clusters (i.e., conversation partnerships) form, to improve efficiency of my estimation by block randomizing the treatment (at the cluster-level), and more.[9]

Figure 2 outlines the five steps of my blocked, randomized cluster design algorithm. I use the hypotheses from this paper as an example. There are a few specifics about this example to highlight before explaining the steps of the algorithm. First, for this example, I want three similar clusters per block because I have three experimental conditions. Second, I want each cluster to have two participants. Third, I want each cluster to feature one Republican and one Democrat.[10] For this reason, I'll call partisanship my "clustering constraint," or the variable the created clusters will be constrained to reflect. Importantly, the clustering constraint must apply to all clusters to ensure the benefits of balance achieved by blocking.[11]

The first step of the algorithm, demonstrated in the first plot of Figure 2, is to identify relevant blocking covariates and the clustering constraint, if any. For simplicity, I plot participants on only two dimensions—education and age. Because these variables likely affect the extent to which

terms of complication, outweigh the benefits, I would not recommend this design. For example, if a large-n study is possible, then completely randomized groups and treatment assignment may be sufficient. Assessing power, efficiency, and more via DeclareDesign will prove helpful in these design decisions (Blair et al. 2019; Blair and Fultz 2019)

[9]Appendix C presents simulations results suggesting the proposed blocked, randomized cluster design improves efficiency and power over a design featuring simple randomization of individuals to partnerships and treatments assignments.

[10]This algorithm is generalizable to any number of experimental manipulations, any number of units per cluster, and any clustering constraint, such as disagreeable attitudes on the topic of discussion, different gender, or none at all.

[11]A researcher may want to consider group composition as a treatment, such as creating same-party groups as a control condition for opposite-party conversation. In this case, one could imagine randomly assigning units to clusters to reflect one of these two clustering constraints (i.e., experimental conditions). However, careful consideration would need to be made to ensure doing so does not induce imbalance across experimental conditions on other covariates likely to impact the outcome of interest. Additionally, considering group composition as treatment (rather than fixed, pre-treatment) complicates consideration of a unit's potential outcomes. This type of design is beyond the scope of this paper.

participants will change their outparty affect, I block on these variables to control for this variation. I also indicate if the participants are Republican or Democrat because this is the variable all created clusters will be constrained to reflect—every cluster will have one Republican and one Democrat.

The second step is to identify temporary groupings of n similar units with respect to the clustering constraint, where n is the number of experimental manipulations.[12] The second plot in Figure 2 shows the groupings of three similar units, conditional on partisan identification. Importantly, these groupings are **not** the clusters; rather, they are temporary groupings of similar units used to facilitate the creation of blocked and randomized clusters in subsequent steps.

The third step finishes the process of creating the blocks. I randomly assign each temporary grouping to another temporary set of units, conditional on having different partisanship. For this example, one group of similar Democrats is randomly assigned to one group of similar Republicans. These six individuals represent one block. Any remaining groups or individuals not assigned to blocks (as shown in gray) are discarded. It may seem counterintuitive to finalize the blocks before finalizing the specific cluster assignments. However, creating the blocks first ensures cluster-level similarity within each block.

The fourth step is to randomly assign clusters. Within each block, I randomly assign one unit from each temporary grouping to a unit from the other grouping. The fourth plot of Figure 2 shows this process for one block—Democrats and Republicans are randomly assigned to each other. The result is three similar, randomly assigned clusters grouped together in a block. Finally, with the created blocks and clusters in hand, treatment is randomly assigned at the cluster-level within each block as in any blocked cluster design. The fifth plot of Figure 2 demonstrates this step.

It is important to stress two features of this algorithm: difference *across* clusters is minimized within each block, but difference *within* clusters is randomized across the blocks. For example, Figure 2 shows that each cluster within this block features a young, less educated Democrat and an older, more educated Republican. The goal of blocking is to achieve this similarity at the cluster-level within a block. On the other hand, I do not want all of the conversation partnerships to be between very dissimilar Republicans and Democrats. The individual-level difference between the Republican and Democrat within a partnership varies across blocks. In summary, while the clusters in the block

---

[12]I create these temporary groupings using the `blockTools` statistical software (Moore 2016) with the optimum greedy algorithm and the Mahalanobis distance metric (Moore 2012). Details on the specific variables used for this step are available in Appendix E.

in Figure 2 feature partners who are all very different in the same ways, another block may feature partners who are all very similar.[13]

4.3.2 Chatter conversation software

In addition to addressing several methodological concerns via the blocked, randomized cluster experimental design, I sought to overcome practical concerns that arise when conducting experiments involving social interaction. To do so, I took a novel approach to how the social interaction amongst participants would occur. I designed a software called "Chatter" so participants can have real-time, written conversations online. Full details on Chatter are available in Appendix B. Chatter allows me to emulate a real social experience for the study participants without an existing laboratory or a participant pool. Moreover, Chatter provides me full experimental control—participants are filtered into chatrooms with the partners and treatments pre-assigned via the blocked, randomized cluster design.

5 Data and results

I fielded the experiment between August 15-22, 2019 and January 13-15, 2020 on MTurk with eight separate rounds of data collection. In total, 1,632 unique MTurk Workers took the pre-treatment survey and a subset of 630 were selected via the blocked, randomized cluster experimental design algorithm. Therefore, the design included 630 participants, formed into 315 partnerships/clusters, and nested within 105 blocks.

Despite the quick timeline and reminders outlined above, chosen participants did attrit between the pre-treatment survey and returning for the experiment. And, a single participant's attrition impacts their cluster and block. The sample contains 238 (75.5%) full partnerships and 44 (41.9%) full blocks.[14] In what follows, all *blocks* for a which any participant did not follow up are dropped.[15] This is an advantage of block randomized designs. Dropping entire blocks allows for the loss of clusters without losing balance on blocked covariates across experimental conditions, which can

---

[13]See Appendix D for a visualization of the randomization of individual-level difference within clusters and the minimization of cluster-level differences within blocks.

[14]Appendix G shows details on attrition by round of data collection.

[15]Appendix L shows results for all full partnerships/clusters which are largely consistent with results when including only full blocks.

generate bias (e.g., King et al. 2007). After doing so, the sample used in subsequent analyses contains a total of 264 participants: 44 pairs in the control condition, 44 pairs in the non-political contact condition, and 44 pairs in the political contact condition. Importantly, no participants attrited post-treatment, which could bias results if participants attrited as a function of treatment assignment, such as after seeing they were assigned to talk politics.

5.1   Manipulation check

Before considering the effects of social interaction on outparty affect and stereotypes, it is important to consider if the participants took the exercise seriously and engaged in their assigned type of social interaction. I have read every short essay and conversation transcript, and the participants do indeed engage with each other and discuss their assigned topic.[16]

Additionally, consider the following results regarding engagement and on-topic conversation. A median number of 17.5 messages were sent across non-political conversation and 14 messages across political conversations. The median number of words exchanged were 222, 269, and 299 in the no contact, non-political contact, and political contact conditions, respectively. These summary statistics demonstrate the participants engaged in the exercise. Moreover, participants discussed their assigned topic. The word "gun" appeared 386 times across the conversations assigned to talk about gun control, and not at all in the other experimental conditions. The word "meaning" appeared 394 times in the control condition, 119 times in the non-political contact condition, and not at all in the political contact condition.[17]

## 5.2   Randomization inference hypothesis tests

Given the evidence that the participants engaged in their assigned type of social interaction, I next assess my claims regarding the consequences of social interaction with an outparty member. I test my hypotheses using randomization inference (Fisher 1935). Because my sample size is relatively small, and as I demonstrate below, my outcomes are not distributed normally, I use randomization inference to avoid appealing to large sample approximations or modeling assumptions.[18] Additionally,

---

[16]Appendix H provides an an example from each of the three experimental conditions of on-topic, active participation in the exercise.

[17]Appendix I presents consistent results across additional words and phrases.

[18]Appendix K reports consistent results using parametric hypothesis test approaches.

Figure 3: Density of outcome variable, change in outparty affect, by experimental condition
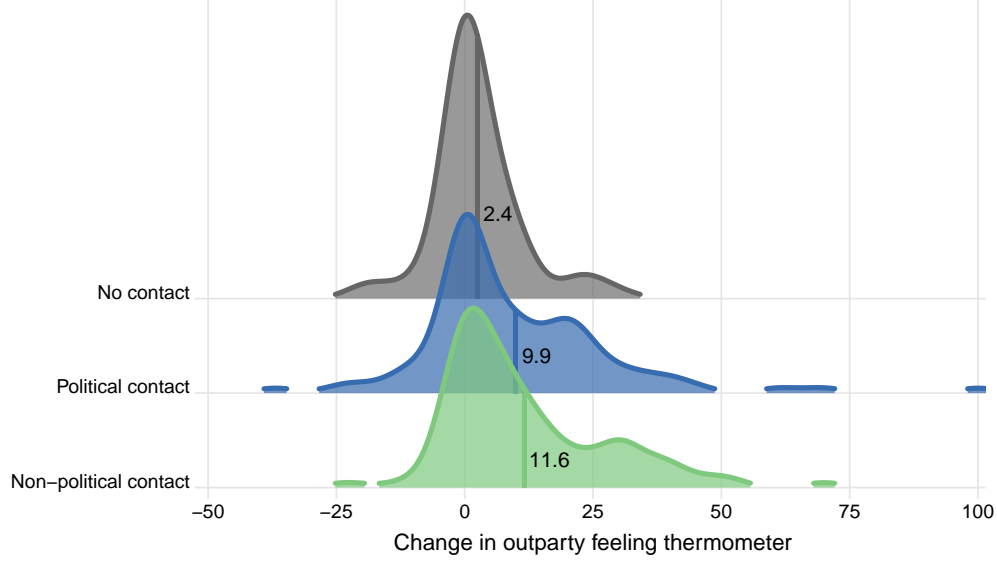
Note: Density of outcome variable, change in outparty affect, by experimental condition considering full blocks. Outcome measured as change in outparty affect measured using pre- and post-treatment ratings of the outparty on the 101-point feeling thermometer. Grey density shows outcome for the no contact condition, blue density shows outcome for the political contact condition, and green density shows outcome for the non-political contact condition. Vertical lines display the mean change for each condition. Change for the no contact condition is tightly concentrated around 0. Change for each contact condition is skewed positively, indicating an improvement, on average, in outparty affect.

randomization inference is straightforward given the blocked cluster randomization procedure I use for assigning conversation partnerships to experimental conditions.

Rather than assume a null hypothesis of regarding the *average* treatment effect, randomization inference depends on a sharp null hypothesis *for all units*. Under the sharp null hypothesis of no effect for all units, I can then state all potential outcomes for all units—the potential outcome for a unit under the control condition is no different than their outcome under treatment. Then, since I know the exact randomization procedure, it is possible to list all permutations of treatment assignment and obtain all hypothetical results of the experiment (using some test statistic) under the sharp null. Given this sampling distribution of the test statistic under the sharp null, I calculate *p*-values to judge if the result of my experiment is unusual (i.e., significant).

I use the suite of tools available in the `DeclareDesign` software to declare my blocked cluster randomization procedure (Blair et al. 2019), and I use the `ri2` software (Coppock 2019) to conduct the randomization inference. Because there are 2^{44} possible random assignments for the following tests, in what follows, I report approximate *p*-values using a random sample of 50,000 random assignments consistent with my design.

Table 2: Significance tests of average treatment effect estimates, outcome change in outparty affect

|	Estimate	p-value
Non-political contact vs. no contact	**9.17**	≈ 0
Political contact vs. no contact	**7.47**	.002
Non-political vs. political contact	1.70	.58

Note: Difference-in-differences estimates of average treatment effect and randomization inference p-values under the sharp null hypothesis of no treatment effect. Results considering full blocks. Dependent variable is individual-level change in outparty affect. Two-sided p-values are reported. Both non-political contact and political contact improved outparty affect relative to no contact.

I first assess contact's effects on outparty affect before turning to contact's effect on outparty trait stereotypes. Figure 3 shows a summary of the change in outparty affect outcome variable. Change is tightly concentrated around zero for the no contact condition, while each of the contact conditions shows considerable positive change in affect. Improvement in outparty affect is notably similar across non-political and political conversations.

I use randomization inference to test my three hypotheses regarding the benefits of contact with an outparty member on outparty affect. Note I consider this a three-armed experiment, so I assess pairwise comparisons between each experimental condition.[19] Table 2 reports the average treatment effect estimate and the p-value associated with a test of this estimate against the sharp null hypothesis of no effect. I use the difference-in-difference estimator for the average treatment effect because my outcome is measuring "change scores"—change in outparty affect from pre-test to post-test. The first row in Table 2 reports results for my first hypothesis—that non-political interparty contact improves negative outparty affect. I find a treatment effect for non-political contact, relative to no contact, of 9.2 degrees. Moreover, the second row in Table 2 reports results for my second hypothesis—that political interparty contact improves negative outparty affect. I also find support for this hypothesis as political contact with a treatment effect of 7.5 degrees. Finally, I assess my third hypothesis—that non-political conversation would be more effective than political conversation at improving outparty affect. Contrary to my expectations, I do not find evidence of non-political conversation being more effective than political conversation at improving intergroup attitudes.

[19]Out of concerns of dropping too many blocks due to attrition, I omitted a no contact and political topic condition. After gaining more experience with the MTurk Workers and scaling the Chatter software, I plan to conduct the full two-by-two factorial design in a replication.

Figure 4: Perceptions of the outparty
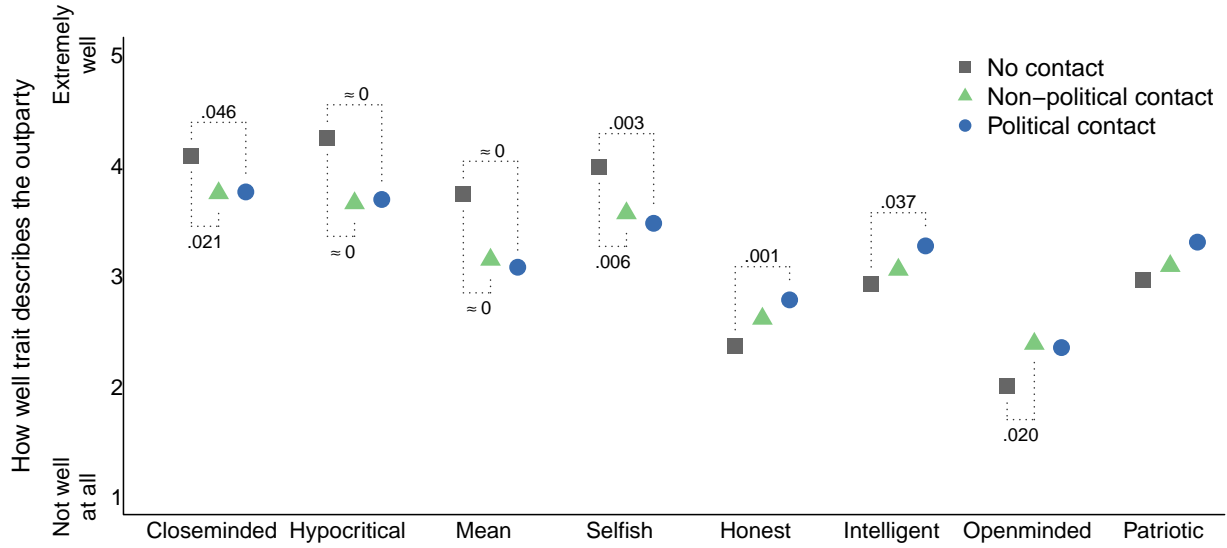
Note: Mean response for each experimental condition for each of the traits surveyed. Randomization inference *p*-values are also reported for significant results ($\alpha = .05$).

To help put these results into context, I summarize participants' post-treatment outparty feeling thermometer ratings relative to meaningful points on the scale (e.g., Levendusky 2018). First, consider the percentage of participants rating the outparty "warmly" after contact, or greater than or equal to 50 on feeling thermometer—44% of those in non-political condition and 38% of those the political condition rated the outparty in this way while only 25% of participants who did not experience interparty contact rated the outparty favorably. Additionally, at the "cold" end of the feeling thermometer, consider a very unfavorable rating of less than or equal to 5. Only 17% of those in the non-political condition and 15% of those in political condition rated outparty in this way, while 28% of those who did not experience outparty contact rated the outparty with such an extremely unfavorable rating.

I've provided evidence that social interaction can alter how partisans feel about the outparty, and I now turn to assess if contact can alter how partisans think about, or perceive, outparty members. The outcomes of interest are respondents' level of agreement (on a five point scale, higher values indicating more agreement) with how well several traits, four negative and four positive, describe supporters of the outparty. I use the difference-in-means estimator to estimate average treatment effects, and I use randomization inference to approximate *p*-values.

Figure 4 plots the mean response of each experimental condition for each of the traits surveyed.

Randomization inference p-values are also reported for significant results ($\alpha = .05$).[20] When respondents had interparty conversation, whether non-political or political, they were less inclined to ascribe the negative traits — closeminded, hypocritical, mean, and selfish — to the outparty. However, there are inconsistent results among the positive traits. Political contact significantly increased perceptions of the outparty as honest and intelligent, and non-political contact increased perceptions of the outparty as openminded. Interestingly, these results differ from Levendusky's (2018) findings that priming the superordinate American identity improves positive trait ratings but does not decrease negative trait ratings. I lack a theoretical expectation for these results. Future research ought to explore these contrasting patterns of outparty prejudice-reduction techniques.

6   Conclusion

Despite the ample evidence documenting a macro-level trend of increasing negativity toward members of the opposite political party, relatively less research has investigated the consequences of micro-level social interactions with outparty members on this trend. This paper addresses this question with an experiment involving actual social interaction among partisans, and shows that interparty conversation can result in a sizable increase in outparty affect and a disinclination to describe outparty members by negative stereotypes. In doing so, this research addresses a concern in the literature to appropriately characterize the extent to which Americans are actually "affectively polarized"—increasingly liking their inparty and disliking the outparty—and what consequences this type of polarization has. While this research characterizes Americans' views of outparty as biased, overgeneralized, and potentially inaccurate, it also illuminates the limits of our partisan identities by showing conversation has the power to break down inparty/outparty categories and correct for heightened outparty negativity.

While I've found that social interactions *can* improve how partisans feel about and think about outparty members, a question left for future research is *when* social interactions improve affect and perceptions outside of the environment constructed for this research. For example, this experiment featured online conversation, limiting the external validity of these findings as applied to in-person interactions where physical appearance and body language are additional guides to social interaction

---

[20]The table of full results is reported in Appendix J.

that may influence how it unfolds and what effects it has on subsequent outcomes.

Moreover, this research does not reflect the role self-selection plays in this process of social interaction, as partisans may try to choose when and when not to engage in social interactions with outparty members. Research shows that anticipating political discussion makes people anxious (Carlson and Settle 2016). It follows that people prefer to avoid political discussion, especially when it is disagreeable (Gerber et al. 2012) or with an outparty member (Settle and Carlson 2019).

But, while research finds some people prefer avoid political conversation or social interaction with outparty members, sometimes these interactions occur beyond our control. Walsh contends that "Much political interaction occurs not among people who make a point to specifically talk about politics but emerges instead from the social processes of people chatting with one another" (pg. 35 Cramer Walsh 2004). Moreover, a recent large-scale, full-network study supports the idea that talking politics is more an *incidental* than it is a *purposive* exercise (Minozzi et al. 2019). If we take the incidental model of political discussion seriously, then talking politics is often unanticipated, it is hard to avoid altogether, and everyone is subject to experiencing some political talk in their daily lives, and this research begins illuminating the consequences of it in regard to how we feel and think about opposing party members.

Another limitation of this experimental design is that it involved only two individuals, one from each party. While this helps satisfy one of Allport's conditions for contact to improve outgroup prejudice—equal status in the contact situation—not all social interactions will avoid having a minority group or minority opinion apparent in the interaction. This is an important consideration because research shows that when politics arises in a discussion, people tend shield their own views and conform to majority opinion (Carlson and Settle 2016). Related to conformity within social interactions, this research does not account for the role social sanctioning may play in political and non-political interactions that occur in Americans' everyday lives. It is left for future research to speak to how different group compositions and pre-existing relationships may impact the effectiveness of social interaction as a strategy for combating negative intergroup attitudes.

Finally, this research theorizes about and studies positive contact. However, research shows that a single negative instance of contact has a stronger, negative effect on affect than does any single instance of positive contact (Barlow et al. 2012). Therefore, future research should consider under what conditions partisan social interactions go awry, such as when a partisan feels like their group

or "team" is being threatened. While this research studied the topic of gun control, electoral politics or issues appealing to more deeply held values may activate stronger emotions and lead partisans to use social interactions to fortify the walls of their inparty/outparty categories.

References

Ahler, Douglas J and Gaurav Sood. 2018. "The parties in our heads: Misperceptions about party composition and their consequences." *The Journal of Politics* 80(3):964–981.

Alford, John R and John R Hibbing. 2007. "Personal, interpersonal, and political temperaments." *The Annals of the American Academy of Political and Social Science* 614(1):196–212.

Allport, Gordon W. 1954. *The nature of prejudice.* Reading, MA: Addison Wesley Publishing Company.

Barlow, Fiona Kate, Stefania Paolini, Anne Pedersen, Matthew J Hornsey, Helena RM Radke, Jake Harwood, Mark Rubin and Chris G Sibley. 2012. "The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice." *Personality and Social Psychology Bulletin* 38(12):1629–1643.

Baron-Cohen, Simon, Jennifer Richler, Dheraj Bisarya, Nhishanth Gurunathan and Sally Wheelwright. 2003. "The Systemizing Quotient: An Investigation of Adults with Asperger Syndrome or High–Functioning Autism, and Normal Sex Differences." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358(1430):361–374.

Berelson, Bernard R, Paul F Lazarsfeld, William N McPhee and William N McPhee. 1954. *Voting: A study of opinion formation in a presidential campaign.* University of Chicago Press.

Bettencourt, B Ann, Marilynn B Brewer, Marian Rogers Croak and Norman Miller. 1992. "Cooperation and the reduction of intergroup bias: The role of reward structure and social orientation." *Journal of experimental social psychology* 28(4):301–319.

Blair, Graeme, Jasper Cooper, Alexander Coppock and Macartan Humphreys. 2019. "Declaring and diagnosing research designs." *American Political Science Review* 113(3):838–859.

Blair, Graeme, Jasper Cooper Alexander Coppock Macartan Humphreys Luke Sonnet. 2019. *estimatr: Fast Estimators for Design-Based Inference.* R package version 0.20.0. https://CRAN.R-project.org/package=estimatr.

Blair, Graeme, Jasper Cooper Alexander Coppock Macartan Humphreys and Neal Fultz. 2019. *DeclareDesign.* R package version 0.20.0. https://CRAN.R-project.org/package=DeclareDesign.

Bond, Robert M, Hillary C Shulman and Michael Gilbert. 2018. "Does Having a Political Discussion Help or Hurt Intergroup Perceptions? Drawing Guidance From Social Identity Theory and the Contact Hypothesis." *International Journal of Communication* 12:21.

Brewer, Marilynn B and Norman Miller. 1988. Contact and cooperation. In *Eliminating racism.* Springer pp. 315–326.

Brewer, Marilynn B. and Norman S. Miller. 1984. Beyond the contact hypothesis: Theoretical perspectives on desegregation. In *Groups in contact: The psychology of desegregation*, ed. Norman S. Miller and Marilynn B. Brewer. New York: Academic Press pp. 281–302.

Carlin, Ryan E and Gregory J Love. 2013. "The politics of interpersonal trust and reciprocity: An

experimental approach." *Political Behavior* 35(1):43–63.

Carlson, Taylor N. 2019. "Through the Grapevine: Informational Consequences of Interpersonal Political Communication." *American Political Science Review* 113(2):325–339.

Carlson, Taylor N and Jaime E Settle. 2016. "Political chameleons: An exploration of conformity in political discussions." *Political Behavior* 38(4):817–859.

Coppock, Alexander. 2019. *ri2.* R package version 0.1.2. https://CRAN.R-project.org/package=ri2.

Cornfield, Jerome. 1978. "Symposium on CHD prevention trials: design issues in testing life style intervention: randomization by group: a formal analysis." *American journal of epidemiology* 108(2):100–102.

Costa, Paul T and Robert R McCrae. 2008. "The revised neo personality inventory (neo-pi-r)." *The SAGE handbook of personality theory and assessment* 2(2):179–198.

Cramer Walsh, Katherine. 2004. *Talking about politics: Informal groups and social identity in American life.* University of Chicago Press.

Dovidio, John F, Samuel L Gaertner and Kerry Kawakami. 2003. "Intergroup contact: The past, present, and the future." *Group Processes & Intergroup Relations* 6(1):5–21.

Druckman, James N and Kjersten R Nelson. 2003. "Framing and deliberation: How citizens' conversations limit elite influence." *American Journal of Political Science* 47(4):729–745.

Druckman, James N and Matthew S Levendusky. 2019. "What do we measure when we measure affective polarization?" *Public Opinion Quarterly* 83(1):114–122.

Druckman, James N, Matthew S Levendusky and Audrey McLain. 2017. "No Need to Watch: How the Effects of Partisan Media Can Spread via Interpersonal Discussions." *American Journal of Political Science* .

Ensari, Nurcan, Julie Christian, Dawn Matsui Kuriyama and Norman Miller. 2012. "The personalization model revisited: An experimental investigation of the role of five personalization-based strategies on prejudice reduction." *Group Processes & Intergroup Relations* 15(4):503–522.

Fang, Marina. 2017. "Obama Uses To Kill A Mockingbird To Remind Americans Of The Importance Of Empathy." *Huffington Post.* January 10. https://www.huffpost.com/entry/obama-farewell-address-to-kill-a-mockingbird_n_58759621e4b03c8a02d3fa49. Last accessed on February 1, 2020.

Fisher, R. A. 1935. *Design of experiments.* Oliver and Boyd.

Gerber, Alan S, Gregory A Huber, David Doherty and Conor M Dowling. 2012. "Disagreement and the avoidance of political discussion: Aggregate relationships and differences across personality traits." *American Journal of Political Science* 56(4):849–874.

Green, Donald P, Bradley Palmquist and Eric Schickler. 2004. *Partisan hearts and minds: Political parties and the social identities of voters.* Yale University Press.

Greene, Steven. 1999. "Understanding party identification: A social identity approach." *Political*

Psychology 20(2):393–403.

Grumet, Jason. 2019. "Why American families must talk about politics this Thanksgiving." *The Hill.* November 27. https://thehill.com/opinion/white-house/472258-why-american-families-must-talk-about-politics-this-thanksgiving. Last accessed on February 1, 2020.

Hetherington, Marc and Jonathan Weiler. 2018. *Prius Or Pickup?: How the Answers to Four Simple Questions Explain America's Great Divide.* Houghton Mifflin.

Hogg, Michael A and Scott A Reid. 2006. "Social identity, self-categorization, and the communication of group norms." *Communication theory* 16(1):7–30.

Huckfeldt, Robert and John Sprague. 1991. "Discussant effects on vote choice: Intimacy, structure, and interdependence." *The Journal of Politics* 53(1):122–158.

Imai, Kosuke, Gary King, Clayton Nall et al. 2009. "The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation." *Statistical Science* 24(1):29–53.

Iyengar, Shanto, Gaurav Sood and Yphtach Lelkes. 2012. "Affect, not ideology: a social identity perspective on polarization." *Public opinion quarterly* 76(3):405–431.

Iyengar, Shanto and Sean J Westwood. 2015. "Fear and loathing across party lines: New evidence on group polarization." *American Journal of Political Science* 59(3):690–707.

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

Karpowitz, Christopher F and Tali Mendelberg. 2014. *The silent sex: Gender, deliberation, and institutions.* Princeton University Press.

Karpowitz, Christopher F, Tali Mendelberg and Lee Shaker. 2012. "Gender inequality in deliberative participation." *American Political Science Review* 106(3):533–547.

King, Gary, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T Moore, Jason Lakin, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila and Héctor Hernández Llamas. 2007. "A "politically robust" experimental design for public policy evaluation, with application to the Mexican universal health insurance program." *Journal of Policy Analysis and Management* 26(3):479–506.

Klar, Samara. 2014. "Partisanship in a social setting." *American Journal of Political Science* 58(3):687–704.

Levendusky, Matthew. 2009. *The partisan sort: How liberals became Democrats and conservatives became Republicans.* University of Chicago Press.

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

Levendusky, Matthew S and Neil Malhotra. 2015. "(Mis) perceptions of partisan polarization in the American public." *Public Opinion Quarterly* 80(S1):378–391.

Mason, Lilliana. 2015. ""I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization." *American Journal of Political Science* 59(1):128–145.

Mason, Lilliana. 2018. *Uncivil agreement: How politics became our identity.* University of Chicago Press.

McClurg, Scott D. 2003. "Social networks and political participation: The role of social interaction in explaining political participation." *Political research quarterly* 56(4):449–464.

Miller, Norman. 2002. "Personalization and the promise of contact theory." *Journal of Social Issues* 58(2):387–410.

Miller, Norman, Marilynn B Brewer and Keith Edwards. 1985. "Cooperative interaction in desegregated settings: A laboratory analogue." *Journal of Social Issues* 41(3):63–79.

Minozzi, William, Hyunjin Song, David MJ Lazer, Michael A Neblo and Katherine Ognyanova. 2019. "The Incidental Pundit: Who Talks Politics with Whom, and Why?" *American Journal of Political Science* .

Moore, Ryan T. 2012. "Multivariate continuous blocking to improve political science experiments." *Political Analysis* 20(4):460–479.

Moore, Ryan T. 2016. *blockTools: Blocking, Assignment, and Diagnosing Interference in Randomized Experiments.* R package version 0.6-3. https://CRAN.R-project.org/package=blockTools.

Mutz, Diana C. 2002. "Cross-cutting social networks: Testing democratic theory in practice." *American Political Science Review* 96(1):111–126.

Mutz, Diana C. 2006. *Hearing the other side: Deliberative versus participatory democracy.* Cambridge University Press.

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

Pettigrew, Thomas F and Linda R Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of personality and social psychology* 90(5):751.

Pew Research Center. 2017. "America's Complex Relationship with Guns." June 22. https://www.pewsocialtrends.org/2017/06/22/americas-complex-relationship-with-guns/.

Pew Research Center. 2018. "Where Americans Find Meaning in Life." November 20. http://www.pewforum.org/2018/11/20/where-americans-find-meaning-in-life/.

Pew Research Center. 2019. "6 facts about U.S. political independents." May 15. https://www.pewresearch.org/fact-tank/2019/05/15/facts-about-us-political-independents/.

Sedikides, Constantine, W Keith Campbell, Glenn D Reader and Andrew J Elliot. 1999. "The relationship closeness induction task." *Representative Research in Social Psychology* 23:1–4.

Settle, Jaime E and Taylor N Carlson. 2019. "Opting out of political discussions." *Political Communication* pp. 1–21.

Sinclair, Betsy. 2012. *The social citizen: Peer networks and political behavior.* University of Chicago Press.

Tajfel, Henri. 1981. *Human groups and social categories: Studies in social psychology.* CUP Archive.

Tajfel, Henri and John C Turner. 1979. "An integrative theory of intergroup conflict." *The social psychology of intergroup relations* 33(47):74.

Tu, Yanping, Alex Shaw and Ayelet Fishbach. 2015. "The friendly taking effect: How interpersonal closeness leads to seemingly selfish yet jointly maximizing choice." *Journal of Consumer Research* 42(5):669–687.

Turner, John C, Michael A Hogg, Penelope J Oakes, Stephen D Reicher and Margaret S Wetherell. 1987. *Rediscovering the social group: A self-categorization theory.* Basil Blackwell.

Westfall, Jacob, Leaf Van Boven, John R Chambers and Charles M Judd. 2015. "Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide." *Perspectives on Psychological Science* 10(2):145–158.

Appendices

A Methodological and practical concerns in the experimental study of social interaction

Political scientists often use experimental research in which participants have actual social interaction to test their hypotheses (e.g., Druckman and Nelson 2003; Karpowitz, Mendelberg and Shaker 2012; Klar 2014).[21] Lab experiments across this broad range of research share a set of methodological and practical concerns. In this section, I outline several of these concerns after first walking through the setup of a typical lab experiment in this field.

The typical design for a lab experiment involving social interaction is the following. The researcher has a list of participants, maybe the students in their department's Introduction to American Politics class. If the experiment requires small group discussion of four participants per group, the researcher randomly assigns participants into groups of four.[22] Or, the experiment might involve a certain group composition, such as discussion amongst opposing party members. If so, the groups are randomly assigned with respect to this constraint. After random group assignment, treatment is assigned at the group-level. For example, the treatment might be whether or not the group engages in discussion (e.g., Klar 2014) or the decision rule governing the group's decision-making (e.g., Karpowitz, Mendelberg and Shaker 2012). Finally, random assignment is typically used to assign groups to experimental conditions.

I discuss four main, interrelated methodological concerns that arise with experiments involving social interaction: sample size, balance across experimental conditions, and unit interference, and attrition. First, these studies often rely on relatively small sample sizes (e.g., 277 in Klar (2014) or 261 in Druckman and Nelson (2003)). Small sample sizes impacts the power of tests, decreasing the probability a researcher can reject the null hypothesis of no treatment effect. Additionally, while simple random assignment or complete random assignment of the treatment are straightforward, they can introduce inefficiencies, especially when coupled with small sample sizes. Because the sample size for these studies is usually small, a researcher could end up with an unlucky, unbalanced randomization. What's more, balance across experimental conditions is not just relevant at the individual level, but it is relevant at the group level, as well. For example, for an experiment involving two individuals engaging in conversation, it may be problematic if all *individuals* who are strong Republicans were, by chance, assigned to the same experimental condition. Additionally, it may be problematic if all *groups* who are ideologically-similar were, by chance, assigned to the same experimental condition. Block randomizing treatment assignment can ensured balanced

---

[21]Beyond the study of conversation, lab experiments in political science featuring constructed interaction amongst participants take the form of trust and ultimatum games, for example (e.g., Alford and Hibbing 2007; Carlin and Love 2013).

[22]Lab experiments involving participant discussion require a great deal of effort and coordination by the researcher and participants. Alternatively, to increase the number of completed discussion groups in the experiment, the researcher may construct groups based on which participants are available and present at any given time, thus without randomization.

randomization and improve inefficiencies (Moore 2012). Relatedly, this class of experiments inherently involves social influence, and therefore interference or spillover, within the assigned groups. Treatment, therefore, ought to be applied at the group level, introducing an added complexity to the experimental design and estimation of treatment effects. Randomizing groups, rather than individuals, also reduces efficiency in this class of experiments (Cornfield 1978).

As a practical matter, studies involving participant interaction are resource-intensive, often prohibitively so, largely requiring an academic lab and existing subject pool. Even with these resources in place, it is difficult to then coordinate participants into complete discussion groups. Therefore resources for hosting social interaction and coordinating the participants, all the while maintaining sample size, are practical concerns of experiments involving social interaction.

B   Chatter conversation software

As a practical matter, in order to study interparty conversations, I designed a software called "Chatter" by which participants can have real-time, written conversations online.[23] Chatter allows 2-10 participants to have a conversation via an interface similar to many messaging apps like iMessage or WhatsApp.

Figure 5 shows an example of Chatter's user interface. Participants see instructions for the conversation above a box containing previously sent messages. A participant's own messages appear unlabeled on the right in blue and other conversation members' messages appear, labeled with a unique identifier, on the left. Participants also see a timer that counts down the time remaining and a "Done" button which illuminates and activates when time expires.

Several features of Chatter facilitate experimentation involving conversations. First, Chatter allows the researcher to set up chatrooms so that the partnerships or groups created using the blocked, randomized cluster design can engage in social interaction. Second, Chatter allows the researcher to customize the conversational instructions shown to each participant. Specifically, in the experiment I explain below, treatment takes the form of the topic at the cluster-level, so conversation partners see the same instructions. But, each participant's instructions are customized to include the partisan identity of their partner. Third, the researcher controls all additional chatroom and user settings, including what usernames are displayed (e.g., random sequence of letters, the same name to control perception of gender, etc.) and how long the conversation should last. Lastly, I use Chatter coupled with the Qualtrics survey software for survey administration and Amazon Mechanical Turk to recruit participants. However, experimentation using Chatter is generalizable to other survey platforms and other participant pools.

Chatter allows researchers to relatively easily emulate a real social experience without an existing laboratory or a participant pool. Moreover, as a software for *online* conversations, Chatter allows for diversity in the participant pool that is hard to come by when using in-person conversations, usually with students, faculty, and staff that are affiliated with a college campus (e.g., Karpowitz and

---

[23]Chatter is a Ruby on Rails application backed by a Postgres database, deployed on Heroku.

Figure 5: Chatter user interface

Note: Chatter user interface. Instructions appear at the top of the page. Akin to other messaging software, an individual's own messages appear on the right. Other users' messages appear on the left. When the timer indicates no time is left, the "Done" button activates and redirects users to a post-conversation survey when clicked.

Mendelberg 2014; Klar 2014) or that can be recruited within a few cities (e.g. Druckman, Levendusky and McLain 2017). A final practical advantage of Chatter is that a researcher can quickly conduct a large-n study involving conversations. Chatter allows for hundreds of conversations to happen simultaneously, which is difficult to achieve in the setting of an academic laboratory.

C Simulation study for proposed blocked, randomized cluster design

I ran a simple simulation to demonstrate the benefits of the blocked, randomized cluster design as opposed to what I will call a "naive" design. The goal of the simulation is to construct clusters and randomly assign treatment at the cluster-level via the blocked, randomized cluster design and the naive design, and to compare the designs' estimation efficiency and hypothesis test power.

Figure 6: Efficiency and power of proposed and naive experimental designs
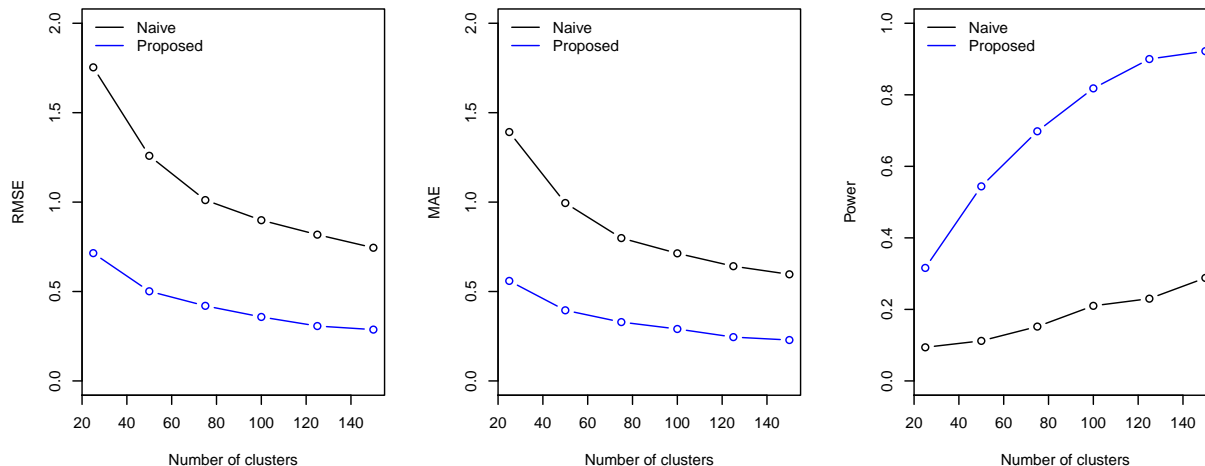
Note: Diagnosands of interest for the proposed blocked, randomized cluster design and the naive design, while varying sample size. Root mean squared error (RMSE) is reported in the first plot, mean absolute error (MAE) is reported in the second plot, and power of randomization inference hypothesis tests is reported in the third plot. Note that the proposed design is more efficient (in terms of RMSE and MAE) and has better power than the naive design.

The naive design creates clusters by (1) randomly choosing a participant who does not already have an assigned partner and (2) randomly choosing a partner of the opposite partisan identification from the remaining un-partnered participants. Then, using simple random assignment, the naive design assigns the cluster either to receive treatment or control with equal probability.

For this simple example, the population is of size $N = 500$, each individual is assigned a partisan identification label, and three variables impact the outcome: $X_1 \sim N(0,1)$, $X_2 \sim Unif(0,1)$, and $X_3 \sim \chi_2^2$. There's both individual-level error $u_i \sim N(0,1)$ and cluster-level error $u_c \sim N(0,1)$. The average treatment effect is 1.

Potential outcomes are a function of these variables: $Y_i = \text{ATE} * Z + X_1 + 2 * X_2 + 3 * X_3 + u_i + u_c$. My estimand is the average treatment effect. Importantly, I vary the number of clusters I sample from the population of fixed clusters from 25 to 150 clusters for the naive design. To vary the sample size in a comparable way for the blocked, randomized cluster design, I sample half as many fixed blocks. I use the difference in means estimate, and calculate *p*-values for hypothesis tests using randomization inference where I can specify the exact randomization procedure used for the blocked, randomized cluster design or the naive design.

I use the `DeclareDesign` and `ri2` R packages (Blair and Fultz 2019; Coppock 2019) to conduct this simulation study. I conduct 500 simulations of the design for each sample size. I assess the performance of the two designs with three diagnosands: root mean squared error (RMSE), mean absolute error (MAE), and power with $\alpha = .05$. Figure 6 shows the results. The proposed blocked, cluster design improves efficiency by having a lower RMSE and MAE across the entire range of sample sizes, and particularly for smaller sample sizes. Additionally, the proposed experimental design rapidly improves the power of randomization inference hypothesis tests in this simulation as the sample size increases.

D Blocked, randomized cluster algorithm and partnership-level and individual-level differences

Figure 7 demonstrates two important features of the blocked, randomized cluster algorithm by adding a second block of clusters. First we see that cluster-level difference is minimized within each block. However, we also see that individual-level difference within partnerships is randomized across blocks. The first block shows Democrats and Republicans that are very different (in the same ways), but the second block shows Democrats and Republicans that are very similar (in the same ways).

Figure 7: Partnership-level difference minimized within blocks and individual-level difference randomized across blocks
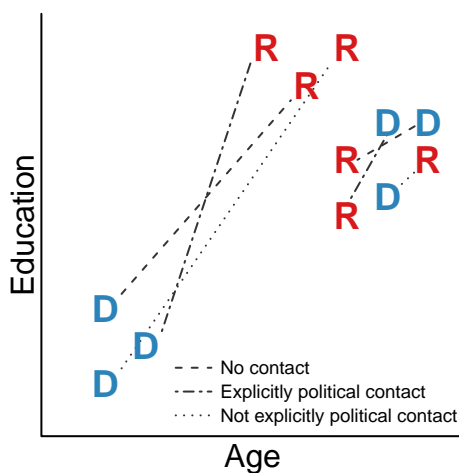
E Details on blocked cluster design for experiment

The following provides more specific, technical details for the block, randomized cluster algorithm used in the experiment, particularly for the blocking variables.

- Create trios of similar individuals within the same party

  - Using the `blockTools::block()` statistical software (Moore 2016) with the optimum greedy algorithm and the Mahalanobis distance metric (Moore 2012). Note is step does not create the blocks *in full*, despite the use of the `blockTools` statistical software. See Section 4.3.

  - Blocking variables are age and education (considered continuous); indicators for gender* and ethnicity, stength of partisan identification*; pre-treatment thermometer ratings of inparty, outparty, and President Trump*; a 6-item battery on pre-treatment gun views (considered continuous); a single pre-treatment item asking overall view regarding gun control*; personality trait estimates including 4 item adaptive versions of each Big 5

Trait (openness*, conscientiousness, agreeableness, neuroticism, and extraversion) (Costa and McCrae 2008), Systemizing and Empathizing Quotients (Baron-Cohen et al. 2003); and latitude and longitude. Variables marked with an ∗ are up-weighted to have twice the weight of the other variables.

– Block within subgroups of partisan identification. Those who indicated "Independent" or "other" for partisan identification are collapsed into the respective party toward which they lean.

- Shuffle individuals within trios.
- Simultaneously create blocks and interparty pairs by randomly assigning one Democratic trios to each Republican trio.
- Within each block of interparty pairs, randomly assign treatment at interparty pair-level.

F   MTurk HIT payment details

All participants who completed the pre-treatment survey were compensated \$1, regardless of whether they expressed interest in returning for the follow-up task.

Those that returned for the follow-up task (including the conversation or short essay and the post-treatment survey) were compensated an additional amount. If their partner joined and they finished the post-treatment survey, they were compensated \$2 + \$1 bonus for engaging in the writing task. If the participant waited for five minutes but their partner did not join (thus they could not complete the task), they were still compensated the \$2.

G   Attrition details

This experiment required multiple rounds of data collection. Early rounds of data collection on MTurk had smaller sample sizes as I assessed the availability of MTurk Workers to immediately complete my posted HITs their willingness to return for the conversation task. I also assessed the scalability of Chatter in these early rounds, scaling up in later rounds of data collection.

Table 3 shows sample sizes across rounds of data collection as well as details on attrition. About 75% of full partnerships completing the task, and about 42% of full blocks completing the task.

H   Example short essays and conversations

Table 3: Sample size and attrition details across eight rounds of data collection

Round	Pre-test Participants (N)	Invited to Return (N)	Full Pairs (%)	Full Blocks (%)
1	124	42	85.7%	71.4%
2	138	30	73.3%	40.0%
3	146	54	63.0%	22.2%
4	174	78	71.8%	30.8%
5	249	84	76.2%	42.9%
6	270	102	74.5%	35.3%
7	311	132	83.3%	59.1%
8	220	108	72.2%	33.3%
	1632	630	75.6%	41.9%

Table 4: Example short essays, conversations for experimental conditions

No contact (control)	Non-political contact	Political contact

No contact (control)

u1 I think that feeling content in where you are is really one of the deepest meanings in life.

Having a beautiful connection with someone you love makes life feel so much more worth it.

Being able to connect spiritually and find your inner peace is a great goal.

Finding someone who is your soul match and a willing partner is the ultimate meaning of life.

Not letting money or materialistic things rule you, but instead letting those things be a byproduct of what you love.

Finding your true family, whether blood or not is a great goal in life.

Realizing that sometimes you have to let go and let God is a great goal.

Making sure your heart and your spirit are in a place of love is incredibly important in life.

Not being around anyone who takes away your joy and your light is an important goal in life.

All of these things together are pertinent to reaching your full potential aka the meaning of your life.

There is no right and wrong answer to what the exact meaning is, but you have to look within yourself to see what your soul and spirit deem important.

The true meaning of life is to feel fully content and at peace with your mind, body and spirit.

The true meaning of life is love.

---

u2 For me life is happiness and love. Happiness is when you are with the people you cared for and love. I think life will be dull or boring if we don't have those people that we will cherish the most. They are the one who makes the ride worthwhile.

Life also is giving to others. It is sharing what you have and not asking in return.

Life is when you cry and laugh the most. Life is when you are hurt but prefer to stand up.

Life is still fighting despite the hardship.. Life is loving all the positive things. Life is protecting our mother nature.

Life is everything.

Non-political contact

u3 What do you think about the meaning of life

u4 I think that the meaning of life is very simply to be good people and to spread love and positivity to others.

u3 I think that the mean of life is to enjoy every second we live on this earth. I think it's important to notice the beauty in the simple things.

u4 I definitely agree with that, especially because we have no idea when our life is going to end and if there is something after this life.

u4 We definitely don't appreciate the small things in life, we tend to take them for granted in my opinion.

u3 EXACTLY. We don't know when our life will end so we have to make everyday importwand not to get hung up on mistakes

u4 Yep, and I think it's human nature to be honest.

u4 We're humans and we tend to only focus on ourselves at times.

u4 I know I sometimes blow things way out of proportion with regards to things going wrong in my life. But in reality, my problems really aren't that big of a deal.

u4 Many people worldwide are way less off than me which is why I try to appreciate everything that I have.

u4 It's difficult at times though, sometimes little things add up and can stress you out and you forget to put that type of stuff into perspective.

u3 Yes I agree

u3 I think that people are so focused on the long term they don't think about the here and now.

u3 I'll admit I'm one of them

u4 I agree. I mean it's good to focus on the future and prepare yourself but I think it's also important to balance the future with the present as well.

u4 I see so many people setting themselves up for future success but in the present they're miserable, working a ton of hours at their jobs and hating their lives because they have no free time.

u3 Exactly. That's why I decided to travel and make the most of my youth

u4 What I try to do is enjoy every single day. I try to find minor things that make me happy such as going out to a restaurant or playing a fun game on my phone.

u3 Perfect example. Little walks down a trail make me happy

u4 I love walking around a lake near my house as well, it's very relaxing.

u4 Lets me ease some of my stress when I'm struggling mentally.

u3 Absolutely. The meaning of life is to bring yourself and the ones around you happiness in my book.

u4 Me too. I had a great time chatting with you. Have a great day!

Political contact

u5 Hi, I am ready when you are. thank you.

u6 Hi there! I just think that something so importat is tough to discuss

u6 Especially with a stranger

u5 My thoughts on gun control are that there are plenty of laws we just need to follow them although maybe some gun types just aren't needed in the general public.

u5 such as assault rifles i mean

u6 I agree, some things are not meant to be had by everyone

u6 yea AK47s are superflous

u6 but some people just like the power they get from owning them

u5 how do you feel about background checks at gun shows...I thought they had those and then someone told me no they do not

u6 also you can have a garage sale and sell your guns, there are many loopholes!

u6 guns are in our culture and will be there in the future

u5 yes...or just give them away...I have known that to happen. I don't know how they will regulate that so yes guns are here to stay no doubt

u5 Maybe we can work harder on making some areas safe.

u6 i am not antigun but have a hard time accepting assault rifles

u5 I know gun enthusiasts that like their assault rifles very much. I don't know how we are going to keep guns out of the hands of the mentally ill...that is a grey area to regulate

u6 sadly people have an 'all or nothing' attitude, so it is hard to compromise

u5 I saw we first expend the money and energy to make sure all the existing laws are followed and see where that leaves us.

u5 sorry say not saw..oops

u6 no prob!! lol,,,,at least it is nice to have a conversation with someone who gets other people's point!

u6 i wish we had more of that

u5 our time is about up....have a wonderful rest of the day and nice talking with you

u6 same!! have a great one!

u6 on to the nest HIT lol :)

I Manipulation check

Table 5 details how many times the following terms appear across all short essays or conversations for each experimental condition. Phrases such as "family," "faith," and "happiness" occur often, as expected, when talking about the meaning of life, and do not occur when participants are asked to discuss gun control. Likewise, "gun," "shooting," and "background check" occur often when discussing gun control but not when participants were asked to discuss the meaning of life.

Table 5: Topic-specific word counts suggesting participants discussed assigned topic

	gun	shooting	background check	family	faith	meaning	happiness
No contact	0	1	0	111	14	394	42
Non-political contact	0	0	0	105	7	119	18
Political contact	386	38	38	13	0	0	0

J Full outparty trait results

Table 6 provides full results corresponding to Figure 4 in the main body of the paper. The table reports randomization inference hypothesis tests with the difference in means test statistic under the sharp null hypothesis of no treatment effect. The dependent variable for each test is a measure of agreement with how well the trait describes supporters of the outparty using a five point scale (higher values indicating more agreement). I report two-sided p-values. Recall these results consider only all full blocks.

Table 6: Significance tests of average treatment effect estimates of outparty stereotypes, with full blocks

| | "Closeminded" | | "Hypocritical" | | "Mean" | | "Selfish" | |
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Non-political contact vs. no contact	**-.34**	.021	**-.59**	≈0	**-.59**	≈0	**-.42**	.006
Political contact vs. no contact	**-.33**	.046	**-.56**	≈0	**-.66**	≈0	**-.51**	.003
Non-political contact vs. political contact	-.01	≈1	-.03	.878	.07	.737	.09	.529
	"Honest"		"Intelligent"		"Openminded"		"Patriotic"	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Non-political contact vs. no contact	.25	.155	.13	.490	**.39**	.020	.13	.444
Political contact vs. no contact	**.42**	.001	**.34**	.037	.35	.053	.34	.063
Non-political contact vs. political contact	-.17	.302	-.22	.201	.03	.890	-.216	.210

Note: Results considering all full blocks. Randomization inference hypothesis tests with the difference in means test statistic under the sharp null hypothesis of no treatment effect. Dependent variable is agreement, on a five point scale with higher values indicating more agreement, with how well the trait describes supporters of the outparty. Two-sided p-values are reported.

K   Robustness with other hypothesis test approaches

As a robustness check, I test my hypotheses by calculating p-values by using standard errors and t−values for matched-pair clustered designs (Blair 2019; Imai et al. 2009). Results are consistent with randomization inference results presented in the main body of the paper.

Specifically, Table 7 reports results from a parametric tests that are consistent with the results in Table 2. Also, Table 8 reports results from a parametric tests that are consistent with the results in Figure 4 in the main body of the paper and Table 6 in the appendix.

Table 7: Parametric significance tests of average treatment effect estimates, outcome change in outparty affect

	Estimate	Std. error	t-value	p-value	Conf. interval	Df
Non-political contact vs. no contact	**9.17**	1.80	5.08	≈0	[5.53, 12.81]	43
Political contact vs. no contact	**7.47**	2.40	3.11	.003	[2.62, 12.31]	43
Non-political vs. political contact	1.70	2.85	.60	.55	[-4.04, 7.45]	43

Note: Results considering all full blocks. Dependent variable is individual-level change in outparty affect. Two-sided p-values are reported. Both non-political contact and political contact improved outparty affect relative to no contact.

Table 8: Parametric significance tests of average treatment effect estimates, trait ratings outcomes

	Estimate	Std. error	t-value	p-value	Conf. interval	Df
Outcome: "Closeminded"						
Non-political contact vs. no contact	**-0.34**	0.14	-2.48	0.017	[-0.62, -0.06]	43
Political contact vs. no contact	**-0.33**	0.15	-2.15	0.037	[-0.64, -0.02]	43
Non-political vs. political contact	-0.01	0.15	-0.07	0.941	[-0.32, 0.3]	43
Outcome: "Hypocritical"						
Non-political contact vs. no contact	**-0.59**	0.12	-4.8	≈ 0	[-0.84, -0.34]	43
Political contact vs. no contact	**-0.56**	0.12	-4.79	≈ 0	[-0.79, -0.32]	43
Non-political vs. political contact	-0.03	0.15	-0.23	0.819	[-0.33, 0.27]	43
Outcome: "Mean"						
Non-political contact vs. no contact	**-0.59**	0.14	-4.31	≈ 0	[-0.87, -0.31]	43
Political contact vs. no contact	**-0.66**	0.16	-4.04	≈ 0	[-0.99, -0.33]	43
Non-political vs. political contact	0.07	0.17	0.41	0.687	[-0.27, 0.41]	43
Outcome: "Selfish"						
Non-political contact vs. no contact	**-0.42**	0.14	-2.99	0.005	[-0.7, -0.14]	43
Political contact vs. no contact	**-0.51**	0.16	-3.24	0.002	[-0.83, -0.19]	43
Non-political vs. political contact	0.09	0.15	0.61	0.543	[-0.21, 0.39]	43
Outcome: "Honest"						
Non-political contact vs. no contact	0.25	0.17	1.51	0.138	[-0.08 , 0.58]	43
Political contact vs. no contact	**0.42**	0.15	2.79	0.008	[0.12, 0.72]	43
Non-political vs. political contact	-0.17	0.15	-1.11	0.272	[-0.48, 0.14]	43
Outcome: "Intelligent"						
Non-political contact vs. no contact	0.12	0.16	0.76	0.449	[-0.2, 0.45]	43
Political contact vs. no contact	**0.34**	0.15	2.23	0.031	[0.03, 0.65]	43
Non-political vs. political contact	-0.22	0.16	-1.37	0.176	[-0.53, 0.1]	43
Outcome: "Openminded"						
Non-political contact vs. no contact	**0.39**	0.15	2.52	0.015	[0.08, 0.7]	43
Political contact vs. no contact	**0.35**	0.17	2.05	0.046	[0.01, 0.7]	43
Non-political vs. political contact	0.03	0.17	0.2	0.841	[-0.31, 0.38]	43
Outcome: "Patriotic"						
Non-political contact vs. no contact	0.12	0.15	0.84	0.404	[-0.17, 0.42]	43
Political contact vs. no contact	0.34	0.17	1.99	0.053	[0-0.004, 0.69]	43
Non-political vs. political contact	-0.22	0.16	-1.35	0.183	[-0.54, 0.11]	43

Note: Results considering all full blocks. Dependent variables are trait ratings of the outparty. Two-sided p-values are reported.

L   Results including all full clusters

As a robustness test, I ignore the blocking element of the randomization procedure and consider only the cluster-level randomization of treatment. This increases the sample size by analyzing all

clusters for which both participants returned to complete the experiment. However, the number of full clusters across conditions now varies with 84 clusters in the no contact condition, 77 clusters in the non-political contact condition, and 77 clusters in the political contact condition. Considering full clusters increases the power of the tests.

First consider change in outparty affect. Table 9 reports treatment effect estimates and randomization inference p-values when using all full clusters. While the effect size estimates are slightly dampened, the patterns of significance are consistent. Both non-political and political contact improve outparty affect, but are not distinguishable from each other.

Table 9: Significance tests of average treatment effect estimates, outcome change in outparty affect, with full clusters

	Estimate	p-value
Non-political contact vs. no contact	**8.41**	≈ 0
Political contact vs. no contact	**6.43**	≈ 0
Non-political vs. political contact	1.98	.28

Note: Results considering all full clusters. Difference-in-differences estimates of average treatment effect and randomization inference p-values under the sharp null hypothesis of no treatment effect. Dependent variable is individual-level change in outparty affect. Two-sided p-values are reported. Both non-political contact and political contact improved outparty affect relative to no contact.

Second, consider stereotypical perceptions of the outparty. Figure 8 plots the mean response of each experimental condition for each of the traits surveyed considering all full clusters. For comparison, ratings of the *inparty* are also plotted for each trait. Randomization inference p-values are also reported for significant results ($\alpha = .05$).

When respondents had interparty social interaction, whether non-political or political, they were less inclined to ascribe the negative traits — closeminded, hypocritical, mean, and selfish — to the outparty. This is consistent with the results when only including full blocks, as presented in the main body of the paper in Figure 4. Additionally, interparty social interaction, whether non-political or political, lead to an increase in ascribing positive traits — honest, intelligent, openminded, patriotic — to the outparty. This is not consistent with the results in Figure 4, likely do to increased power when including all full clusters. Table 10 reports full details on the inparty trait rating results.

However, it is not the case that respondents are feeling better about other people, in general, after a contact experience. Figure 8 also plots the mean trait ratings of the inparty, where largely, perceptions of the inparty are not affected by contact with an outparty member. Table 11 reports full details on the inparty trait rating results.

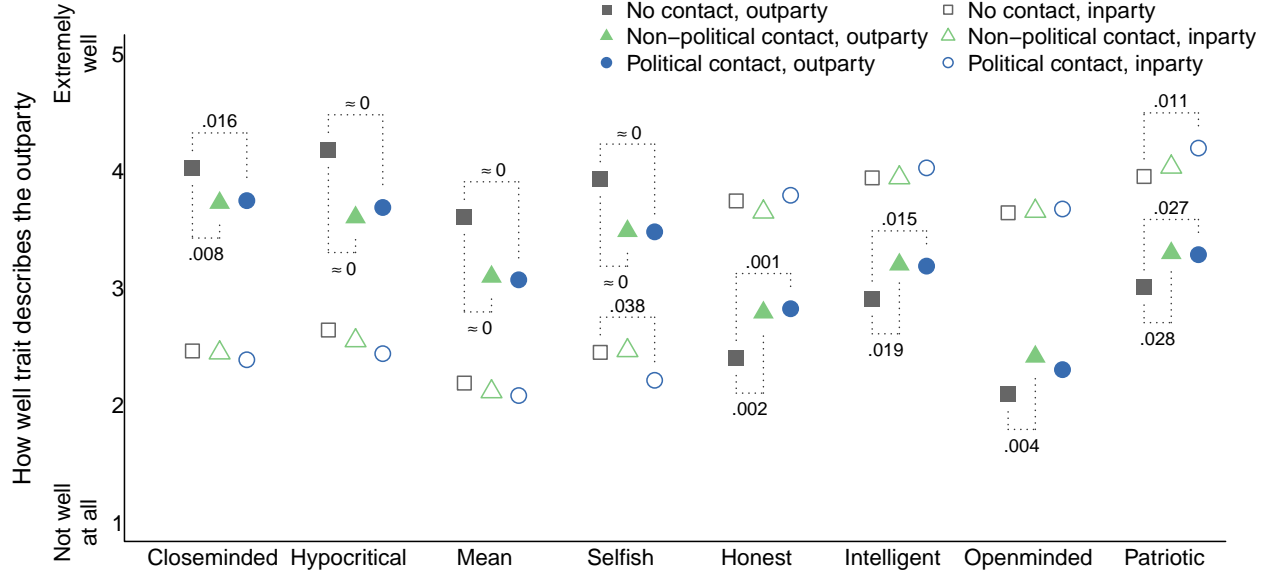Figure 8: Outparty and inparty trait ratings considering all full clusters

Table 10: Significance tests of average treatment effect estimates of **outparty** stereotypes, with full clusters

| | "Closeminded" | | "Hypocritical" | | "Mean" | | "Selfish" | |
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Non-political contact vs. no contact	**-.30**	.008	**-.57**	≈0	**-.51**	≈0	**-.44**	≈0
Political contact vs. no contact	**-.28**	.016	**-.49**	≈0	**-.54**	≈0	**-.45**	≈0
Non-political contact vs. political contact	-.02	.92	-.084	.52	.026	.87	.006	≈1
	"Honest"		"Intelligent"		"Openminded"		"Patriotic"	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Non-political contact vs. no contact	**.39**	.002	**.29**	.019	**.32**	.004	**.29**	.028
Political contact vs. no contact	**.42**	.001	**.28**	.015	**.21**	.07	**.28**	.027
Non-political contact vs. political contact	-.032	.83	.013	.96	.11	.40	.013	.96

Note: Results considering all full clusters. Randomization inference hypothesis tests with the difference in means test statistic under the sharp null hypothesis of no treatment effect. Dependent variable is agreement, on a five point scale with higher values indicating more agreement, with how well the trait describes supporters of the outparty. Two-sided *p*-values are reported.

Table 11: Significance tests of average treatment effect estimates of **_inparty_** traits, with full clusters

	"Closeminded"		"Hypocritical"		"Mean"		"Selfish"	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Non-political contact vs. no contact	-.016	.880	-.091	.433	-.074	.459	.015	.919
Political contact vs. no contact	-.075	.539	-.201	.089	-.106	.328	**-.238**	.038
Non-political contact vs. political contact	.058	.665	.110	.389	.032	.810	.253	.053
	"Honest"		"Intelligent"		"Openminded"		"Patriotic"	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Non-political contact vs. no contact	-.095	.337	.001	$\approx$1	.013	.902	.080	.427
Political contact vs. no contact	.048	.596	.085	.360	.032	.774	**.242**	.011
Non-political contact vs. political contact	-.143	.126	-.084	.407	-.019	.898	-.162	.087

Note: Results considering all full clusters. Randomization inference hypothesis tests with the difference in means test statistic under the sharp null hypothesis of no treatment effect. Dependent variable is agreement, on a five point scale with higher values indicating more agreement, with how well the trait describes supporters of the inparty. Two-sided p-values are reported.

M Partnership agreement

Because the treatment not only involved exposure to an outparty member, but also conversation with them, it is interesting to consider both pre-treatment levels of agreement on the issue of gun control might lead to heterogeneous treatment effects. I code a partnership as having "similar" or "different" views by splitting partnership-level measures of agreement at the mean. The agreement measure is the absolute difference between each individual's mean response to six gun control proposal questions. In Figure 9, we see that partners who agree pre-treatment are no more likely to improve their affect than those who disagreed.

To further assess the impact of agreeing on the topic of gun control, I hand coded each message sent in the political and non-political conversations for whether the message expressed agreement, disagreement, or neither. For each transcript, I then created an agreement score: (# of messages expressing agreement - # of messages expressing disagreement)/(# of messages in transcript). Finally, I split this variable at the mean to dichotomize conversation-level agreement. Figure 10 presents boxplots of this result. We see that agreement and disagreement within the non-political conversations does not impact changes in outparty affect. However, conversations that disagree about gun control improve their outparty affect less than those that agree.
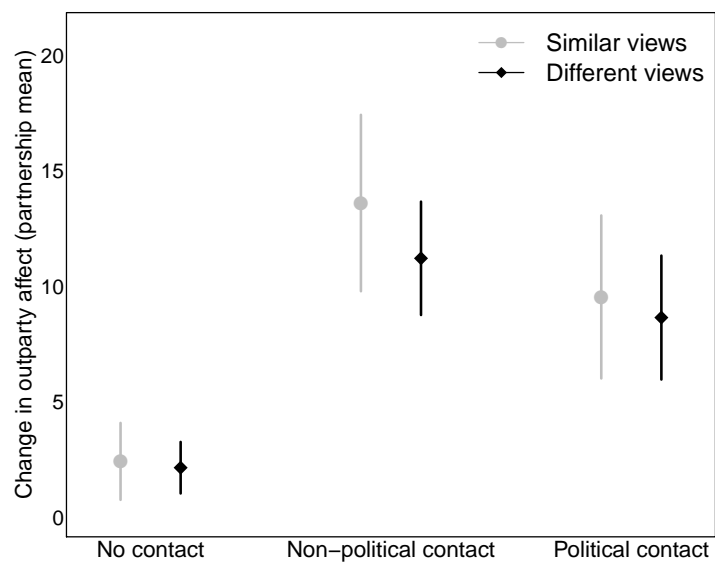
Figure 9: Pre-treatment agreement on gun control

Figure 10: Agreement within conversations