To Discuss or Not to Discuss? How Selective Exposure to Political Discussion Conditions Experimental Findings on Polarization[1]

Erin L. Rossiter[2] &Taylor N. Carlson[3]

October 25, 2023

Abstract

How do political discussions affect polarization? Extensive research suggests that disagreeable (heterogeneous) political discussions can reduce polarization, while agreeable (homogeneous) political discussions can increase polarization. Much of this evidence comes from experiments in which participants are asked to interact with other people in a randomly assigned group discussion setting (e.g., homogeneous or heterogeneous). However, Americans have strong preferences for avoiding political discussion, especially with those who disagree. By assigning people to discussion settings that they might never choose themselves, prior experimental findings may have limited generalizability. In this study, we incorporate self-selection into the standard political discussion experimental design. We use a Preference Incorporating Choice and Assignment (PICA) design to estimate a discussion setting's effect on polarization, conditional on the choice to opt into (or avoid) that setting. We find that heterogenous discussion reduces polarization for people who would and would *not* choose that social setting. In contrast to previous work, we do not replicate that homogeneous discussion increases polarization. Results are durable for at least three days and robust to alternate measures of discussion preferences and polarization. Although homogeneous discussion did not increase polarization, people who prefer homogeneous discussion, but are forced into hearing the other side or reflecting individually, were more likely to opt out of discussions in the future. Our findings have greater external validity than prior experimental work. This study helps us understand how to best scale discussion-based interventions to reduce polarization.

[1]We thank the Alexander and Diviya Magaro Peer Pre-Review Program at Harvard's Institute for Quantitative Social Science for valuable feedback during the design of this study. For generous research funding we thank the following the Social Science Research Council, the Francis and Kathleen Rooney Center for the Study of American Democracy at the University of Notre Dame, and the Institute for the Scholarship in the Liberal Arts, College of Arts and Letters at the University of Notre Dame. This research was approved by Washington University in St. Louis (ID 202304031) and University of Notre Dame (ID 23-04-7805) Institutional Review Boards and adheres to APSA's Principles and Guidance for Human Subjects Research. This experiment's design, hypotheses, and pre-analysis plan were pre-registered with OSF Registry (ID 052523) prior to data collection.

[2]Assistant Professor, University of Notre Dame, erinrossite@nd.edu, erossiter.com

[3]Associate Professor, Washington University in St. Louis, tncarlson@wustl.edu, sites.wustl.edu/tncarlson

> *We can join forces, stop the shouting and lower the temperature. For without unity there is no peace, only bitterness and fury...This is our historic moment of crisis and challenge, and unity is the path forward.*
>
> Joe Biden

> *It does not matter if you are liberal, moderate or conservative. All of us must resist the temptation of further polarization. Instead, we must unite once again as Americans.*
>
> Kevin McCarthy

Politicians, journalists, and voters are increasingly concerned about polarization in the United States. In the face of what appears to be increasing ideological and affective polarization, political leaders on both sides of the aisle have called upon the nation to come together. While many may agree with this sentiment, it seems to be a task that is easier said than done. Civic organizations and social scientists have considered many interventions to reduce polarization, often focusing on what happens when people interact with others who disagree. Yet, most people tend to primarily interact with others who *agree* with them, reflecting widespread concerns about "echo chambers."

To use disciplinary jargon, the question of how politically homogeneous (when people in the group agree with each other) and heterogeneous (when people in the group disagree with each other) group discussion polarizes attitudes has received a great deal of scholarly attention (e.g., Klar 2014; Druckman, Levendusky and McLain 2017; Jost, Baldassarri and Druckman 2022). Much of this work finds that heterogeneous group discussion can reduce polarization, while homogeneous group discussion can increase polarization. The idea is that heterogeneous discussion exposes people to other viewpoints, which can lead them to moderate their own views and become more tolerant of those on the other side of the aisle. In contrast, homogeneous discussion only exposes people to ideas similar to their own, which can reinforce their preferences or even lead them to be more extreme. These theoretical

expectations are intuitive and have been tested extensively by social scientists. Indeed, many methodological approaches have been used to examine the relationship between political discussion composition (homogeneous vs. heterogeneous) and polarization (attitudinal, ideological, and affective), including: observational data of online social media networks (e.g., Barberá et al. 2015), survey data characterizing political discussion networks sometimes elicited via name generators and snowball sampling (Huckfeldt, Johnson and Sprague 2004; Mutz 2006; Minozzi et al. 2020), and deliberative polling (e.g., Fishkin et al. 2021; Amsalem, Merkley and Loewen 2022). The most dominant approach, however, has been randomized experiments in which participants are asked to interact with other people in a randomly assigned political group context (Klar 2014; Druckman, Levendusky and McLain 2017; Levendusky and Stecula 2021; Ahn, Huckfeldt and Ryan 2014; Rossiter 2022; Broockman, Kalla and Westwood Forthcoming).

Experiments are a rigorous choice to estimate the effects of different discussion contexts because they allow researchers to estimate the causal effect of the features of the social experience of interest (e.g., the partisan composition of the group) on polarization. Other approaches, such as measuring the correlation between the partisan homogeneity within one's political discussion network and their political attitudes, suffer from endogeneity issues. For example, people who *choose* to surround themselves with people from the other side of the political aisle might *already* hold less extreme attitudes or might be more open to changing their preferences than people who choose to surround themselves with only those who share their partisan identity. In this case, the political discussions themselves do not cause variation in polarization; rather the observed effects are due to self-selection. Experiments overcome the problem self-selection causes when using observational data to understand the causal effects of political discussion contexts. By randomly assigning people to distinct group contexts, often varying whether groups are politically homogeneous or heterogeneous, randomized experiments remove the threat of people selecting into treatment (i.e., preferable discussion settings). Removing the element of self-selection allows experiments to estimate a clean

causal effect of a given political discussion on polarization outcomes.

This experimental design, sometimes called a "forced exposure" design because it assigns people to discussion contexts without considering if they would ever choose that social setting themselves, leads to an important limitation in interpreting the real-world implications of the results. If treatment effects of discussion settings are only observed amongst groups who would otherwise *not* join those settings in the real world, then findings from forced-exposure designs have limited generalizability beyond the experimental setting. Indeed, previous research has demonstrated that forced exposure designs sometimes overestimate treatment effects, relative to situations in which self-selection is allowed (e.g., De Benedictis-Kessner et al. 2019; Arceneaux and Johnson 2015; Arceneaux, Johnson and Murphy 2012; **?**; Gaines and Kuklinski 2011). For example, De Benedictis-Kessner et al. (2019) use a Preference Incorporating Choice and Assignment (PICA) design to examine the effects of exposure to partisan media on polarization, finding that partisan media can increase polarization both among people who would select into partisan media and those who would not, but the magnitude of this difference varied.

Political discussion is a particularly important context in which selection likely matters. Recent work demonstrates that political discussion is a *social* process in which people make many decisions about whether, how, and with whom to discuss politics (Carlson and Settle 2022). In the 4D Framework outlined by Carlson and Settle (2022), the *decision* stage is a crucial component for understanding the types of people who are willing to opt in to a political discussion and how that choice might affect the impact the discussion has on their subsequent attitudes. The authors show that Americans have clear preferences for avoiding political discussions in the first place, with about one-third of Americans reporting that they prefer to avoid political discussion. If they must discuss, most people report strong preferences for having those discussions with like-minded others (homogeneous discussion). Moreover, these preferences are not uniformly distributed throughout the population. Demographic, political, and psychological characteristics correlate strongly with these preferences. This

further complicates the importance of selection in studies of political discussion, which have largely gone overlooked in forced exposure political discussion experiments.

In short, forced exposure designs could be overestimating the real-world effects of political discussion on polarization outcomes. In this study, we extend this line of work by seeking to understand differential effects of discussion based on individuals' real-world tendencies to self-select into these settings, or even opt-out. When given the choice, people may choose like-minded discussants, people may choose to engage with a variety of viewpoints, or they may choose to avoid political discussion altogether. We expect whether and the extent to which discussion type (heterogeneous, homogeneous, or no discussion) polarizes or depolarizes attitudes will depend on people's choice to engage in these social encounters.

In this study, we apply new experimental methods designed to incorporate self-selection to the political discussion experiment context. Because we expect that political discussion will have different effects for people who would and would not choose a given social experience, a central goal of this study is to estimate treatment effects of political discussion that incorporate this choice, and therefore likely have greater external validity than prior experimental work on the effects of political discussion. Specifically, we use an experimental design suited to understanding political discussion's effects among different sub-groups of people who would choose to participate in the discussion or not. We apply the Preference Incorporating Choice and Assignment (PICA) experimental design (Knox et al. 2019; De Benedictis-Kessner et al. 2019), which allows a researcher to estimate an experimental intervention's effect conditional on choice to receive it.

In summary, we measured participants' pre-treatment preferences for political discussion and several days later invited them to participate in an experiment. In the experiment, participants were first randomly assigned to either a forced exposure design, in which participants were randomly assigned to either a homogeneous discussion, homogeneous discussion, or individual writing task, or a free choice design, in which participants were free to choose which of the three tasks they wanted to complete. The discussions focused on energy

policy and occurred between four participants via an online chat platform called Chatter (Rossiter 2022) and lasted approximately ten minutes. In homogeneous discussions, all four participants had similar pre-treatment energy policy preferences; in heterogeneous discussions, two participants had liberal energy policy preferences and two participants had conservative energy policy preferences. The individual writing task also focused on energy policy and lasted ten minutes. After completing their assigned (or chosen) tasks, participants answered questions about their policy preferences and whether they would be willing to engage in another task. We followed up with participants three days later to measure durability.

We find that heterogeneous political discussion had a durable, depolarizing effect for people who would self-select into this setting *and* people who would self-select out of it. This finding is consistent across several measures of pre-treatment political discussion preferences. This result bolsters previous work suggesting that heterogeneous discussion can reduce polarization, demonstrating that the results are not limited to those who would be unlikely to experience heterogeneous discussion in the real world. However, we find no evidence that homogeneous discussion increases polarization, which goes against the conventional wisdom that echo chambers exacerbate polarization. Preferences for different political discussion settings (or no discussion at all) did not condition the effect of that setting on attitudinal polarization. However, political discussion preferences did affect the extent to which people would be willing to engage in future conversations. Participants who prefer to opt out of conversations, but participated in a discussion in the experiment, were more likely to opt in to a political discussion in the future. On the other hand, however, people who prefer homogeneous political discussion, but are not assigned that discussion setting, were more likely to *opt out* of future discussions.

Our paper makes three key contributions to the study of political discussion and polarization. First, we introduce a new methodological approach to group discussion experiments that carefully incorporates the "decision" stage of political discussion. Second, our findings suggest that we can have more confidence in the external validity of group discussion

experiments utilizing forced exposure designs because we replicate previous findings that heterogeneous group discussion reduces polarization, regardless of self-selection into this (or other) settings. Third, we highlight that there might be an important distinction between attitudinal and behavioral consequences of political discussion. While heterogeneous political discussion reduced polarization for people with all types of discussion setting preferences, these preferences shaped the extent to which people would opt in to discussion in the future. People who prefer homogeneous discussion often have more extreme policy preferences at the outset, suggesting that they have the most to "gain" from heterogeneous discussion. Experiencing a heterogeneous discussion might lead them to moderate their views, but it also makes them less willing to engage in future discussions, which could create a vicious cycle that perpetuates polarization. These nuanced findings about attitudinal vis-a-vis behavioral consequences of political discussion further highlight the need for broader consideration of polarization outcomes and the contours of political discussion.

1 Incorporating Political Discussion Preferences into its Polarizing Effects

This project reinvestigates the state of the literature on group discussion and polarization. Broadly, we argue that selection into different political discussion contexts is crucially important for interpreting the generalizability of treatment effects from forced-exposure group discussion experiments. In this section, we detail the dominant hypotheses from the literature, argue that self-selection could condition these effects, and describe the implications of understanding how self-selection does (or does not) condition the effects of group discussion contexts on polarization.

Because many group discussion contexts and types of polarization interest social scientists, we want to be clear about our language from the outset. First, we define "homogeneous" political discussion as conversations about political issues, candidates, or events with people

who share similar views on the topics being discussed. In contrast, we define "heterogeneous" political discussion as conversations about political issues, candidates, or events with people who share different views on the topics being discussed. Note that we do not define homogeneity or heterogeneity in terms of partisan identity, as some past work has done (e.g., Klar 2014). Although partisan identity likely overlaps with views on the topics discussed in a given political discussion, we are reinvestigating claims about how discussion with others who are like-minded (not necessarily others who share an identity) affects policy preferences. When we say "polarization," we are referring to attitudinal polarization as the primary downstream outcome affected by different group discussion contexts.[4]

Evidence from the political discussion and polarization literatures have generally points to two key findings: heterogeneous political discussion reduces polarization, while homogeneous political discussion increases polarization. However, previous research has not considered the potential differing effects of these distinct discussion contexts based on peoples' choices to join that kind of social setting. While forced exposure to heterogeneous discussion has positive effects for reducing polarization on average; and exposure to homogeneous discussion has negative effects for increasing polarization on average, the tendency to self-select into this exposure is an important moderator to for social scientists to understand because it structures the real-world implications of engaging in different political discussion settings. Moreover, the decision to avoid political discussion altogether is often overlooked entirely. Some people are so averse to politics that any political discussion, regardless of (dis)agreement, could be a "treatment" people would be unlikely to select in the real world.

[4]In the appendix, we investigate affective polarization. Our study was designed to examine attitudinal polarization and choices, such as the topic of conversation, were made with this in mind. If the primary goal was to understand affective polarization, we might have selected different topics or chosen to group people based on partisan identity rather than policy agreement. Still, we investigate affective polarization as a secondary outcome, acknowledging that this was not the primary outcome of interest in the study, but the conversations could still affect it.

1.1 Preferring Heterogeneous Discussion

First, consider people who actively choose to join heterogeneous discussions. These people may have higher political interest if they prefer talking politics to opting out of political discussion. Or, they may have political heterogeneity in their social networks that leads them to be comfortable with choosing mixed-view political discussion. Carlson and Settle (2022) provide evidence for both of these possibilities. Despite the abundance of strong evidence that heterogeneous political discussion reduces polarization, it may be that those who opt-in to heterogeneous discussion enter the conversation having already reaped the benefits of being exposed to diverse views, and thus the effect of it will modest or even negligible.

When people who prefer heterogeneous discussion are forced into a homogeneous discussion, however, the results could be similarly interesting. Perhaps their inexperience with interactions with like-minded others could have a particularly strong polarizing effect. It may be that these people generally prefer heterogenous settings because they like the opportunity to persuade their peers, enjoy a good sparring match, or live for chaos. When the opportunity to argue is no longer there because the people in the discussion largely agree, this could push people farther to the extremes as they look for a way to create conflict. A different explanation could be that people who prefer heterogeneous discussion like it because they are particularly open-minded, persuadable people who like hearing and learning from others. If so, a homogeneous discussion may depolarize these peoples' attitudes as they shift in light of seeing group unanimity on the issue.

1.2 Preferring Homogeneous Discussion

Next, consider people who would prefer homogeneous political discussion. Carlson and Settle (2022) show that this is actually the most common political discussion preference, which is consistent with extensive previous research on political discussion networks indicating a strong pattern of politically homogeneous networks (Sinclair 2012; Huckfeldt, Johnson and Sprague 2004; Ahn, Huckfeldt and Ryan 2014; Carlson, Abrajano and Bedolla 2020; Minozzi

et al. 2020).If people who prefer homogeneous discussion settings are less likely to be treated with heterogeneous discussion in the real world, this experience in a research setting could be particularly powerful, leading to large depolarizing effects. On the other hand, however, it is possible that this group, consisting largely of strong partisans, entrenches into their views and becomes resistant to the opposing views in the discussion. In this scenario, heterogeneous discussion could lead to no effects or even polarizing effects.

1.3 Preferring to Opt Out

Finally, consider people who would opt-out of political discussion. Carlson and Settle (2022) show that about one-third of Americans fall into this category, reporting that they avoid political discussion at all costs, regardless of the views of the potential discussants. The desire to opt out of political discussion could be due to a distaste for politics (Klar and Krupnikov 2016; Klar, Krupnikov and Ryan 2018; Krupnikov and Ryan 2022; Carlson and Settle 2022) or the anticipation of confrontation (Ulbig and Funk 1999; Carlson and Settle 2022; Mutz 2006; Connors and Howell 2022), thus these people avoid political discussion when possible. However, in reality, political discussion may be more civil and informative than anticipated. This potential reality, coupled with the fact that people who prefer no discussion may not have strong political attitudes, could mean that heterogeneous discussion serves its hypothesized role in reducing attitude polarization.

Similarly, undesired exposure to homogeneous discussion could result in strong polarizing effects. Some people prefer to avoid political discussion because they do not feel sufficiently knowledgeable about the topic. If someone would opt out of a homogeneous political discussion because they do not have enough information on the topic, the homogeneous conversation becomes a crucial information source for them. The information to which they are exposed in the discussion is likely to heavily favor one side, which could further polarize their attitudes.

In sum, there are strong theoretically-motivated reasons why selection into treatment could be particularly important in group discussion and polarization experiments. People

who prefer heterogeneous, homogeneous, or no political discussion are fundamentally different from one another in ways that could moderate the effect each discussion context has on polarization.

2 Expectations

As discussed above, incorporating the role of self-selection into our understanding of the effects of discussion on polarization is important. The most rigorous evidence explaining the effects of political discussion on polarization comes from experiments, yet the "forced exposure" design of these experiments may limit how generalizable their findings are beyond the experimental setting. In this section, we outline expectations for how discussion may polarize beyond the experimental setting by incorporating that people have real-world preferences for selecting *into* or *out of* different discussion contexts.

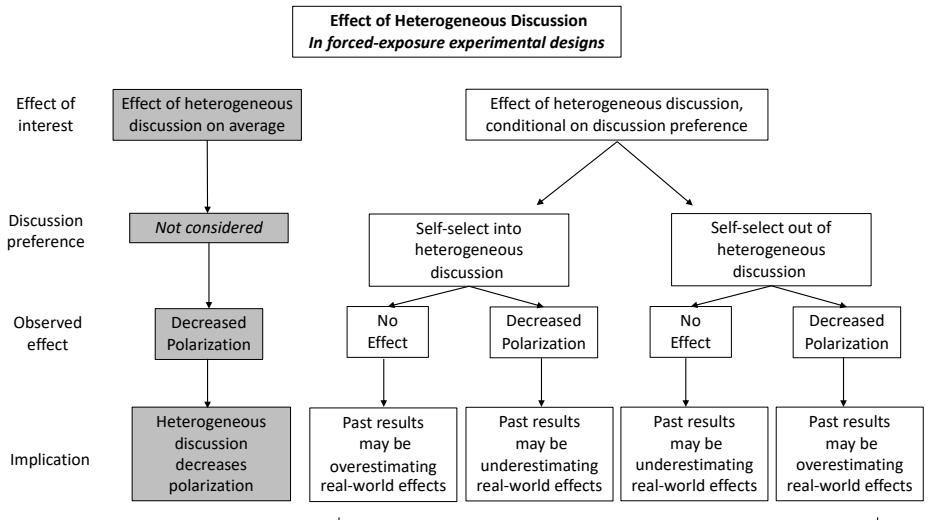
We visualize our main argument in Figures 1 (heterogeneous discussion) and 2 (homogeneous discussion). Both figures show that incorporating self-selection into our understanding of discussion's ability to (de)polarize is an important pursuit regardless of whether we find doing so replicates prior findings or shows prior findings are over- or under-estimating the real-world effects of discussion. Specifically, Figure 1 visualizes how the real-world effects of heterogeneous discussion on reducing polarization may be over- or under-estimated in the literature depending on how the effects relate to self-selection. Figure 2 is a similar visualization for the real-world effects of homogeneous discussion on increasing polarization. We will use Figures 1 and 2 to outline our expectations and their implications.

2.1 Effect of Heterogeneous Discussion on Polarization

First, we will outline our expectations for the treatment effects of heterogeneous discussion amongst the different discussion preference groups discussed in Section 1—those who prefer heterogeneous discussion, homogeneous discussion, or opting out entirely. We illustrate our

expectations in Figure 1. The left-most column in the flow chart, shaded in gray, represents the dominant trends from previous research. The remaining, unshaded portion of the chart reflects what we expect to observe if we were to consider individuals' discussion preferences in the design.

Figure 1: Effect of Heterogeneous Discussion Incorporating Role of Self-Selection



Note: Gray boxes reflect the approach and dominant findings from previous research using forced exposure experimental designs. The remainder of the flowchart illustrates our expectations after accounting for discussion preferences.

As we highlight in gray in Figure 1, the dominant finding in the literature is that heterogeneous discussion is an important social setting for **depolarizing** attitudes. Importantly, this finding is an average treatment effect of heterogeneous discussion, thus is cannot speak to whether people with real-world tendencies to self-select into this setting might have a different treatment effect from heterogeneous discussion than those who would avoid this setting in real life. It may be that the findings from previous experiments overestimate the depolarizing effect for people who already prefer heterogeneous discussion, which is the very group most likely to engage in the behavior by choice.

Therefore, we first consider the effect of heterogeneous political discussion on polarization among those who prefer this type of discussion in the real world (the second path). In other

words, this path of the flowchart examines possible outcomes of heterogeneous discussion for those who would self-select into this discussion context. The key implication is that if we find that heterogeneous discussion has **no effect** when we examine those who would opt-in to this setting in the real-world (the second path down the flowchart), past forced-exposure experimental findings are overestimating the real-world effects of heterogeneous discussion on reducing polarization.

However, if we instead observe that heterogeneous discussion **reduces polarization** for this group (the third path), then our results suggest that past results may be underestimating the real-world effects of heterogeneous discussion. If heterogeneous discussion reduces polarization among the very people most likely to be "treated" outside the experimental setting, past results may be downplaying the real-world implications of heterogeneous discussion for decreasing polarization.

Next, Figure 1 considers the effects of heterogeneous discussion on those who would not select into this setting—either they would prefer homogeneous discussion or would prefer to opt-out of political discussion entirely. We emphasize that this part of the flowchart replicates prior work by "forcing" exposure to people who would not encounter these settings in the real-world. If we find no effect of heterogeneous discussion among those who would not opt-in to heterogenous discussion (fourth path of the flowchart), then prior experimental findings fail to replicate in this group. In this sense, observing no effect of heterogeneous discussion among this group would also imply prior work has underestimated the polarizing effects of heterogeneous discussion. This finding means the "treatment" does not work among people who would not receive it anyway in the real-world, so prior results, when being used to understand how heterogeneous discussion's effects generalize to the real-world, would be dampened by including this group. On the other hand, if we find heterogeneous discussion reduces polarization among this group (fifth path of the flowchart), then prior experimental findings were successfully replicated in this group. This suggests prior work may be overestimating real-world effects of heterogeneous discussion because the effects are

being driven by people who otherwise would not engage in this kind of discussion in the real world.[5]

All together, we are interested in the treatment effects of different discussion settings conditional on the choice to join a heterogeneous discussion, homogeneous discussion, or no discussion at all. Knox et al. (2019) calls these causal quantities of interest the average choice-specific treatment effects (ACTEs). We are interested in three ACTEs of heterogeneous discussion. Specifically, we are interested in the effects of heterogeneous discussion, relative to no discussion, given discussion preference (i.e., the effect for those who would choose heterogeneous, homogeneous, and no discussion).

1. Effect of heterogeneous discussion, relative to no discussion, among those who prefer heterogeneous discussion (would self-select in, if given the choice)

2. Effect of heterogeneous discussion, relative to no discussion, among those who prefer homogeneous discussion (would self-select out, if given the choice)

3. Effect of heterogeneous discussion, relative to no discussion, among those who prefer no discussion (would self-select out, if given the choice)

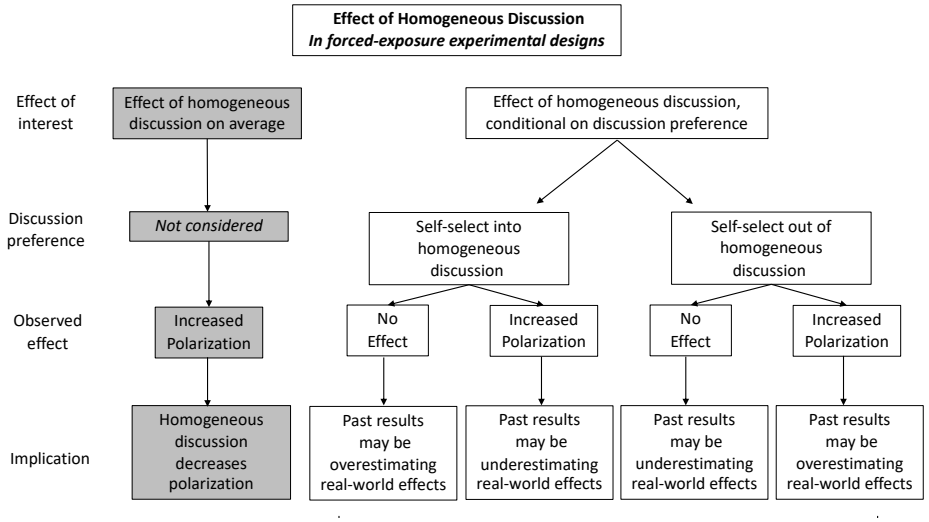
2.2 Effect of Homogeneous Discussion on Polarization

Figure 2 similarly presents the implications of possible outcomes for the effect of homogeneous political discussions. The dominant finding in the literature, highlighted in gray on the left-hand side of Figure 2, is that engagement in homogeneous political discussion increases

[5]We have not yet discussed the possibility that heterogeneous discussion actually polarizes attitudes. Our main expectations in Figure 1 consider the possibility that the depolarizing effect of heterogeneous discussion found on average in the literature either does or does not hold within different subgroups who have different discussion preferences. There is, of course, the possibility that these discussions backfire and polarize attitudes. We do not include this possibility explicitly in Figure 1, however, Figure 1 can be used to demonstrate the implications of this outcome as well. If heterogeneous discussion polarizes, we follow the logic of the "no effect" branches of the diagram because not only was there no effect in the expected direction, but it was in the opposite direction. Thus, the listed implications would only be even stronger if we found backfire effects in heterogeneous discussion.

polarization. We investigate whether this pattern holds once we account for selection into homogeneous discussion.

Figure 2: Effect of Homogeneous Discussion Incorporating Role of Self-Selection



Note: Gray boxes reflect the approach and dominant findings from previous research using forced exposure experimental designs. The remainder of the flowchart illustrates our expectations after accounting for discussion preferences.

If we consider individuals' proclivity for such homogeneous discussion, we could come to different results. We present these possibilities in the unshaded portion of Figure 2. For example, consider the second path down the flowchart in Figure 2. Here, we consider people who prefer homogeneous discussion in the real world, so they would be likely to self-select into this context in reality and to be "treated" outside the experimental setting. If we observe that homogeneous discussion does not increase polarization among those who would self-select into this context, then experimental findings from forced-exposure designs likely overestimate the real-world implications of homogeneous discussion for increasing polarization. That is, the popular narrative that echo-chambers increase polarization might be overstated because the very group most likely to be in an echo chamber in reality might be unaffected by any given homogeneous discussion. In contrast, if we observe that homogeneous discussion increases polarization among those who would self-select into this setting, then we have a

successful replication of previous findings within this subgroup and provided evidence that forced-exposure experimental results may be underestimated.

Now, consider people who would not self-select into homogeneous discussion in reality, shown in the right-hand portion of Figure 2. If we observe that homogeneous discussion increases polarization among people who would not self-select into homogenous discussion (i.e. people who prefer heterogeneous discussion or no discussion), then we have a successful replication of previous findings from forced-exposure experimental designs within this subgroup. This also has important real-world implications, suggesting that if people are inadvertently exposed to homogeneous group settings (i.e., they prefer no political discussion or heterogeneous political discussion, but are forced into a homogeneous setting, perhaps through social media algorithms or geographic sorting), polarization could indeed increase. However, if we observe no effect of homogeneous discussion on polarization among those who would self-select out of such a treatment in the real world, then we have failed to replicate past experimental findings within this subgroup, questioning how generalizable past results are beyond the experimental setting.

Like with heterogeneous discussion, we are interested in three ACTEs for homogeneous discussion. We are interested in the effects of homogeneous discussion, relative to no discussion, given discussion preference (i.e., the effect for those who would choose heterogeneous, homogeneous, and no discussion):

4. Effect of homogeneous discussion, relative to no discussion, among those who prefer homogeneous discussion (would self-select in, if given the choice)

5. Effect of homogeneous discussion, relative to no discussion, among those who prefer heterogeneous discussion (would self-select out, if given the choice)

6. Effect of homogeneous discussion, relative to no discussion, among those who prefer no discussion (would self-select out, if given the choice)

3 Research Design

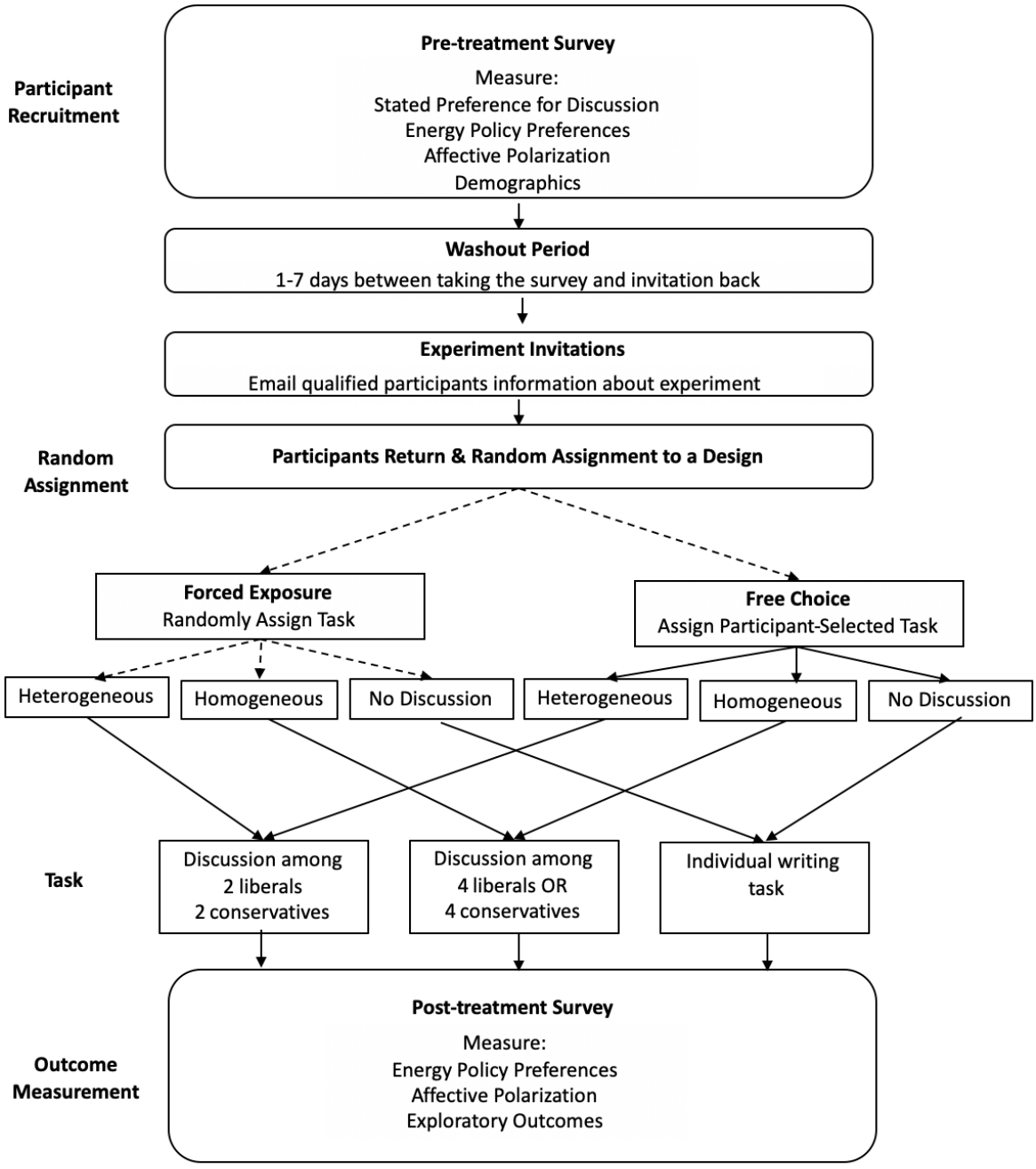
In order to understand how self-selection might condition the real-world effects of different discussion settings, we use a Preference-Incorporating Choice and Assignment (PICA) experimental design, closely resembling the design in De Benedictis-Kessner et al. (2019) and introduced by Knox et al. (2019). This experimental design allows researchers to observe and compare the effects of their stimuli under free choice and forced exposure (i.e., randomization). This allows us to estimate the effects of discussion when people are "forced" to be there and when people select into the experience.

Figure 3 summarizes this design. Broadly, the design includes a recruitment survey where we ask about political discussion preferences, a washout period, and then an invitation to the experiment. When invited participants return, they are randomly assigned to a design: forced-exposure or free choice. Those in the **forced-exposure design** are randomized a second time. They are randomly assigned to heterogeneous, homogeneous, or no discussion, regardless of their pre-treatment stated preference for discussion. Those in the **free-choice design** are not randomized, but instead can choose their task (and will be assigned what they choose): heterogeneous, homogeneous, or no discussion. Participants then engage in either their randomly assigned or chosen task, and then take a post-treatment survey so we can learn about attitudinal polarization from the discussions.

3.1 Recruitment, Washout Period, and Experiment Invitation

First, we begin by administering a recruitment survey in which we measure participants' stated preferences for three types of discussion (heterogeneous, homogeneous, or no discussion). We also collect demographics and pre-treatment measures of our outcomes of interest, which focus primarily on attitudes regarding the conversation topic in the design (energy policy) to measure attitudinal polarization. Specifically, we use an additive index based on four energy policy questions listed in Appendix B.1.

Figure 3: Visualization of Group Political Discussion PICA Design



Key for the PICA design is collecting participants' stated preference for different discussion settings. This question allows us to understand what people who are randomized into a

discussion setting would have freely chosen, if given the chance. Therefore, this question is our key operationalization of what discussion setting each participant prefers, and would choose if given the chance. The question wording is in Appendix B.2. It outlines that we will be posting a task in the future where participants read about energy policy, then engage with the topic in one of three ways: talking with people who agree with you about energy policy, talking with a mix of people who agree and disagree with you about energy policy, or writing about your energy policy views alone. After explaining the future task, we ask "If given the choice, in which of the following tasks would you choose to participate?" We deliberately wrote this question to describe the task options as clearly as possible in order to obtain precise estimates within the context of a PICA design. Although the questions directly measure whether someone would prefer to discuss politics with a group who agrees, a group with mixed views, or have no discussion at all within the context of an online study, this is unlikely to be how people think about political discussion settings in the real world. As such, we also include additional measures of individuals' real-world discussion preferences, outside of the online study context, in order to demonstrate the preference for the study task correlates with these real-world behaviors.[6]

The next step of the design is a washout period between the recruitment survey and the experimental portion of the study. Although (De Benedictis-Kessner et al. 2019) had a washout period of only a few minutes within the same survey, we have a longer washout period of approximately 1-7 days. The washout period has a range because the recruitment

[6]Specifically, we ask three survey questions measuring individuals' broader discussion preferences. We ask people whether all, more than half, about half, less than half, or none of the people they talk to about politics are from their political party, whether people tend to avoid, enjoy (or fall somewhere in between) political discussions, and whether they will talk about politics with anyone, only if they know they have the same views, only if they know they have different views, or if they avoid discussion at all costs. These questions may suffer from response bias, but they allow us to speak more directly to real world discussion preferences and habits, outside the context of a study. We also build a model trained on the choices of participants who were randomized into the free choice arm of the study. We use this model to predict choices for participants in the forced choice arm. This approach avoids the stated preference question, which suffers from systematic switching between certain stated preferences to observed choices, allowing for a good robustness check. However, the free choice arm has a small sample, so the training set for the predictive model is limited and could suffer from overfitting. Many people switched from a stated preference of homogeneous discussion into opting out of discussion, which means we have little training data for the homogeneous choice.

survey was live for several days, varying the length of time between when a participant takes the pre-treatment survey and returns for the experimental portion of the study.

During the washout period, we determined which participants from our recruitment survey are qualified for the experimental portion of the study. The selection criteria are as follows: (1) the participate answers "yes" to a question asking if we can contact them via the online panel notification to participate in future study, (2) the participant's energy policy preference could be classified as liberal or conservative,[7] and (3) the participant passes the open-ended text response quality check question. Moreover, we invited an equal number of participants we classified as "liberal" and "conservative" on the energy policy index. After the washout period, qualified participants were invited via a notification in the platform to return to the experimental task.

3.2 Experimental Randomization

The next section of Figure 3 demonstrates the randomized portion of the study. When invited participants returned after the washout period, they were randomly assigned to a design—either the forced exposure condition (assigned with probability .80) or the free choice condition (assigned with probability .20). If assigned to the forced exposure condition, participants will be randomly assigned to one of the three discussion types with equal probability—heterogeneous discussion, homogeneous discussion, or a solo task in lieu of discussion. If assigned to the free choice condition, participants chose which type of task to join: a heterogeneous discussion, homogeneous discussion, or a solo writing task.[8]

After we randomly assigned a participant to either the forced-choice or free-choice arm

[7]Each energy policy question had a 7-point likert scale for response options. We calculated the average of these four items for each participant. Participants with averages greater than 4 were considered conservative, participants with averages less than 4 were considered liberal, and participants with averages equal to 4 were not invited to participate in the experiment.

[8]Note that we made it clear to participants that all three tasks take the same amount of time to complete, have the same compensation rates, and are anonymous. The discussion tasks do not require extensive additional coordination effort or time spent waiting for discussion partners. We intend for these clarifications to remove extraneous motivations that might correlate with the kind of task, such as thinking one task will be much faster than another, from the participants' decision calculus.

of the experiment, participants in the free-choice arm were asked one survey question that participants in the forced-choice arm were not asked. Participants in the free-choice arm were asked to choose their preferred task. This question exactly mirrors the preference question asked in the recruitment survey, except it makes clear the participant is indeed choosing what task to join in just a few moments (rather than asking which task they would join hypothetically as in the recruitment survey).

As a final step prior to joining the assigned or chosen task, all participants in the experiment were presented with a short description of U.S. energy policy tradeoffs shown in Appendix B.4. We modeled these policy descriptions after previous work on attitude polarization and group discussion by Klar (2014). This step was designed to give participants some background information on which to base their discussions or essays.

Next, participants joined their assigned or chosen task. Participants in the forced-choice design were told at this point what their randomly assigned task is (i.e. that they will be having a heterogeneous or homogeneous discussion, or writing individually). Participants in the free-choice design were likewise informed of the specifics of their chosen task. Participants were also filtered into a specific chatroom according to the pre-treatment measure of respondents' attitudes on the topic (either liberal or conservative on the issue of energy policy). Specifically, liberal (conservative) participants assigned to or choosing homogeneous discussion joined a group with three other participants who we determined also have liberal (conservative) views on the issue. Heterogeneous discussions had two participants with conservative views and two participants with liberal views, which follows closely from previous group discussion research (Klar 2014; Druckman, Levendusky and McLain 2017). Those who were assigned to or chose no discussion wrote about the topic by themselves. The specific task prompts are shown in Appendix D.

Participants discussed or individually wrote about the issue for 10 minutes. All discussions occurred via text-chat using the Chatter app (Rossiter 2022). Chatter is an app that facilitates online chats. Its flexibility makes it uniquely compatible for studies like this that involve

complex interactive designs.

3.3 Post-treatment Surveys

Finally, after completing their discussion or essay, participants were asked to complete a short post-treatment survey. We again asked about energy policy preferences, and for exploratory purposes we asked about perceptions of and reactions to the task (was it a positive experience, did you learn about the outparty, and did you learn about energy policy). At the end of the post-treatment survey, participants were once again given the chance to participate in another bonus task. We presented participants with the same three options (a discussion with people who have similar views, a discussion with people with mixed views, or an individual essay), in addition to an option to simply not participate in a bonus task and complete their work now. This bonus task allowed us to measure an exploratory outcome investigating whether political discussion settings – and individuals' preferences for them – affected willingness to actually engage in the same task again.

Three days after completing the experiment, participants were recontacted via the platform's notification feature to complete a short follow-up survey to determine the durability of treatment effects on policy preferences. Specifically, we asked participants the same four core dependent variables about energy policy preferences. We also included an exploratory question in which participants were given the option to learn more about energy policy via a selection of liberal, conservative, and moderate videos about U.S. energy policy.

4 Results

We conducted this experiment in June 2023 on CloudResearch's online crowdsourcing platform, similar to Amazon Mechanical Turk, called Connect. Throughout one week, 2,434 participants took our recruitment survey and qualified to be invited to the experiment. Of those participants, 940 were classified as having conservative attitudes on energy policy and

1,494 were classified as having liberal attitudes. We invited all 940 conservative respondents and a random sample of 940 liberal participants. Of this sample, 997 participants returned and fully completed their assigned task, which also means their discussion group had all four members participate.

4.1 Describing Preference Groups

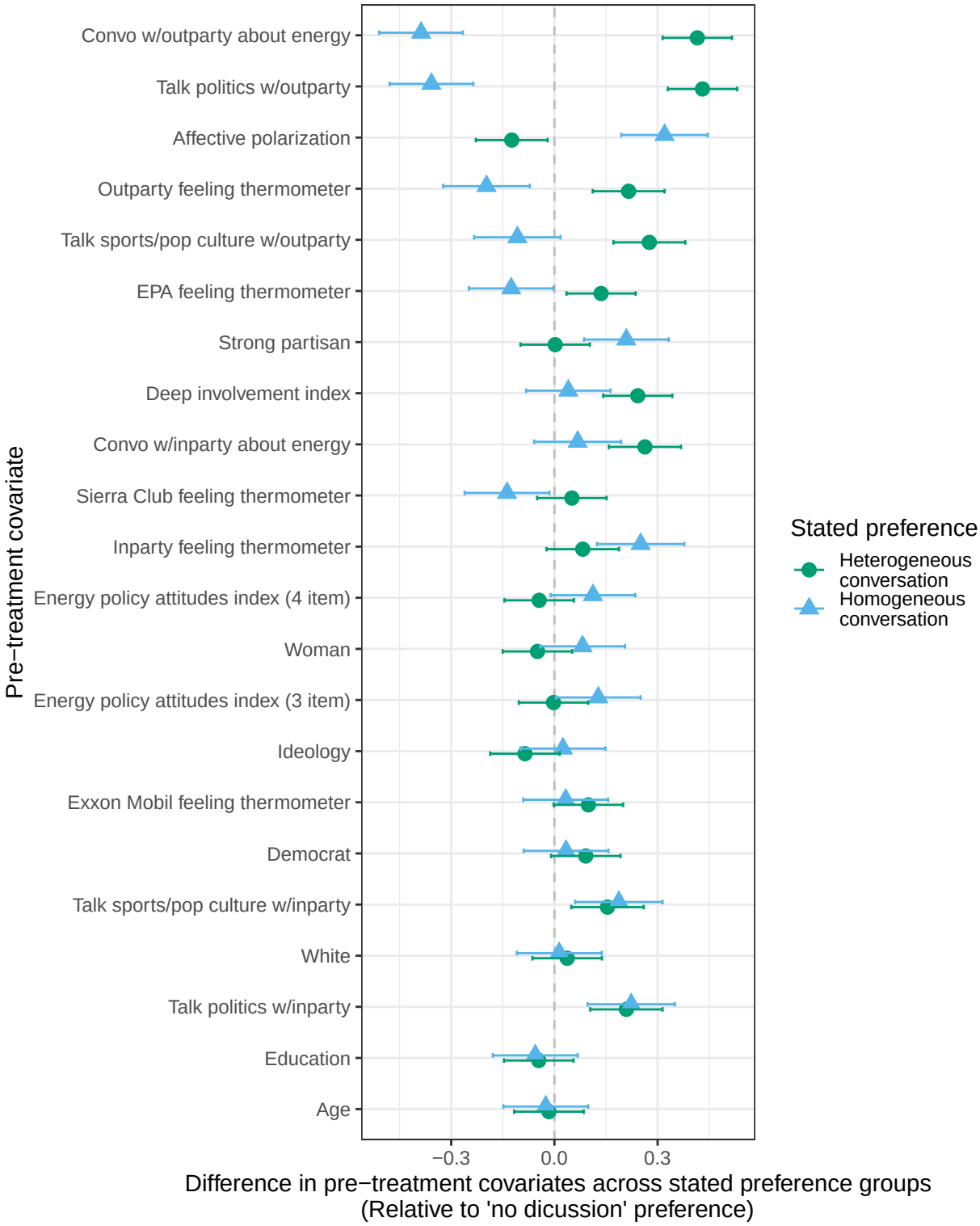
First, we investigate the *free choice* arm of the experiment in which participants got to choose their task. In our main sample of 997 participants, 39%, 20%, and 41% stated they would prefer heterogeneous, homogeneous, and no discussion respectively.[9] This pattern is different from what we would expect based on previous research on political discussion. Notably, the *least* common preference was for homogeneous discussion, although previous work would suggest that opting out would be most common, followed by homogeneous discussion.[10] First, we will assess how this preference correlates with other characteristics before we turn to examining the extent to which stated preference for a given discussion setting equates to what participants actually choose to do when given the choice.

Figure 4 reports linear regression coefficients describing the relationship between stated discussion preference and all pre-treatment covariates asked in the survey, standardized to have mean 0 and standard deviation of 1. We find that preferences for different discussion settings are most strongly correlated with willingness to talk to the outparty about both political and non-political topics. People who prefer heterogenous discussion, relative to the group who prefer to opt-out of political discussion, are more likely to be willing to talk to the outparty about politics; whereas, those who prefer homogeneous discussion are less likely to be willing to talk politics with ouparty members. Likewise, both the groups that prefer heterogeneous and homogeneous conversation are more likely to be willing to talk

[9] These numbers are consistent with the full sample invited to the study.

[10] We suspect that some of this is due to the nature of the task itself and online survey panelists. It is possible that an online discussion is more engaging for people taking many online surveys and that they are looking for a more entertaining task. We view this as an external validity limitation for group discussion experiments more broadly, but we suspect this pattern would be similar in other online samples. Thus, we suspect that our results would generalize to other research studies, but may not reflect real-world preferences.

Figure 4: Pre-treatment Covariates by Preference Group



with inparty members about politics, relative to those who prefer to opt-out. These results validate our survey question measuring participants' stated preference for different discussion settings. (See Appendix C for additional validation of the stated preference question.)

We also find these discussion setting preferences to be highly correlated with partisan strength and affective polarization. People who prefer homogeneous discussion are stronger partisans and more affectively polarized. Finally, people who prefer homogeneous discussion also hold stronger views on energy policy. Notably, demographics such as age, education, race, gender, and partisanship are not strong predictors of political discussion preferences.

Figure 4 demonstrates the many ways selection into discussion settings correlates with covariates that also explain discussion's potential effects. Researchers who use randomized experiments to understand the effects of different discussion settings on political attitudes and behaviors alleviate the threat self-selection imposes on these causal effects. However, these key differences across groups who prefer heterogeneous discussion, homogeneous discussion, and opting out entirely suggest that their reactions to different discussion settings might also vary in important ways.

However, people *saying* they prefer a certain discussion setting may not actually *choose* to engage in it when the opportunity presents itself. The free choice arm of the PICA design allows us to investigate whether stated preferences align with actual behavior. In the free choice arm, we find that 65%, 44%, and 87% of those who stated preferring heterogeneous, homogeneous, and no conversation respectively chose that option in the experiment. When deviating, the people who stated they would prefer homogeneous or heterogeneous conversation instead chose to opt-out (47% and 31% respectively). Very few people opted *for* a discussion setting that was against their stated preference. (Full results are in Appendix E.)

People who deviated from their stated preference of homogeneous or heterogeneous conversation, choosing to opt out when faced with a real choice, may have behaved inconsistent with their stated preference because the reality of having to share their views with others—likeminded or not—was intimidating or anxiety inducing, as political discussion can be in

the real-world. Alternatively, it could have been features of the task at hand that led people to want to switch. For example, in the pre-treatment survey, people could have reported that they would prefer heterogeneous discussion because they thought it would increase their chances of being selected for the follow-up study, but then when given the actual choice, they chose to do the individual task because it seemed less demanding. Regardless of whether the reason people switched was because of the experiment or real, theoretically motivated reasons related to discussion, an assumption of the PICA design is that a participants' stated preference can substitute for observing their actual behavior. In a future draft, we will conduct sensitivity analyses introduced by Knox et al. (2019) to assess if our results hold under violations of this assumption.

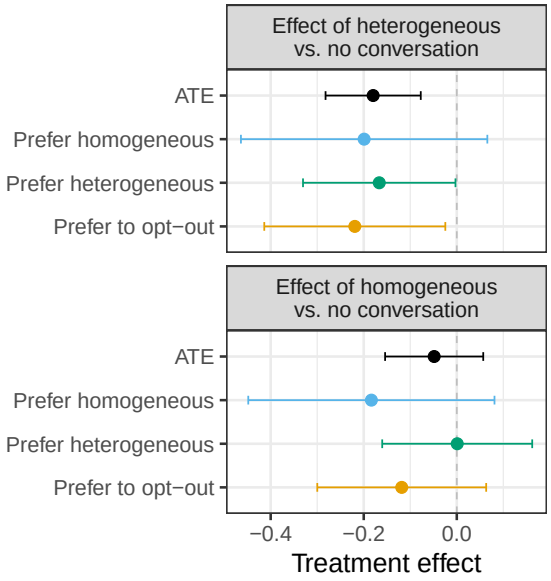
4.2 Forced Choice Results

Next we turn to the forced choice arm of the experiment. We estimate average treatment effects (ATEs) to replicate prior experimental work. We also estimate the ACTEs for the six quantities outlined in Section 2 by estimating subgroup treatment effects given stated preferences amongst participants in the forced-exposure arm of the experimental design (Knox et al. 2019).[11] We use a linear model estimator for each test and cluster standard errors for participants in the same conversation. In all models, we control for pre-treatment covariates to increase precision. We control for participants' pre-treatment measure of the outcome variable if applicable and the following demographics: age, gender, education, race and ethnicity, party identification, ideology.

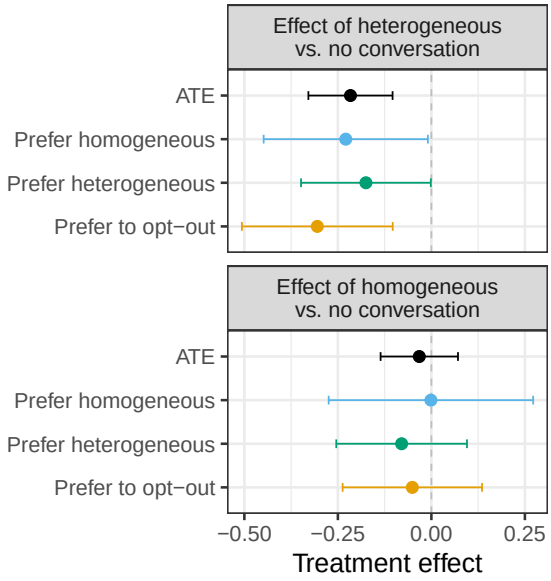
Our main outcome of interest is attitude polarization. To measure polarization, we follow previous research and fold our energy policy index so that larger values mean the participant's views are more polarized and smaller values mean less polarized. We fold the variable based on our pre-treatment determination of if the participant has liberal or conservative views on

[11] This is the naïve estimator of ACTEs because it assumes any deviations between stated preferences and actual choice are not associated with potential outcomes of the discussion treatments. We will probe the sensitivity of our results to deviations in this assumption in future drafts.

Figure 5: ACTE Estimates for Polarization of Energy Policy Attitudes for Each Discussion Preference Group



(a) Immediate Effects



(b) Effects Three Days Later

the issue.

Figure 5 shows results for our main pre-registered outcome of interest—attitudinal polarization regarding energy policy. The outcome is standardized to have a mean of 0 and a standard deviation of 1, so treatment effects can be interpreted in terms of standard deviations from the mean. First, we see that the average treatment effect (ATE) of heterogeneous discussion, relative to no discussion, depolarizes attitudes. This replicates prior experimental work. However, we see that the ATE for the effect of homogeneous discussion relative to no discussion is not significant, which fails to replicate prior findings on the polarizing effect of like-minded discussions.[12]

The core of our inquiry is whether incorporating discussion preferences will change our understanding of discussion's effects. We find that heterogeneous conversation depolarizes regardless of whether one would self-select into this setting or not. Despite Figure 4 demonstrating that the preference subgroups being markedly distinct in ways that could affect their

[12]Heterogeneous discussion's depolarizing effect is stronger than homogeneous (p=0.013).

response to a discussion, the ACTE for each discussion preference group is negative. (The ACTE for those that prefer homogeneous discussion has high levels of uncertainty (p=0.146) because this is the smallest subgroup in our sample).[13]

Turning to the effect of homogeneous discussion among preference groups, we see suggestive evidence that those who want to be there depolarize the most. This runs contrary to the echo chamber literature that suggests homogeneous discussion can create a self-reinforcing cycle where people's opinions are amplified, rarely challenged, and features group think, which taken together can contribute to polarization. We note these results feature high levels of uncertainty, and future work should replicate if homogeneous discussion can indeed depolarize the people that choose this setting.[14]. Ultimately, our findings suggest that the conventional wisdom that echo chambers are harmful because they can increase polarization might be misplaced, at least within the context of political discussion. We find no evidence that homogeneous discussion affects polarization at all among people who would not opt into homogeneous discussion in the real world, and we find some suggestive evidence that homogeneous discussion actually *decreases* polarization among people who would opt into

[13]In the appendix, we present these results using other measures of political discussion preferences. These results are purely exploratory as we did not pre-register these robustness checks. Among those who would choose heterogeneous discussions, heterogeneous discussion reduces polarization, relative to an individual essay on the same topic. This is the case if we measure discussion preferences using the formal stated preference (-.13, p=.073), predicted choice (-.20, p=.031), heterogeneous social networks with respect to partisanship (-.18, p=.02), and a general habit of talking politics with heterogeneous groups with respect to attitudes (-.22, p=.008). Among people who would prefer homogeneous discussion, but were assigned heterogeneous discussion, the point estimates are all negative and of similar magnitude except one, but only two operationalizations are significant: stated preference (-.26, p=.034), predicted choice (-.26, p=.17), homogeneous social networks with respect to partisanship (-.17, p=.007), and a general habit of talking politics with homogeneous groups with respect to attitudes (-.07, p=.414). Finally, among those who would prefer to opt-out of discussing politics entirely—whether operationalized as in our experimental context or in their everyday lives—heterogeneous discussion depolarizes relative to writing a short essay, including when operationalized as stated preference (-.22, p=.011), predicted choice (-.20, p=.01), a preference to avoid of political discussions (-.20, p=.023), and a habit of opting out of political discussions whether they know the distribution of attitudes in the group or not (-.14, p = .178).

[14]In the appendix, we present these results using other measures of political discussion preferences. For those who would self-select into homogeneous settings, the point estimates are all negative and of similar magnitude except one, but only two operationalizations are significant: stated preference (-.24, p=.031), predicted choice (-.23, p=.261), homogeneous social networks with respect to partisanship (-.07, p=.272), and a general habit of talking politics with homogeneous groups with respect to attitudes (-.22, p=.024). For those would not self-select into homogeneous settings, we consistently fail to find evidence of any effect on attitudes. Therefore, homogeneous settings only suggest a depolarization effect for those who would self-select into them.

that setting in reality. While more work is needed to validate these findings, our work at minimum suggests that the conventional wisdom should be questioned.

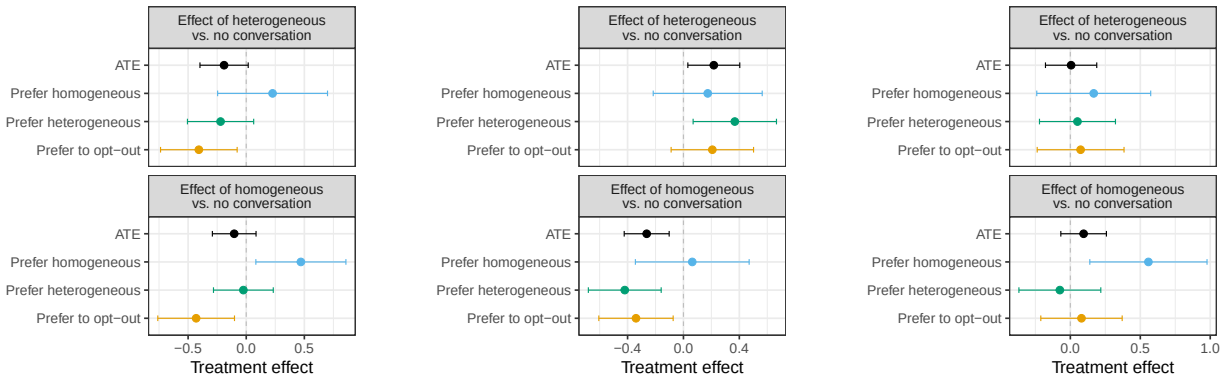
We resurveyed participants at least three days after their conversation to ask the same energy policy questions. We find that the depolarizing effect of heterogeneous conversation is durable across all preference subgroups. The null effects of homogeneous discussion on polarization remained null over time, and the depolarizing effect of homogeneous discussion among those who prefer homogeneous discussions decayed to 0.

4.3 Exploratory Outcomes: Mechanisms and Repeating the Task

Despite finding no differences in treatment effect depending on one's preference for the social setting, we find stark differences by preference subgroup regarding their experience in the discussion. Figure 6 reports exploratory outcomes we asked to probe potential mechanisms of discussion's effects. People who prefer to opt-out of political discussion felt either kind of discussion was a negative or bad experience, relative to short essays. In stark contrast, there was a strong association for reporting homogeneous conversations as positive or good for those who wanted to be there. Likewise, the only group who had a positive effect on learning about energy policy were those who were exposed to homogeneous discussion and wanted to be there. Finally, people who wanted to be in a homogeneous setting had null effects on learning about the outparty, but the effect was negative for those who were forced into the likeminded setting. Despite these stark contrasts in how people experienced and perceived their discussion, we find no evidence of distinct effects of discussion on polarization based on their preference to be in that setting or not.

Finally, we investigated whether political discussion preferences and experiences affected whether someone would be willing to engage in a discussion again. At the end of the post-treatment survey, we asked participants if they were interested in completing a bonus task for an additional financial incentive. The task would be completed right away, rather than needing to come back at a different time, allowing us to investigate whether people would

Figure 6: ACTE Estimates for Possible Mechanisms for Each Discussion Preference Group



(a) Valence (Positive or Negative)

(b) Learned about the Outparty

(c) Learned about Energy Policy

be immediately interested in engaging in another discussion. We presented people with the same options, in addition to just ending their participation altogether (with no penalty).

For people who would choose heterogeneous discussions before our randomized intervention, the probability of choosing heterogenous discussion for the bonus task was about .5 due to having any discussion—heterogeneous or homogeneous. The echoing of similar views and comfortability of agreement in homogeneous discussion did not dissuade this group from their preference for hearing alternative views. Writing an essay instead of discussing did significantly increase the probability of choosing the essay, however.

For people who would choose homogeneous discussion, we see some of our only evidence of possible backfire effects of people being forced into settings they wouldn't self-select into otherwise. If assigned to a setting they would not self-select into, this group is more likely to choose to opt out in the future (none of the above or essay) than to join future discussions. If given the homogeneous setting they prefer, the probability of choosing that setting again is much higher (.5).

For people who would prefer to opt-out before our randomized intervention, the probability they would choose to do an essay is high regardless of treatment assignment. However, assignment to discussion led to a sharp decrease in predicted probability of choosing the essay option (about .4) relative to if they were assigned to their preferred setting of the essay

(about .75). This is good news—when people experience political discussion despite their aversion to it, it increases their willingness to join these settings in the future.

5 Conclusion

This study makes several contributions. First, from a theoretical perspective, this study seeks to better understand the extent to which preferences for different kinds of political discussions, or an aversion to political discussion altogether, can contribute to polarization. Previous research on political discussion highlights the importance of these decisions in explaining which political discussions are likely to occur (Carlson and Settle 2022), but research on the relationship between political discussion and polarization has not yet adequately considered these preferences. We find that, despite preferences for discussion settings being correlated with important political characteristics, like affective polarization and partisan strength, heterogeneous discussion depolarizes attitudes regardless of tendency to self-select into this setting in the real world. Contrary to prior experimental findings, homogeneous discussion fails to polarize attitudes in our study, and we fail to find evidence for a polarizing effect among all preference subgroups, even those who would normally opt-out of political conversation and thus might have the most malleable views.

Second, from a more practical standpoint, understanding heterogeneity based on political discussion preferences will help us understand how interventions to reduce polarization may scale in the real world beyond the experimental setting. Many political discussion experiments are justified as important because of their potential to help reduce polarization in the real world. However, if treatment effects are only observed among people who would not otherwise choose discussion in the real world, then interventions need to consider how to get people to come together for discussion in the first place. However, we do not find this to be the case.

Our results suggest heterogeneous discussion was a powerful depolarizing experience regardless of one's interest in joining this setting. This is important for depolarization

practitioners. These results suggest that people willingly joining opportunities for open dialogue and diverse perspectives—the people walking through the door at depolarization events—may still see positive effects from depolarization interventions. It is not the case that their tendency to expose themselves to these kinds of settings in their everyday lives has put a ceiling on the positive effects of engaging with diverse perspectives and information. These findings are equally important for our understanding of compelling people to partake in heterogeneous settings and political discussion in general. Heterogeneous discussion had durable depolarizing effects among people who preferred a likeminded discussion or to avoid political talk altogether.

References

Ahn, TK, Robert Huckfeldt and John Barry Ryan. 2014. *Experts, activists, and democratic politics: Are electorates self-educating?* Cambridge University Press.

Amsalem, Eran, Eric Merkley and Peter John Loewen. 2022. "Does talking to the other side reduce inter-party hostility? Evidence from three studies." *Political Communication* 39(1):61–78.

Arceneaux, Kevin and Martin Johnson. 2015. "How does media choice affect hostile media perceptions? Evidence from participant preference experiments." *Journal of Experimental Political Science* 2(1):12–25.

Arceneaux, Kevin, Martin Johnson and Chad Murphy. 2012. "Polarized political communication, oppositional media hostility, and selective exposure." *The Journal of Politics* 74(1):174–186.

Barberá, Pablo, John T Jost, Jonathan Nagler, Joshua A Tucker and Richard Bonneau. 2015. "Tweeting from left to right: Is online political communication more than an echo chamber?" *Psychological science* 26(10):1531–1542.

Broockman, David, Joshua Kalla and Sean Westwood. Forthcoming. "Does Affective Polarization Undermine Democratic Norms or Accountability? Maybe Not." *American Journal of Political Science* .

Carlson, Taylor N and Jaime E Settle. 2022. *What goes without saying*. Cambridge University Press.

Carlson, Taylor N, Marisa Abrajano and Lisa García Bedolla. 2020. *Talking Politics: Political Discussion Networks and the New American Electorate*. Oxford University Press.

Connors, Elizabeth C and Christopher Howell. 2022. ""You Need to Calm Down": How Tone Shapes Political Discussion.".

De Benedictis-Kessner, Justin, Matthew A Baum, Adam J Berinsky and Teppei Yamamoto. 2019. "Persuading the enemy: Estimating the persuasive effects of partisan media with the preference-incorporating choice and assignment design." *American Political Science Review* 113(4):902–916.

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* .

Fishkin, James, Alice Siu, Larry Diamond and Norman Bradburn. 2021. "Is deliberation an antidote to extreme partisan polarization? Reflections on "America in one room"." *American Political Science Review* 115(4):1464–1481.

Gaines, Brian J and James H Kuklinski. 2011. "Experimental estimation of heterogeneous treatment effects related to self-selection." *American Journal of Political Science* 55(3):724–736.

Huckfeldt, Robert, Paul E Johnson and John Sprague. 2004. *Political disagreement: The survival of diverse opinions within communication networks*. Cambridge University Press.

Jost, John T, Delia S Baldassarri and James N Druckman. 2022. "Cognitive–motivational mechanisms of political polarization in social-communicative contexts." *Nature Reviews Psychology* 1(10):560–576.

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

Klar, Samara and Yanna Krupnikov. 2016. *Independent politics*. Cambridge University Press.

Klar, Samara, Yanna Krupnikov and John Barry Ryan. 2018. "Affective polarization or partisan disdain? Untangling a dislike for the opposing party from a dislike of partisanship." *Public Opinion Quarterly* 82(2):379–390.

Knox, Dean, Teppei Yamamoto, Matthew A Baum and Adam J Berinsky. 2019. "Design, identification, and sensitivity analysis for patient preference trials." *Journal of the American Statistical Association* 114(528):1532–1546.

Krupnikov, Yanna and John Barry Ryan. 2022. *The other divide*. Cambridge University Press.

Levendusky, Matthew S and Dominik A Stecula. 2021. *We Need to Talk: How Cross-Party Dialogue Reduces Affective Polarization*. Cambridge University Press.

Minozzi, William, Hyunjin Song, David MJ Lazer, Michael A Neblo and Katherine Ognyanova. 2020. "The Incidental Pundit: Who Talks Politics with Whom, and Why?" *American Journal of Political Science* .

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

Rossiter, Erin. 2022. "The Consequences of Interparty Conversation on Outparty Affect and Stereotypes." Working paper. https://erossiter.com/files/conversations.pdf.

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

Ulbig, Stacy G and Carolyn L Funk. 1999. "Conflict avoidance and political participation." *Political Behavior* 21:265–282.