

Rank-heterogeneous effects of political messages: Evidence from randomized survey experiments testing 59 video treatments

Luke B. Hewitt ^a Ben M. Tappin ^b

Massachusetts Institute of Technology

^a lhewitt@protonmail.com

^b benmtappin@gmail.com

This is a preprint and has not yet been published in a peer-reviewed journal.

Abstract

Central to theories of political persuasion is treatment effect heterogeneity—the idea that people respond to political messages in different ways—so persuasion is easier when different messages are targeted to different audiences. The standard approach to testing for heterogeneity is to examine whether the effect of an individual message differs between subgroups of people (such as liberals versus conservatives). We describe the shortcomings of this approach, and propose an alternative: jointly examining *many* messages on the same political issue, and assessing whether the rank-order of their effects differs between subgroups (which we call “rank-heterogeneity”). Implementing this approach, we conduct two large-scale survey experiments spanning two policy issues, 59 message treatments, and over 40,000 American adults. Across experiments we find mixed evidence of rank-heterogeneity, suggesting that it depends upon the particular issue in question. However, in the case where we do observe strong evidence of rank-heterogeneity, its primary cause is consistent with the predictions of moral reframing theory, an influential account of heterogeneity in political persuasion. Alongside these implications for theory, our results have implications for political persuasion in practice.

Keywords: Persuasion, attitudes, heterogeneity, American politics, experiment

Author note

We thank Jack Blumenau and Alex Coppock for feedback on a previous version of this manuscript. Any errors are ours. COI statement: we are founders of a research organization that conducts public opinion research.

Introduction

What types of messages persuade people in politics? A fundamental assumption of many theories of persuasion is that the answer depends on the person receiving the message—because different types of people find different types of messages persuasive. As Hornikx and O’Keefe (2009) explain, “[for] centuries, students of persuasion have taken it to be a commonplace that, in order to be effective, persuasive messages should be adapted to the audience” (p.3). Indeed, major theories of political attitude formation posit that variables such as people’s existing attitudes (Lord, Ross, and Lepper 1979; Taber, Cann, and Kucsova 2009; Taber and Lodge 2006), moral values (Feinberg and Willer 2015, 2019) and political and social identities (Kahan 2016; Van Bavel and Pereira 2018), among many others, moderate the extent to which they change their attitudes in response to political messages. Notably, this assumption—that there exists reliable heterogeneity across people in the effects of messages—underpins the logic of targeting different types of messages to different types of people in order to maximize their persuasive impact (Teeny et al. 2021).

The prominence of academic theories of heterogeneity, combined with the apparent prevalence of message-targeting in the political advertising industry (Dobber, Fathaigh, and Borgesius 2019; Privacy International 2020), suggests that there exists substantial heterogeneity across people. In other words, the political message that best persuades one type of person is perhaps *generally unlikely* to be that which best persuades another, different type of person. While this could be the case, in this paper we argue that there is actually limited existing evidence (in the academic literature at least) to speak to this question. This lack of evidence has important implications, both for prominent theories of heterogeneity and for political persuasion in practice.

Our argument rests on a simple premise: the most common method of examining heterogeneity is ill-equipped to determine whether or not the political message that best persuades one type of person is that which also best persuades another type of person. Therefore, in this paper we propose an alternative method, and apply it in two large-scale survey experiments spanning two U.S. policy issues, 59 message treatments, and over 40,000 American adults. In doing so, we simultaneously provide a new and comprehensive test of *moral reframing theory*, an influential account of persuasive communication which claims that political messages cause greater attitude change when they are matched to people’s moral values (Feinberg and Willer 2019). In the next section, we lay out our argument in greater detail.

Limitations of the common operationalization of heterogeneity

A common approach for examining heterogeneity in the effects of political messages is to conduct an experiment in which people are randomly exposed to a political message, and then to estimate whether the average effect of the message on people’s attitudes varies across those of different demographics, political identities, moral values, or any number of other covariates. Typically, the inferential threshold used to conclude heterogeneity in such cases is observing an interaction between treatment assignment and the relevant covariate—indicating that the average treatment effect of the political message differs between subgroups of the sample.

While the norm, this operationalization of heterogeneity cannot determine how often the political message that best persuades one type of person is that which also best persuades another type of person. To illustrate why, consider the following case. Imagine that a researcher conducts a survey experiment like that described above, and observes that the average treatment effect of a political message **M** is half as large

among people in subgroup 1 (e.g., liberals) than in subgroup 2 (conservatives). The difference between subgroups is statistically significant—evidence of heterogeneity by the standard interaction effect definition. This result is depicted in Figure 1A.

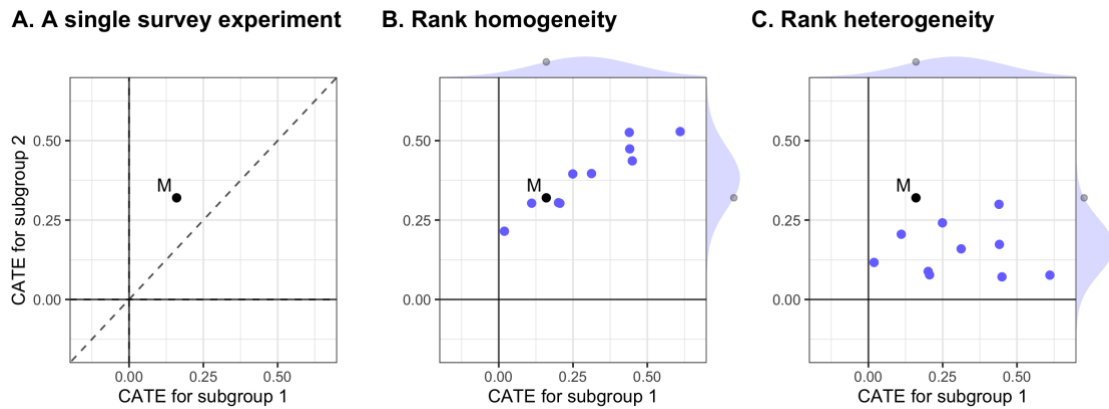


Figure 1. Illustrative example of conditional average treatment effects (CATEs) for a hypothetical set of experiments. The overall distribution of treatment effects for each subgroup is shown at the margins.

Now imagine that the researcher conducts ten further survey experiments, using a different type of political message as the treatment in each. In almost every experiment, they observe that the average treatment effect of the message is different in subgroup 1 versus subgroup 2. Sometimes the differences are larger, and other times they are smaller. However, because the researcher is diligent and rigorous, they recruited a large sample of respondents in each experiment and are thus able to statistically detect even small differences between subgroups. The results of these ten imaginary experiments are plotted in Figure 1B alongside the original experiment.

Figure 1B reveals a striking pattern: while there is clear evidence of heterogeneity in many of the individual experiments (an interaction effect), the

subgroup treatment effects are strongly correlated *across* experiments—such that the strongest and weakest treatments are essentially the same for subgroups 1 and 2. This pattern indicates that, despite responding differently to the individual messages, people in either group were nevertheless receptive to *similar* features of the messages in general—and this shared receptivity swamps the heterogeneity at the level of individual messages. Put differently, the pattern in Figure 1B reveals evidence of *homogeneity in the rank order of message effects*. Thus, we call this pattern rank-homogeneity. When considered relative to the overall distribution for each subgroup (shown at the margins of Figure 1B), the original message **M** is unremarkable; it is a somewhat weak treatment for people in subgroup 1, and a similarly weak treatment for people in subgroup 2. In other words, the message that best persuaded one type of person tended to be that which also best persuaded a different type of person.

Contrast this with the pattern depicted in Figure 1C. There, the subgroup treatment effects are *uncorrelated* across experiments, and so the rank ordering of treatment effects differs substantially between subgroups. If the effect of a message on one subgroup is large, this does not mean it is large for other subgroups, and may even suggest it is small among other subgroups. This is clear evidence of *heterogeneity in the rank order of treatment effects*—a pattern that we call *rank-heterogeneity*. Relative to the overall distribution for each subgroup, the original message **M** is a somewhat weak treatment for subgroup 1, yet it is an unusually strong treatment for subgroup 2. The message that best persuaded one type of person was generally unlikely to be that which best persuaded a different type of person.

Are we in a world more like that depicted in Figure 1B or 1C? That is, should we expect to see rank-homogeneity or rank-heterogeneity in message persuasiveness? Answering this question is important for understanding whether people are receptive to

similar or different features of messages in general. As a result, the answer to this question has important implications for political persuasion theory and practice.

Implications for persuasion theory and practice

The persuasive advantage of targeting and micro-targeting political messages scales with rank-heterogeneity; if the message that best persuades one type of person is rarely that which best persuades a different type of person—i.e., there is high rank-heterogeneity—then targeting different messages to different people can result in greater returns to persuasion (compared with not targeting). Thus, the extent of rank-heterogeneity has practical implications for public messaging campaigns. For example, suppose that a campaign wants to persuade both liberals and conservatives to change their behavior. Should they use different messaging for each audience, or use the same messaging for both? If rank-heterogeneity is low, then there may be a single “best” message, and targeting will generate little benefit. Furthermore, in light of recent public concerns over the power of political micro-targeting to influence voters’ behavior (BBC 2018; Cadwalladr 2017; Scott 2018), better understanding the extent of rank-heterogeneity can shed empirical light on the validity of those concerns.

Understanding the extent of rank-heterogeneity also has implications for influential theories of political communication. For example, moral reframing theory is one such theory, claiming that political messages are most persuasive when they are framed to appeal to people’s moral values (for a review, see Feinberg and Willer 2019). More specifically, moral reframing theory is based on a more general theory of human moral psychology—known as moral *foundations* theory—which posits that there are five core moral value foundations (care/harm, fairness, loyalty, authority, and sanctity) that different types of people endorse to different degrees (Graham et al. 2013). Most

notably, liberals in the U.S. are claimed to endorse the “individualizing” foundations of care/harm and fairness more than conservatives, while putting lesser weight on the “binding” foundations of loyalty, authority, and sanctity (Graham, Haidt, and Nosek 2009). As applied in moral reframing theory, previous research has found evidence to suggest that messages based on the individualizing foundations are uniquely persuasive among liberals, while messages based on the binding foundations are uniquely persuasive among conservatives (Feinberg and Willer 2013, 2015, 2019).

However, a key question left open by previous work on moral reframing theory is the extent to which liberals and conservatives are receptive to different features of messages *in general*. On one view, the values of liberals and conservatives are sufficiently different that appealing to their distinct sets of values is all but essential when trying to successfully persuade them—it will rarely be the case that they find the same types of messages persuasive (characteristic of rank-heterogeneity). On another view, while liberals and conservatives may put different emphasis on some values, there is sufficient overlap among their values as a whole that they respond largely similarly to most types of messaging appeal (characteristic of rank-homogeneity). Previous work has rarely evaluated the persuasiveness of morally-framed appeals alongside other types of messaging appeals—such as appeals to scientific evidence, commonsense, or expert opinion—and thus cannot distinguish between these views.

Relevant previous research

To estimate rank-heterogeneity (versus rank-homogeneity) and shed light on the foregoing questions, it is necessary to estimate the persuasive effects of many different messages for various subgroups of people, and examine the correlation between the estimates (i.e., Figure 1). This demands a very large sample of messages and people.

To our knowledge, Coppock (2016, 2022) has conducted the most relevant analysis of this type to date. He analyzed twenty-three existing experiments in which Americans were randomized to receive a political message. An important feature of his analysis was that the topics targeted by the messages varied across the experiments, as did various other features of the messages and experiments (e.g., message length, delivery format, experiment subject pools, etc.). Coppock created plots of the average treatment effects of the messages by different subgroups, and they most closely resemble Figure 1B; that is, he finds that the average effects of the messages are strongly positively correlated across subgroups. For example, he estimates a correlation coefficient of 0.82 for the treatment effects across messages for subgroups defined by ideology: when the effects of the messages were large for liberals, they very often also tended to be large for conservatives. Similarly-large correlations were reported for subgroups defined by partisanship, race, education, and gender. Contrary to influential theories of political attitude formation, and the apparent prevalence of message-targeting in the political advertising industry, these results suggest that the effects of political messages are best characterized by rank-*homogeneity*, not heterogeneity.

However, while insightful, the foregoing analysis is limited in two crucial ways. First, by virtue of being taken from existing experiments, the political messages targeted a variety of different policy issues and were delivered in a variety of different formats (e.g., text vignettes, long-form op-eds, videos, etc.). This plausibly inflates the correlation in message effects across subgroups, because attitudes on some issues are generally easier to change than others (e.g., issues that are lower salience), and because some message formats may be generally more engaging than others. This diminishes the theoretical implication of the observed rank-homogeneity, because theories of heterogeneity typically argue that the “match” between message content and subgroup

characteristic is what is most relevant—not the policy issue to which the message speaks, nor the format in which it is delivered. For example, moral reframing theory holds that messages must appeal to the values of the audience in order to persuade; when there is a discordance between message frame and audience values, the effect of the message is diminished. Thus, a strong test of rank-homogeneity is to hold fixed the policy issue and delivery format while varying only the message *content*.

A second and related limitation of Coppock’s analysis is that the political messages were not designed to vary over dimensions that, according to theory, are most relevant for inducing subgroup heterogeneity. This is crucial because a lack of variation over theoretically-relevant message dimensions could also inflate the correlation in message effects across subgroups. For example, if the messages lack variation in the types of moral arguments that are theorized to appeal to liberals versus conservatives then it is perhaps unsurprising to observe minimal heterogeneity by ideology. Thus, a strong test of rank-homogeneity is to vary message content along dimensions identified by theory as most relevant for inducing subgroup heterogeneity.

In this paper, we conduct two large-scale randomized survey experiments designed to address the above limitations, comprising two policy issues, 59 message treatments and over 40,000 U.S. adults in total. Importantly, half of the messages are modelled explicitly on moral reframing theory. Our experiments thus provide (1) a stronger test of whether people are receptive to *similar* or *different* features of political messages in general, and (2) a new and comprehensive test of moral reframing theory.

Methods

Policy issues

Each of our two experiments focuses on attitudes towards a single policy issue. The policies are *Universal Basic Income* (UBI; Experiment 1) and the *U.S. Citizenship Act of 2021* (Experiment 2), an immigration reform bill introduced by Joe Biden on his first day in office. We chose these policy issues for several reasons. First, they are both somewhat low salience (at the time the experiments were conducted), meaning respondents would be less likely to have “made up their mind” on the issue. This increases our chance of detecting message effects. Second, both issues are relatively complex and thus amenable to a variety of different arguments—enabling us to develop and test a large diversity of different message treatments. Finally, the two issues are different in character: UBI may be seen primarily as an economic issue, while the U.S. Citizenship Act is more akin to a social issue because it relates to immigration policy. Studying both issues allows us to better assess the generalizability of our results.

Message treatments

We created a large sample of message treatments for each experiment by first searching online news sources and articles for arguments relating to the policy in question. We subsequently used these arguments to write a set of messages supporting/opposing the policy, exemplifying different persuasive strategies of interest. In particular, as mentioned previously, we drew heavily upon moral foundations/reframing theory, developing half our messages using this framework. We modelled the remaining messages after a variety of other dimensions that commonly appear in theories of political persuasion—including appeals to religion, scientific/historical evidence, expert opinion, public opinion, commonsense reasoning, moral arguments based on

liberty, appeals to compromise, and ad-hominem attacks on those supporting/opposing the policy (Blumenau and Lauderdale 2021). All messages were developed to exemplify a single persuasive strategy, and were coded as such prior to testing. Finally, each message was then edited into a short video treatment comprising 1 or 2 background images in a slideshow, a voiceover with subtitles, and music that was common to all videos. We used video treatments to facilitate engagement.

To illustrate the treatments, Figure 2 shows a screenshot from a pro-UBI treatment (Experiment 1), whose argument was coded as appealing to the “care/harm” moral foundation. The full argument transcript of this treatment reads:

The idea is seen by supporters as a way to live up to a compassionate ideal that society, as a first priority, should look out for its people’s survival. The COVID-19 pandemic has exposed vulnerabilities of huge populations, and without a social safety net, the social costs can be unbearably high. Universal Basic Income can provide citizens with the money to get necessities like food, to ensure that, at the very least, nobody should be made to go hungry or homeless after losing their jobs.

The full video can be viewed at https://mit.edu/~lbh/www/ubi_for-care.mp4. The complete list of arguments used in each treatment is in Appendix sections 0.6 and 0.7.



Figure 2. Screenshot from pro-UBI message treatment (Experiment 1).

Experiment design

To estimate the persuasive effect of our video treatments, in each experiment we recruited U.S. adults online via a survey platform that is a common supplier to the Lucid platform (Coppock and McClellan 2019). While our sample is thus a convenience sample and unrepresentative of the general U.S. population, this does not undermine the generalizability of our estimates: formal comparisons between treatment effects estimated using samples from Lucid or Mechanical Turk, versus those in national samples, indicate the two often correspond quite closely (Coppock 2019; Mullinix et al. 2015). Notably, our samples were approximately balanced on ideology and party identification (demographic characteristics of our sample are reported in Appendix section 0.1 that contains balance and differential attrition checks). Finally, studies conducted since the COVID-19 pandemic have found that online survey respondents are recently more diverse and representative than they were before, but

also substantially less attentive (Aronow et al. 2020). Thus, respondents were required to pass an attention check question before entering our experiments.

After answering an initial survey containing demographic covariates, respondents were randomly assigned to condition: either (1) a control group, (2) a single argument for or against the policy, or (3) two arguments: one for and one against the policy. Respondents in all conditions viewed a single video (that is, all videos contain a common introduction describing the policy, followed by zero, one, or two arguments). We included the two-sided treatment condition to improve the generalizability of our results—as the public may often be exposed to conflicting arguments on an issue—and to improve the robustness of our experiment to possible demand characteristics. An important point for our analysis is that respondents who received both a for- and against-argument were assigned each argument independently. Finally, after viewing the video, respondents were asked to rate their attitude towards the policy on a 7-point Likert scale. Below we briefly detail the specific policy description, set of treatment conditions, and outcome question used in each experiment.

Experiment 1: Universal Basic Income (UBI)

Date fielded: 2021-01-02.

Sample size: A total of 17,418 U.S. adults were assigned to watch a video in the UBI experiment.

Treatment assignment: This experiment tested 10 arguments in favor of UBI and 10 arguments against it. Each respondent was assigned to see one (or no) argument in favor of the policy and independently assigned to see one (or no) argument against the policy. When two arguments were assigned to one respondent, their order in the video was randomized—producing a total of 221 possible conditions.

Video transcript: “*Universal Basic Income is a policy proposal that's recently gained public attention. But what does it mean? A Universal Basic Income would replace the current welfare system, so the government would cut all the existing means-tested programs like food stamps, or earned income tax credit and instead pay a fixed amount to everybody in the United States, with a monthly cheque of \\$1000 for every citizen.* [TREATMENT ARGUMENTS]. *What do you think? Should the United States implement a Universal Basic Income?*”

Outcome variable: “Do you think the U.S. federal government should create a Universal Basic Income of \$1000 per month for every citizen?” (7-point Likert scale: [1] Definitely no - [4] Not sure - [7] Definitely yes).

Experiment 2: U.S. Citizenship Act

Date fielded: 2021-07-08.

Sample size: A total of 26,472 U.S. adults were assigned to watch a video in the U.S. Citizenship Act experiment.

Treatment assignment: This experiment tested 26 arguments in favor of the U.S. Citizenship Act and 13 arguments against it. Each respondent was assigned to see one argument in favor of the policy, or to a control group. Then, among only those who were assigned a for-argument, a subset was also assigned an argument against the policy. When two arguments were assigned to one respondent, their order was randomized—producing a total of 703 possible conditions (including control, for-only, and for + against). Excluding the “against-only” condition from the design of Experiment 2 allowed us to test a wider variety of arguments (in the “for” direction)—and thus better characterize the variability between them—at the expense of adding some additional complexity to our analysis (as described in the Results section).

Video transcript: “*What is the U.S. Citizenship Act, and why is it important? The U.S. Citizenship Act is an immigration bill introduced to congress this year. Among other things, the bill would increase the yearly limit on visas; create pathways to citizenship for many undocumented immigrants; and invest money in smart border control technology. [TREATMENT ARGUMENTS]. What do you think? Should congress pass the U.S. Citizenship Act?*”

Outcome variable: “Do you support or oppose the US Citizenship Act?” (7-point Likert scale: [1] Strongly oppose - [4] Not sure - [7] Strongly support).

Results

Examining rank-heterogeneity

We begin by estimating the correlation between message effects across different subgroups of people in order to identify the extent of rank-heterogeneity (versus rank-homogeneity) in each experiment. Thus, we estimate the conditional average treatment effect (CATE) of each treatment for the demographic subgroups of age, gender, partisanship, and ideology, and we plot the estimates in a scatterplot.¹ The estimates are computed separately for each subgroup in a linear regression which includes both for- and against-treatments, and all 3 remaining demographic variables as covariates to improve precision (Gerber and Green 2012). Due to our two-sided treatment experiment design, the interpretation of the estimates differs slightly between experiments. For UBI, treatment effects are estimated in the context of a (randomly present or absent) counterargument. For the U.S. Citizenship Act, the effects of for-

¹ In Appendix section 0.1, we present balance and differential attrition checks in which we (i) provide demographic characteristics of our sample and (ii) confirm that the effects we report are not likely to be driven by any differences in sample composition between treatment conditions.

arguments are estimated when those arguments are viewed alone, while the effects of against-arguments are estimated in the presence of a random for-argument.

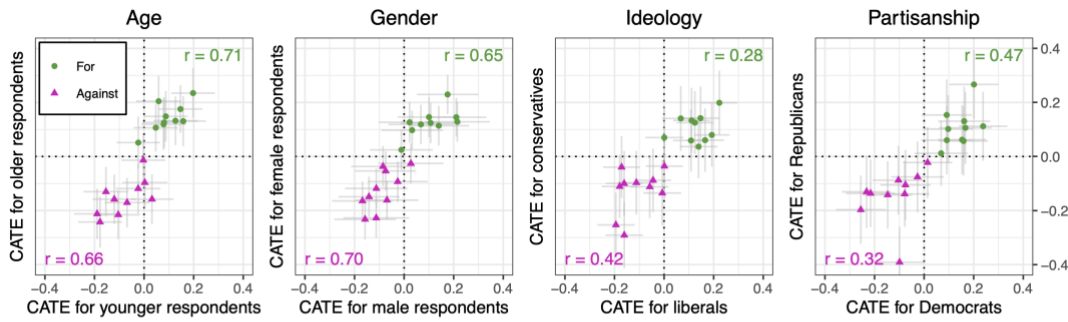
Figures 3A and 3B show the scatterplots of the estimates for UBI and the U.S. Citizenship Act, respectively. All the estimates are standardized.

To a first approximation, we find limited evidence of rank-heterogeneity of message effects by age or by gender in either experiment. On the contrary, the effects of the messages appear relatively well correlated for respondents older than 40 vs. younger, and for male vs. female respondents. These correlations are all the more impressive given that the individual message effect estimates contain substantial noise, which will tend to attenuate the correlation towards zero.² Furthermore, for UBI we find evidence of only moderate rank-heterogeneity by partisanship and ideology: broadly speaking, message effects remain positively correlated for Democrats vs. Republicans, and for liberals vs. conservatives. This occurs despite the fact that our message treatments were designed explicitly to include features that theory says should strongly distinguish these subgroups. Notably, we also observe no evidence of backfire: the for-arguments generally increased support among all subgroups, while the against-arguments generally decreased support among all subgroups on average.

In sum, the results from our UBI experiment appear to be best characterized by a pattern of rank-*homogeneity*, most similar to the pattern depicted in Figure 1B. In contrast to theories of heterogeneity, various different types of people—including Democrats and Republicans, as well as liberals and conservatives—were receptive to broadly similar features of messages regarding the issue of Universal Basic Income.

² Disattenuated correlations are reported in Appendix section 0.4.

A. Universal Basic Income experiment



B. US Citizenship Act experiment

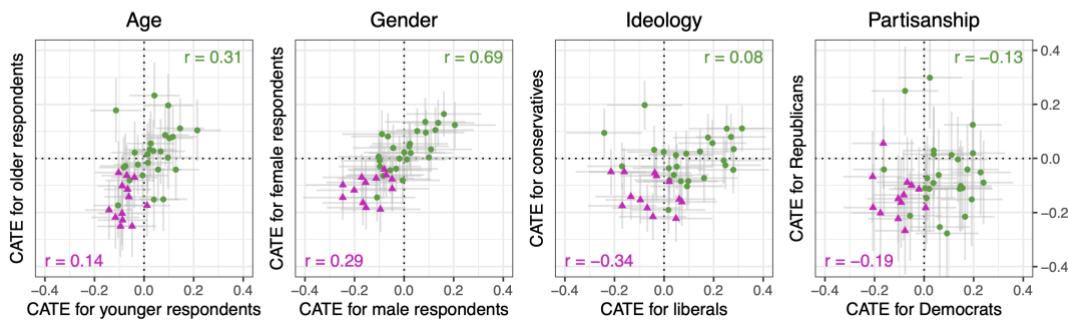


Figure 3. Standardized conditional average treatment effects (CATEs) of all message treatments among demographic subgroups in both experiments. Error bars are 95% CI.

However, the results from our experiment on the U.S. Citizenship Act paint a strikingly different picture (Figure 3B). In particular, far from the moderate positive correlations observed in the UBI experiment, here the correlations between message effects are substantially weaker for subgroups of age, and typically *negative* in the case of ideology and partisanship subgroups. Thus, the messages with the largest persuasive effects for liberals/Democrats were often among those with the weakest effects for conservatives/Republicans. The results of this experiment are therefore best characterized by rank-*heterogeneity*, most similar to that depicted in Figure 1C.

The lack of positively-correlated message effects among partisans and ideologues in the U.S. Citizenship Act experiment is consistent with theories which hold that, at least under some conditions, different types of people are indeed receptive

to different features of political messages in general. Given this result, a central question of interest is whether the rank-heterogeneity we observe among liberals and conservatives in particular is explained by their differential receptiveness to messages that were based upon the moral foundations—as predicted by moral reframing theory.

Examining Figure 3B offers suggestive evidence that this may be the case. In the upper left quadrant of the Ideology panel, two for-arguments stand out as having unusually large treatment effects on conservatives, yet lower-than-average effects on liberals. These arguments were based on the moral foundations of *sanctity* and *authority*. The fact that they rank substantially more persuasive among conservatives is consistent with moral reframing theory, because both are classified as “binding” foundations; those theorized to be particularly persuasive among conservatives. Nevertheless, properly testing the theory’s prediction requires testing (1) whether this pattern holds for every single morally-framed argument, and (2) whether it holds *averaging across* all such arguments. In the next section, we examine these questions.

Explaining rank-heterogeneity by testing moral reframing theory

We conduct two analyses to examine whether rank-heterogeneity among liberals and conservatives is explained by their differential receptiveness to messages that were based upon the moral foundations, as predicted by moral reframing theory. First, we use OLS to estimate the average treatment effect of messages that contain moral frames that are aligned vs. misaligned with respondents’ self-reported ideology. Recall that, according to the theory, “individualizing” frames are aligned for liberals, whereas “binding” frames are aligned for conservatives. This analysis provides a basic overall picture of whether aligned messages are more persuasive on average, collapsing across all the relevant (i.e., morally framed) messages. Second, we fit a multilevel linear

regression (Gelman and Hill 2006) in order to quantitatively examine the effect of *individual* messages. This allows us to ask whether *all* the morally-framed messages conform to the predictions of moral reframing theory, or whether only some do.

The estimates from our first analysis—the OLS regressions—are shown in Figure 4. The estimates are broadly consistent with the theory: on average, messages that used moral frames aligned with respondents’ ideology were qualitatively more persuasive—in both experiments and in both directions (*for* and *against*). However, the pattern is much more pronounced and only statistically significant at the .05 level in the U.S. Citizenship Act experiment. This difference between the experiments implies that the greater rank-heterogeneity we see in the U.S. Citizenship Act experiment may indeed be partly due to the fact that liberals and conservatives were differentially persuaded by the moral frames on average, as predicted by the theory.

To investigate whether this is true of every individual morally-framed message in our set, we proceed to fit a multilevel linear regression model to the data. This is preferable to using OLS regression to estimate the individual treatment effects, because some amount of the variability in the individual treatment effect estimates is due to sampling variability rather than true underlying variation in the treatment effects. The multilevel model takes this into account, thus providing estimates of the effect of each individual message that are more accurate on average (McElreath 2020, chapter 13).

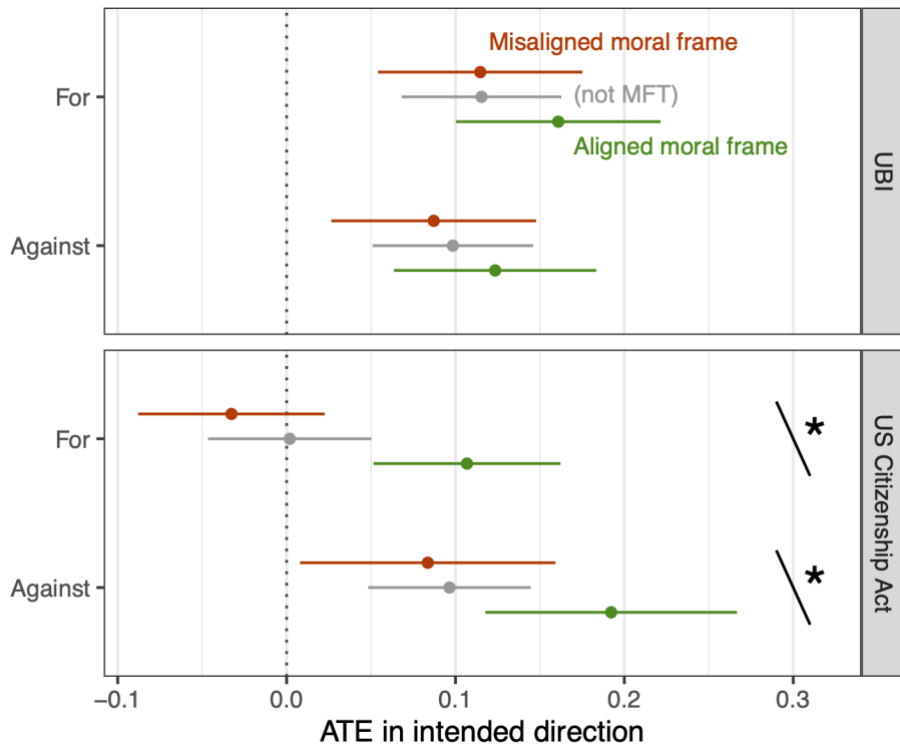


Figure 4. Average treatment effects of messages, disaggregated by whether the moral frame is aligned or misaligned with the respondent's ideology (moderates excluded). OLS regressions were fitted separately on liberals and conservatives and then the estimates were averaged, to ensure that ideology and frame were orthogonal. Asterisks indicate a (95%) significant difference between aligned and misaligned frames.

The informal description of the multilevel model is that it estimates respondents' policy support using fixed effects for (i) whether they were treated (by either a *for-* or *against-* argument) vs. in the control group, (ii) their self-reported ideology, and (iii) a linear interaction between (i) and (ii). We specify random effects for the individual treatments, allowing (i) and (iii) to vary across them. Since it would be incoherent to pool the *for-* and *against-*treatments together, we include separate fixed and random effects for each of these treatment types. The formal specification of the model is:

$$\begin{aligned}
Y_i &= \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta_1 \text{ideo}_i + \beta_{2j[i]} \text{for}_i + \beta_{3k[i]} \text{against}_i \\
&\quad + \beta_{4j[i]} (\text{for} \times \text{ideo})_i + \beta_{5k[i]} (\text{against} \times \text{ideo})_i \\
\begin{bmatrix} \beta_{2j} \\ \beta_{4j} \end{bmatrix} &\sim \text{MVNormal} \left(\begin{bmatrix} \beta_2 \\ \beta_4 \end{bmatrix}, \Sigma_1 \right) \\
\begin{bmatrix} \beta_{3k} \\ \beta_{5k} \end{bmatrix} &\sim \text{MVNormal} \left(\begin{bmatrix} \beta_3 \\ \beta_5 \end{bmatrix}, \Sigma_2 \right)
\end{aligned}$$

Where Y_i is policy support (standardized), subscript i indexes observations, j indexes the for-treatments, k indexes the against-treatments, and Σ is a covariance matrix. The variable *ideo* is scaled to have zero mean and unit range so that the interaction coefficients are interpretable as the difference in treatment effects between “very conservative” and “very liberal” respondents; the *for* and *against* variables are dummy variables for whether the respondent received a for- or against-treatment, respectively. We fit the model in a Bayesian framework, and specify weakly-informative prior distributions over all model parameters—allowing the data to “speak for itself.” (Prior distributions and model diagnostics are in Appendix section 0.2).

Table 1 shows a summary of the model results. The results confirm that the average treatment effects were in the expected direction: in both experiments, the *for* dummy is positive and the *against* dummy is negative. Furthermore, there was substantial variability across treatments in both experiments—shown by the models’ estimate of the standard deviation (SD) in the treatment effects. For example, the average treatment effect of the treatments in favor of UBI was 0.11, with an estimated standard deviation of 0.06 across the treatments. Using the mean \pm 2SDs as a heuristic,

this implies that we should expect the true effect of the messages in favor of UBI to range from zero ($0.11 - 2*0.06 = -0.01$) to larger than 0.20 ($0.11 + 2*0.06 = 0.23$).

Moving onto the interaction between the treatments and ideology, there is some evidence of interaction effects on average across treatments, but only one of these average effects is significant at the 0.05 level: *For x Ideology* in the U.S. Citizenship Act experiment. Notably, however, all of the average interaction effects are negatively signed. This tells us that for-treatments tended to produce the greatest persuasive effects among liberals, while against-treatments tended to produce the greatest persuasive effects among conservatives. This highlights the potential pitfalls of considering only individual interaction effects when drawing conclusions about heterogeneity—as we outlined in the introduction. That is, a for-treatment with a larger persuasive effect among liberals vs. conservatives may still be an unusually weak treatment among liberals, and an unusually strong treatment among conservatives.

Table 1. Summary of results from multilevel models.

Term	UBI	U.S. Citizenship Act
Ideology	-0.72 (0.06)	-0.68 (0.05)
For	0.11 (0.03)	0.03 (0.02)
Against	-0.12 (0.03)	-0.12 (0.02)
For x Ideology	-0.12 (0.07)	-0.21 (0.09)
Against x Ideology	-0.11 (0.08)	-0.12 (0.10)
SD (For)	0.06 (0.02)	0.07 (0.01)
SD (Against)	0.08 (0.03)	0.03 (0.02)
SD (For x Ideology)	0.06 (0.05)	0.34 (0.06)
SD (Against x Ideology)	0.12 (0.08)	0.27 (0.09)
Cor (For, For x Ideology)	0.03 (0.55)	-0.10 (0.24)
Cor (Against, Against x Ideology)	0.24 (0.46)	0.20 (0.48)

Note. Standard errors are in parentheses. SD = standard deviation; Cor = correlation.

Now we move onto our main quantity of interest: the models' estimate of the standard deviation in interaction effects across treatments: $SD(For\ x\ Ideology)$ and $SD(Against\ x\ Ideology)$. This quantity tells us how variable the effects of individual treatments were for liberals vs. conservatives. In the U.S. Citizenship Act experiment, this variability is substantial — $SD(For\ x\ Ideology) = 0.34$ — swamping the average interaction effects observed in that experiment. For example, despite the for-treatments being more persuasive among liberals on average ($For\ x\ Ideology = -0.21$), some individual for-treatments were nevertheless substantially more persuasive among conservatives (i.e., the mean + 2SD = $-0.21 + 2*0.34 = 0.47$). This was not the case for UBI. Consistent with Figure 3A, in the UBI experiment the variability was much more muted; interaction effects, where they existed, were mostly predictable based only on the direction of the treatment rather than its content.

Our key question is whether the variability in treatment effects across liberals and conservatives coheres with the predictions of moral reframing theory. In other words, is it true that (i) every single treatment that uses a “binding” frame is more persuasive among conservatives than liberals, and (ii) every single treatment that uses an “individualizing” frame is more effective among liberals than conservatives?

To answer this question, we examine the model-estimated interaction effects for each of the 59 individual treatments tested in our two experiments (Figure 5). The interpretation of the effects is: the difference in each treatment's effect size between “very liberal” and “very conservative” respondents, after adjusting for the mean effect size of each subgroup. Positive values indicate that the treatment was ranked more persuasive among conservatives than among liberals, while negative values indicate the treatment was ranked more persuasive among liberals than among conservatives.

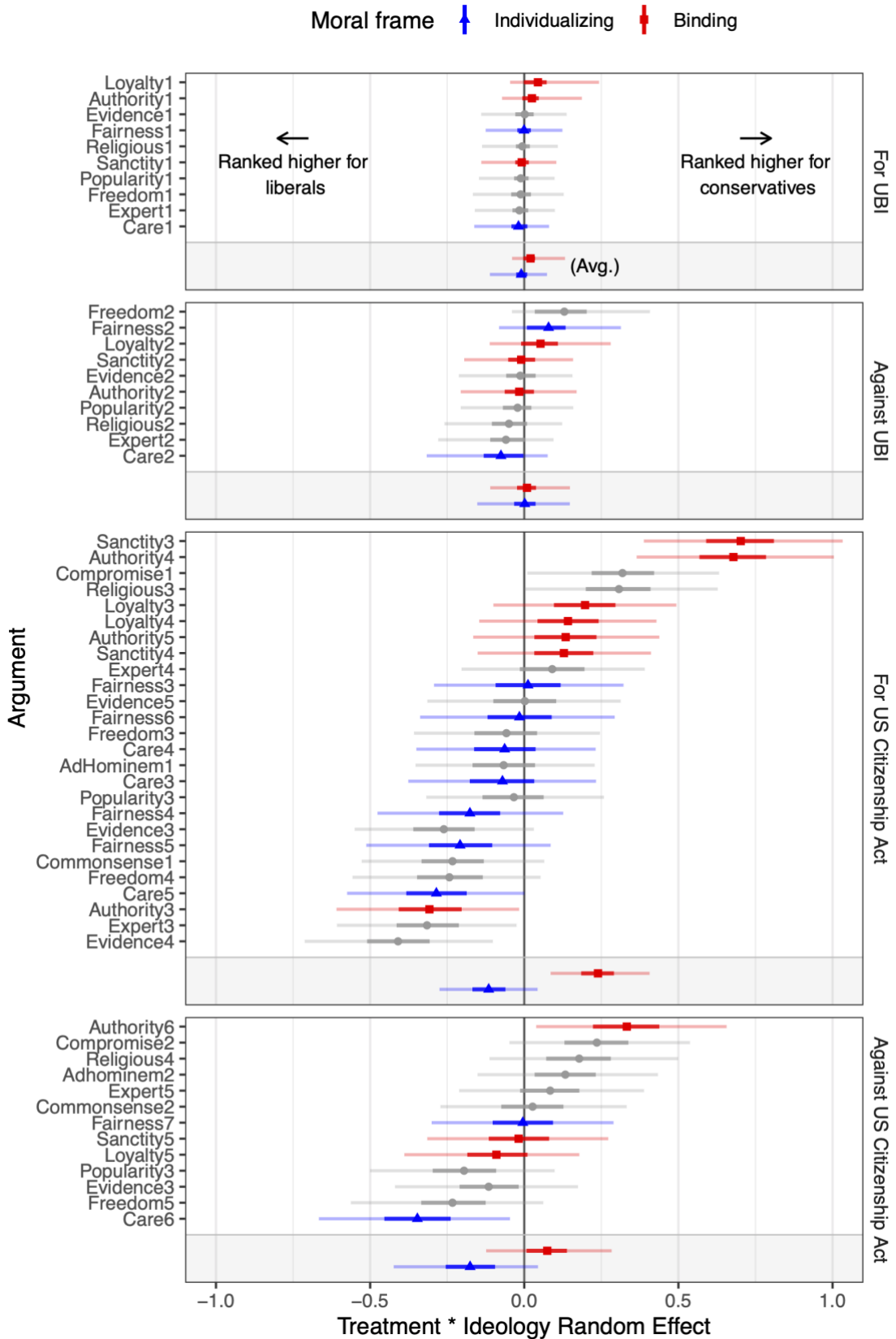


Figure 5. Interaction effects for each argument (treatment) from the multilevel models. Coefficients are interpretable as the difference in treatment effect size between “very liberal” and “very conservative” respondents, after adjusting for the mean effect size of each subgroup. Error bars are 95% and 50% credible intervals.

Treatments that appealed to the “binding” moral foundations are colored red, while those that appealed to the “individualizing” foundations are colored blue (treatments that made other types of appeal are in grey).

The estimates in Figure 5 provide mixed support for moral reframing theory. In particular, examining the treatments from the U.S. Citizenship Act experiment shows that there are some treatments that clearly map onto the predictions of the theory. For example, “binding” treatments *Sanctity3* and *Authority4* ranked substantially more persuasive among conservatives than liberals, while an “individualizing” treatment like *Care5* ranked more persuasive among liberals than conservatives.³ Furthermore, averaging across all treatments, the pattern predicted by theory is generally borne out. However, a number of treatments depart from the theory’s predictions. For example, *Authority3* actually ranked higher among liberals than conservatives, and several other treatments ranked similarly among both sets of respondents—such as *Sanctity5* and *Loyalty5*—despite theory implying they should rank higher among conservatives.

We thus draw two main conclusions from the results reported in this section. First, the greater rank-heterogeneity observed in the U.S. Citizenship Act (vs. UBI) experiment does indeed map onto the predictions of moral reframing theory, *on average across treatments*. Second, however, we observe a number of cases where the effects of individual treatments clearly depart from the theory’s predictions. This qualifies support for the theory, and highlights an important implication of our results for generalizability—which we return to in our concluding comments below.

³ Recall that the treatment texts can be viewed in Appendix sections 0.6 and 0.7.

Conclusion

In this paper, we aimed to advance understanding of the extent to which people are receptive to similar or different features of political messages in general. To that end, we conducted two large-scale experiments to study how correlated message effects are between different subgroups of U.S. adults, across a large and diverse set of messages on the same issue. A strong positive correlation suggests that people are primarily susceptible to similar features of messages when updating their attitudes, and this shared susceptibility swamps any differences in how they respond to individual messages. We call this pattern *rank-homogeneity*. In contrast, a weak, null or negative correlation indicates that people are primarily susceptible to distinct features of messages when updating their attitudes. We call this pattern *rank-heterogeneity*.

Does rank-homogeneity or rank-heterogeneity better characterize people's response to political messages? The results of our experiments provide nuanced and revealing evidence on this question. In the UBI experiment, message effects were moderately to strongly positively correlated across various demographic subgroups, including partisanship and ideology. This is significant because the messages were designed explicitly and on the basis of theory to induce heterogeneity among these subgroups. Thus, the results offer solid evidence of rank-homogeneity. By contrast, the U.S. Citizenship Act experiment revealed weak negative correlations among these subgroups—thus, the most persuasive messages for conservatives were among the least persuasive for liberals, and vice versa. This offers strong evidence of rank-heterogeneity. Furthermore, we found evidence of a message-level cause of this heterogeneity, consistent with moral reframing theory: messages ranking more persuasive among liberals typically employed “individualizing” frames, whereas those ranking more persuasive among conservatives typically employed “binding” frames.

Our results have several implications for persuasion theory and practice.

First, the contrasting pattern of results between our experiments could be due to a number of factors. However, the clearest and perhaps most obvious candidate is the different policy issues that were targeted (UBI vs. U.S. Citizenship Act). A growing body of evidence suggests that the effects of political messages are prone to considerable variation across policy issues (Blumenau and Lauderdale 2021; Clifford, Leeper, and Rainey 2021; Tappin 2020). Our results are consistent with this evidence, and further demonstrate the importance of including multiple policy issues or topics in studies on political communication and persuasion. This is important for building cumulative and generalizable knowledge, and for avoiding a published literature that consists of many studies that canvas a small number of policy issues and reach different, mutually exclusive conclusions due to an idiosyncratic sample of issues (Yarkoni 2020). Of course, in a design like ours, the resources required to conduct such studies can sometimes be substantial. The field may thus need to hasten its shift toward larger scale collaborations to meet these requirements (e.g., Moshontz et al. 2018).

A related implication of our results arises from the variability in treatment effects observed for the U.S. Citizenship Act issue. Specifically, despite the predictions of moral reframing theory being supported *on average across treatments*, we observed a number of cases where the effects of individual treatments clearly departed from the theory's predictions (e.g., an appeal to authority ranking more persuasive among liberals than conservatives). The implication is that, when theories make predictions about the effects of a *latent* treatment—such as “appeals to authority will be more persuasive among conservatives than liberals”—the effects of particular *instantiations* of that treatment may still depart from the theory's prediction. The upshot is that tests of theory will be more generalizable when they rely on multiple instantiations of the

latent treatment, as we did here (see also Blumenau and Lauderdale 2021). Thus, our results offer stronger evidence in favor of moral reframing theory than previous studies, which tend to rely on just one or two treatments from a single moral foundation.

Our results are broadly consistent with the idea that U.S. conservatives and liberals hold different patterns of moral values (Graham et al. 2013), explaining why they are receptive to different moral arguments on some issues. However, we hasten to add that there are other explanations of the mechanism here. For example, another possibility is that moral arguments provide information about which groups (e.g., political parties) support the policy in question. Perhaps for-arguments that appeal to “binding” values are more diagnostic of Republican than Democratic Party support, explaining why conservatives are more receptive to such arguments.⁴ However, we leave it to future research to more rigorously investigate the mechanism(s) underlying moral reframing’s impact on rank-heterogeneity.

Our results additionally offer insights for political communication in practice: for some political issues and demographic subgroups, targeting messages to different subgroups may not be worthwhile because the rank-ordering of message effectiveness is similar. However, as our results also show, clearly this may not always be the case. Thus, perhaps the clearest practical implication of our results is that it is worthwhile to test the heterogeneous effectiveness of messages before disseminating them.

Finally, our results directly extend the analyses conducted by Coppock (2016, 2022). Recall that those analyses identified very strong positive correlations across subgroups—powerful evidence of rank-homogeneity—even for subgroups defined by partisanship and ideology. While our results qualitatively agree with those findings in

⁴ In Appendix section 0.3, we report some evidence of this possibility from a smaller-scale follow up study: heterogeneity was somewhat attenuated when people were told the positions of the parties on the U.S. Citizenship Act.

many cases, we did not consistently replicate such strong correlations. Our findings thus lend credibility to a concern described in the introduction, that the very strong correlations reported by Coppock may be somewhat inflated by variation in policy issues/message format across experiments, rather than the content of the messages themselves. That said, there are other possible explanations. The discrepancy may reflect a difference in the emphasis of the treatments: our experiments explicitly focused on arguments that theory suggests should induce subgroup heterogeneity, an emphasis that is arguably less pronounced in the treatments analyzed by Coppock. Furthermore, we focused on lower-salience policy issues for which many people may be unaware of party positions. It is possible that the heterogeneity we observe is due to respondents interpreting the arguments as implicit cues as to which parties support the policy. Therefore, future research should more thoroughly test whether and to what extent rank-heterogeneity is robust to the presence of explicit party cues, as well as how common it is among treatments (e.g., ads) produced by actual political campaigns.

References

- Aronow, Peter Michael, Joshua Kalla, Lilla Orr, and John Tervovski. 2020. *Evidence of Rising Rates of Inattentiveness on Lucid in 2020*. SocArXiv. preprint. <https://osf.io/8sbe4> (September 16, 2020).
- BBC. 2018. "Vote Leave's Targeted Brexit Ads Released by Facebook." *BBC News*. <https://www.bbc.com/news/uk-politics-44966969> (April 23, 2022).
- Blumenau, Jack, and Benjamin E Lauderdale. 2021. "The Variable Persuasiveness of Political Rhetoric." *American Journal of Political Science*. <http://benjaminlaurerdale.net/files/papers/BlumenauLaurerdalePersuasion.pdf>.
- Cadwalladr, Carole. 2017. "The Great British Brexit Robbery: How Our Democracy Was Hijacked." *The Guardian*. <https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy> (November 29, 2019).

- Clifford, Scott, Thomas J Leeper, and Carlisle Rainey. 2021. "Increasing the Generalizability of Survey Experiments Using Randomized Topics: An Application to Party Cues." *Working Paper*.
- Coppock, Alexander. 2016. "Positive, Small, Homogeneous, and Durable: Political Persuasion in Response to Information." Columbia University. <https://doi.org/10.7916/D8J966CS> (June 27, 2019).
- . 2019. "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach." *Political Science Research and Methods* 7(3): 613–28.
- . 2022. *Persuasion in Parallel*. Chicago: University of Chicago Press.
- Coppock, Alexander, and Oliver McClellan. 2019. "Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Research & Politics* 6(1): 2053168018822174.
- Dobber, Tom, Ronan Ó Fathaigh, and Frederik J. Zuiderveen Borgesius. 2019. "The Regulation of Online Political Micro-Targeting in Europe." *Internet Policy Review* 8(4). <https://policyreview.info/articles/analysis/regulation-online-political-micro-targeting-europe> (June 21, 2022).
- Feinberg, Matthew, and Robb Willer. 2013. "The Moral Roots of Environmental Attitudes." *Psychological Science* 24(1): 56–62.
- . 2015. "From Gulf to Bridge: When Do Moral Arguments Facilitate Political Influence?" *Personality and Social Psychology Bulletin* 41(12): 1665–81.
- . 2019. "Moral Reframing: A Technique for Effective and Persuasive Communication across Political Divides." *Social and Personality Psychology Compass*: e12501.
- Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Gerber, Alan, and Donald Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. W. W. Norton.
- Graham, Jesse et al. 2013. "Chapter Two - Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism." In *Advances in Experimental Social Psychology*, eds. Patricia Devine and Ashby Plant. Academic Press, 55–130. <https://www.sciencedirect.com/science/article/pii/B9780124072367000024> (April 19, 2022).
- Graham, Jesse, Jonathan Haidt, and Brian A. Nosek. 2009. "Liberals and Conservatives Rely on Different Sets of Moral Foundations." *Journal of Personality and Social Psychology* 96(5): 1029–46.
- Hornikx, Jos, and Daniel J. O’Keefe. 2009. "Adapting Consumer Advertising Appeals to Cultural Values: A Meta-Analytic Review of Effects on Persuasiveness and

- Ad Liking.” *Annals of the International Communication Association* 33(1): 39–71.
- Kahan, Dan M. 2016. “The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It.” In *Emerging Trends in the Social and Behavioral Sciences*, , 1–16.
<https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118900772.etrds0417>
 (June 4, 2019).
- Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. “Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence.” *Journal of Personality and Social Psychology* 37(11): 2098–2109.
- McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. CRC Press.
- Moshontz, Hannah et al. 2018. “The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network.” *Advances in Methods and Practices in Psychological Science* 1(4): 501–15.
- Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman, and Jeremy Freese. 2015. “The Generalizability of Survey Experiments*.” *Journal of Experimental Political Science* 2(2): 109–38.
- Privacy International. 2020. “Why We’re Concerned about Profiling and Micro-Targeting in Elections.” *Privacy International*.
<http://privacyinternational.org/news-analysis/3735/why-were-concerned-about-profiling-and-micro-targeting-elections> (November 26, 2021).
- Scott, Mark. 2018. “Cambridge Analytica Helped ‘Cheat’ Brexit Vote and US Election, Claims Whistleblower.” *POLITICO*. <https://www.politico.eu/article/cambridge-analytica-chris-wylie-brexit-trump-britain-data-protection-privacy-facebook/>
 (April 23, 2022).
- Taber, Charles S., Damon Cann, and Simona Kucsova. 2009. “The Motivated Processing of Political Arguments.” *Political Behavior* 31(2): 137–55.
- Taber, Charles S., and Milton Lodge. 2006. “Motivated Skepticism in the Evaluation of Political Beliefs.” *American Journal of Political Science* 50(3): 755–69.
- Tappin, Ben M. 2020. *Estimating the Between-Issue Variation in Party Elite Cue Effects*. PsyArXiv. preprint. <https://osf.io/p48zb> (October 8, 2020).
- Teeny, Jacob D., Joseph J. Siev, Pablo Briñol, and Richard E. Petty. 2021. “A Review and Conceptual Framework for Understanding Personalized Matching Effects in Persuasion.” *Journal of Consumer Psychology* 31(2): 382–414.
- Van Bavel, Jay J., and Andrea Pereira. 2018. “The Partisan Brain: An Identity-Based Model of Political Belief.” *Trends in Cognitive Sciences* 22(3): 213–24.

Yarkoni, Tal. 2020. "The Generalizability Crisis." *Behavioral and Brain Sciences*.
<https://osf.io/jqw35> (December 10, 2019).