Extremely Partisan Samples Distort Perceptions of Group Beliefs

Alexandra M. van der Valk[1,2], Alexander C. Walker[3], Jonathan A. Fugelsang[1], and Derek J. Koehler[1]

[1] Department of Psychology, University of Waterloo, Waterloo, ON, Canada

[2] Geriatric Medicine, Nova Scotia Health & Dalhousie University, Halifax, NS, Canada

[3] Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Providence, RI, USA

Author Note

Alexandra van der Valk https://orcid.org/0000-0001-5176-3516

Alexander C. Walker https://orcid.org/0000-0003-1431-6770

Jonathan A. Fugelsang https://orcid.org/0000-0002-6342-7023

Derek J. Koehler https://orcid.org/0000-0002-3596-7882

Correspondence concerning this article should be addressed to Alexandra van der Valk, 1314-5955 Veterans' Memorial Lane, Halifax, NS, Canada, B3H 2E1. Email: Alexandra.vanderValk@nshealth.ca

Abstract

How do we infer the beliefs of an entire group (e.g., Democrats) after being exposed to the beliefs of only a handful of group members? What if we know that the beliefs we encountered were selected in a biased manner? Across two experiments, we recruited 640 U.S. residents and assessed whether they could correct for known sample bias. Some participants viewed biased samples that exclusively presented the political beliefs of extreme partisans while others viewed representative samples free from selection biases. We find evidence that people correct for known sample bias, but do so insufficiently, leading them to make inaccurate inferences that are aligned with sample bias. Specifically, participants overestimated the ideological extremity of Democrats and Republicans to a greater extent when exposed to explicitly biased—compared to representative—samples. We discuss how peoples' tendency to insufficiently adjust from transparently biased samples leads to partisan misperceptions that amplify political polarization.

Keywords: social inference, partisan perceptions, bias, sample selection, political polarization

Research Transparency Statement

General Disclosures

Conflicts of interest: The authors declare no conflicts of interest. Funding: This research was supported by a Social Sciences and Humanities Council of Canada Banting Postdoctoral Fellowship (to ACW) and by grants from The Natural Sciences and Engineering Research Council of Canada (to JAF and DJK). Artificial intelligence: No artificial intelligence assisted technologies were used in this research or the creation of this article. Ethics: This research complies with the Declaration of Helsinki (2013). All experiments were reviewed and received ethics clearance from a University of Waterloo Research Ethics Committee. Computational reproducibility: The authors are applying for a Computational Reproducibility Badge which will be awarded pending checks by the STAR Team.

Experiment One

Preregistration: Experiment one was not preregistered. Materials: All study materials can be viewed in the supplementary materials (Part A). Data: All data are publicly available (https://osf.io/g9tmz/). Analysis scripts: All analysis scripts are publicly available (https://osf.io/g9tmz/).

Experiment Two

Preregistration: Experiment two was not preregistered. Materials: All study materials can be viewed in the supplementary materials (Part A). Data: All data are publicly available (https://osf.io/g9tmz/). Analysis scripts: All analysis scripts are publicly available (https://osf.io/g9tmz/).

Introduction

Political opinions are not hard to find. Politicians and political pundits regularly engage in public discussions surrounding polarizing issues. Meanwhile, social media bombards us with the political attitudes of friends, family, and strangers. People attend to what others believe and adjust their own beliefs in response. For example, a message informing Republicans that a majority of Republicans agree that the climate is changing increases the likelihood that they themselves endorse this belief (Bayes et al., 2020). How people *perceive* the attitudes of others similarly shapes their viewpoints. Overestimating the ideological extremity of others' increases the extremity of one's own political attitudes (Ahler, 2014). Likewise, exaggerating the degree to which opposing partisans dislike one's political in-group facilitates reciprocal feelings of out-group animosity (Moore-Berg et al., 2020). Fortunately, these misperceptions can be corrected, with interventions correcting partisan misperceptions reducing cross-party animus and the extremity of individuals' political views (Ahler, 2014; Lees & Cikara, 2020).

Despite the ubiquity of political expressions, people cannot directly observe the normative beliefs of a group. Rather, such beliefs must be inferred from available evidence. For example, when contemplating a polarizing issue, a Democrat cannot access the full distribution of beliefs of fellow Democrats, instead having to make inferences about this distribution by cobbling together observations of political expressions (e.g., from everyday conversations, social media, etc.). We can think of this set of observations as a sample of beliefs drawn from a population whose distribution we wish to infer.

Often, the samples available to us are subject to biased selection processes, resulting in them not being entirely representative of the population from which they are drawn. In the political domain, polarizing content captures attention, ensuring its overrepresentation on many platforms (Brady et al., 2020). On social media, most political content comes from users who are more politically engaged and ideologically extreme than the average person (Hughes, 2019). Individuals may be rewarded for expressing more polarizing views. For example, compared to their more moderate peers, American politicians with extreme ideological positions acquire more followers on social media (Hong & Kim, 2016). Similarly, politically-biased language increases the perceived trustworthiness of in-group speakers (Walker et al., 2024) while negative tweets about one's out-group are rewarded with enhanced audience engagement (Rathje et al., 2021). Social and monetary incentives rewarding polarizing political expressions ensure that the political attitudes people encounter—whether through social media or other mediums (e.g., cable news)—are frequently more extreme than those of the average person.

Exposure to biased samples may systematically distort how people perceive the political attitudes of others. On social media, the processes that amplify some political expressions—and not others—are often opaque, concealing biases in content selection. Thus, people may fail to realize that the polarizing views they encounter do not represent the attitudes of the population at large. Consistent with this claim, past work has linked consuming political content with less accurate (and more negative) perceptions of opposing partisans (Yudkin et al., 2019). Regardless of their source, the pervasiveness of partisan misperceptions is well-established. People not only overestimate the ideological extremity of in- and out-party members (Levendusky & Malhotra, 2016), but also the extent to which these individuals are politically engaged (Druckman et al., 2022) and belong to partisan-stereotypical groups (Ahler & Sood, 2018). Moreover, people overestimate the extent to which the average partisan dislikes and dehumanizes their political opponents (Moore-Berg et al., 2020) and underestimate the degree to which they agree with the views of out-party members (Dorison et al., 2019). These misperceptions have consequences, amplifying partisan animosities and deepening ideological divides (Lees & Cikara, 2021).

Ideally, if people were aware of sample bias, they would adjust their inferences about the population distribution to correct for it. Yet, people seldom know enough about the particular biases shaping sample selection to adequately correct for them, at times lacking awareness of sample bias altogether. The present research constructs an idealized scenario in which people are explicitly described the biases shaping the samples they encounter. Past work suggests that people fail to discount evidence based on knowledge that it was selected in a biased manner. For example, Hamill and colleagues (1980) found that even when participants were informed that they would be presented with an interview of a prison guard chosen for his unusual cruelty, they nevertheless generalized his characteristics to all prison guards. Thus, encountering biased samples of political opinion may give rise to partisan misperceptions even when people are explicitly described the biases involved in sample selection.

Across two experiments, participants estimated Democrat's [Republican's] average level of agreement with various political statements (e.g., "*The US has loose gun laws*") after viewing biased samples of agreement ratings from Democrats [Republicans]. Notably, participants were described a sampling process that selectively displayed the agreement ratings of more extreme partisans. As such, they were informed of the sampling process that produced biased samples. By comparing participants' estimates (i.e., of the average rating given by all Democrats/Republicans) to the sample mean, the true population mean, and to estimates from different comparison groups, we test whether people adjust for known sample bias and, if so, whether this adjustment adequately corrects for such bias. We hypothesized that participants' estimates would be less extreme than the mean agreement rating displayed in biased samples (indicating adjustment) but that they would be more extreme than the true population mean and estimates provided by participants viewing unbiased samples (indicating insufficient adjustment). Thus, we hypothesized that viewing biased samples featuring the political attitudes of extreme partisans would lead participants to overestimate the ideological extremity of the average partisan, even when sample biases were explicitly described. In this way, we investigate one mechanism (i.e., insufficient adjustment from biased samples) that can explain why people misperceive the political attitudes of others in ways that exacerbate partisan animosities and deepen political divides.

Experiment 1

Methods

Participants

We recruited 300 United States residents from Amazon Mechanical Turk (MTurk). To maximize data quality, participants were exclusively recruited from CloudResearch's pool of approved participants (Hauser et al., 2023) and were required to a) pass two pre-study attention checks and b) possess an MTurk approval rating equal to or greater than 95%. We excluded data from 19 participants who failed a post-task comprehension check and one participant who failed to provide sufficient study data, leaving data from 280 participants (60% Male; 135 Democrats, 66 Republicans, 75 Independents[1]) to be analyzed.

Materials

Experiment 1 featured 24 politically polarizing statements adapted from Vlasceanu and colleagues (2021). Of these 24 statements, twelve *Democrat-leaning* statements were shown to elicit agreement from Democrats and disagreement from Republicans (e.g., "*The US has loose*

[1] Four participants did not report an affiliation.

gun laws") while twelve *Republican-leaning* statements produced agreement from Republicans and disagreement from Democrats (e.g., "*The US justice system is fair to racial minorities*"). Vlasceanu and colleagues had a representative sample of 352 Democrats and 352 Republicans rate their agreement with each statement on a 101-point scale that ranged from 0 (*Completely disagree*) to 100 (*Completely agree*). In the present work, we presented participants with 12 Democrat-leaning [Republican-leaning] statements and asked them to estimate the average rating given by Democrats [Republicans] who participated in this survey (Vlasceanu et al., 2021; See Figure 1). Instructions described this survey to participants and informed them whether they would be shown Democrat- or Republican-leaning statements (see Supplementary Materials Part A). On each trial, participants were presented a statement and asked to "Please estimate the average agreement rating given by all [Democrat/Republican] participants in the survey." Participants provided these estimates using the aforementioned 101-point scale.

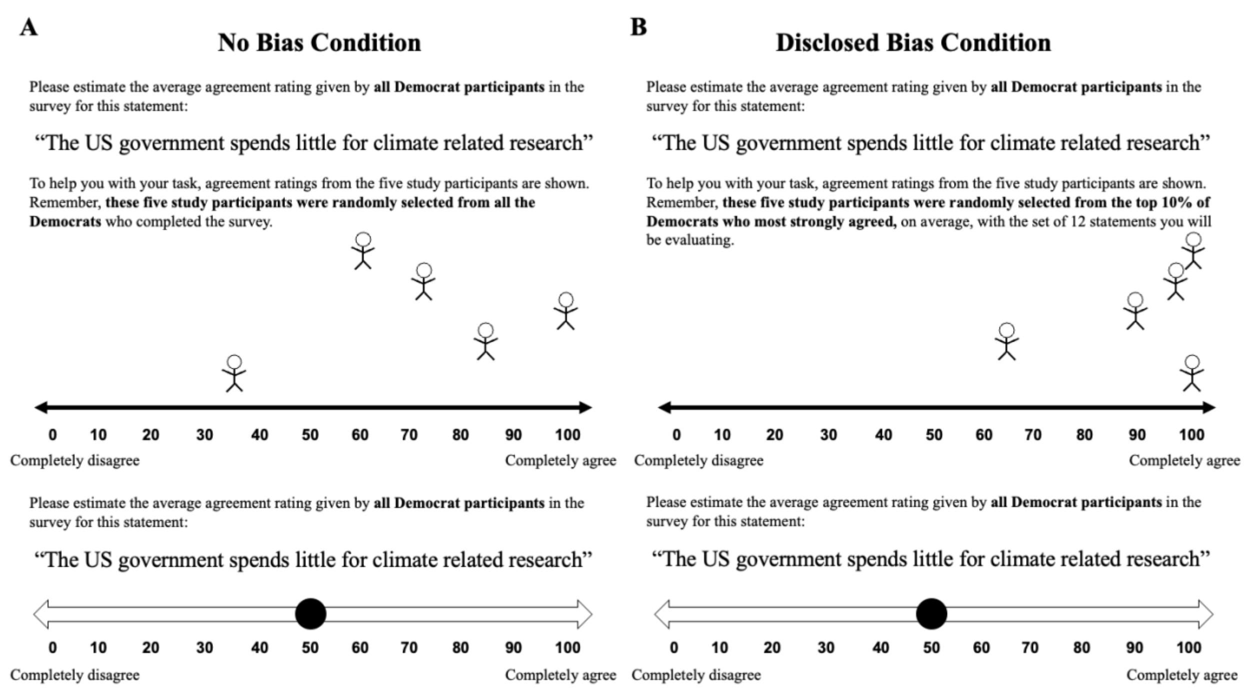

Figure 1. Example of an item presented to participants in No Bias (Panel A) and Disclosed Bias (Panel B) conditions.

With the exception of those randomly assigned to a No Sample condition, participants were presented the attitudes (i.e., agreement ratings) of five survey respondents on each trial.

No Bias Sample. For each statement, we obtained a sample of five agreement ratings that accurately reflected the attitudes of a target sample (i.e., Democrats or Republicans). Specifically, for each statement, we divided the ratings of target respondents into quintiles and then extracted the median of each quintile to obtain five agreement ratings that were representative of the attitudes endorsed within the target political party. For example, for each Democrat-leaning statement, we divided the ratings of Democrats into quintiles and depicted the ratings of five survey respondents whose attitudes represented the median rating present within each quintile (see Figure 1A). Prior to experimental trials, participants randomly assigned to view representative samples were informed that the samples they would be shown were selected in "an unbiased manner" and thus could be "considered representative of [Democrats/Republicans] survey respondents as a whole."

Disclosed Bias Sample. Experiment 1 also featured biased samples that, for each statement, depicted the attitudes of five more extreme partisans. When creating biased samples we considered only the attitudes of Democrats and Republicans whose mean agreement rating across ideologically-congruent statements was in the top 10% of respondents. That is, we considered only the attitudes of Democrats who most strongly endorsed Democrat-leaning statements and Republicans who most strongly endorsed Republican-leaning statements. Mirroring the procedure used to generate representative samples, we divided the ratings of these more extreme partisans into quintiles and then extracted the median rating of each quintile, resulting in ratings from five survey respondents that were representative of the attitudes endorsed among the top 10% of extreme Democrats (or Republicans) but unrepresentative (i.e., more extreme and homogenous) of the attitudes of the average partisan respondent. Prior to experimental trials, participants presented biased samples were informed that the samples they would be shown were "selected in a biased manner" and therefore could *not* be considered representative of [Democrat/Republican] survey respondents as a whole. Likewise, during each experimental trial, these participants were reminded that the survey respondents presented were "randomly selected from the top 10% of [Democrats/Republicans] who most strongly agreed, on average, with the 12 statements" they would be evaluating (see Figure 1B).

Measures

Belief Correction. For the two conditions in which participants were presented with a sample of agreement ratings, we calculated the difference between participants' estimates and the mean of the samples they had been shown. A belief correction measure was created by averaging these differences over all 12 statements. Positive [negative] values indicate that a participant's mean estimate was higher [lower] than the mean rating of the presented sampled respondents.

Accuracy. After making their 12 estimates, we assessed participants' beliefs about their accuracy. We aimed to assess whether participants shown biased samples would expect to be less accurate than those viewing unbiased samples. We defined a "hit" as a statement for which a participant's estimate of the average agreement rating among all Democrats [Republicans] fell within 10 points of the actual mean agreement rating among that group. Following this description, we asked participants: "Across your 12 estimates, how many hits do you think you scored?" Participants responded to this question by selecting a number between 0 and 12, with this response representing their subjective accuracy. We also calculated an objective accuracy score for each participant that was equal to the number of hits they achieved.

Perceived Party Consensus. At the end of the task, we asked participants to estimate the global probability (0-100%) that, for any particular issue, two randomly selected Democrats (for those viewing Democrat-leaning statements) or Republicans (Republican-leaning statements) would give agreement ratings that fall within 10 points of each other. Because participants in the Disclosed Bias condition were shown samples that were not only more extreme but also more homogenous than those viewing unbiased samples, we aimed to assess whether they would perceive greater consensus within political groups.

Design and Procedure

Experiment 1 featured a 2 x 3 between-subjects design. Participants were presented either 12 Democrat- or 12 Republican-leaning statements and estimated the extent to which the average Democrat or Republican, respectively, agreed with each statement. Furthermore, participants were randomly assigned to provide each estimate either without sample information, after viewing a representative sample of five survey respondents, or after viewing a biased sample of five more ideologically extreme respondents. Those presented with either representative or biased samples were informed of the relevant sample selection method and completed pre- and post-task comprehension checks. Following experimental trials, participants responded to accuracy and perceived consensus questions and provided demographic information.

Results

Mean agreement estimates—collapsed across Democrat- and Republican-leaning statements—in the Disclosed Bias, No Bias, and No Sample (control) conditions are shown in Figure 2. First, we ask if participants shown biased samples adjust their estimates in attempt to correct for sample biases, by comparing their belief correction scores to those of participants in the No Bias condition. Consistent with such adjustment, belief correction was more pronounced in the Disclosed Bias (M = -12.94, SD = 12.53) compared to the No Bias condition (M = 2.59, SD = 7.55), $t(179) = 9.65$, $p < .001$, $d = 1.45$, 95% CI [1.12, 1.78], with participants in the Disclosed Bias condition providing less extreme estimates than the mean of the biased samples they were shown, $t(103) = 11.10$, $p < .001$, $d = 1.09$, 95% CI [0.84, 1.33].

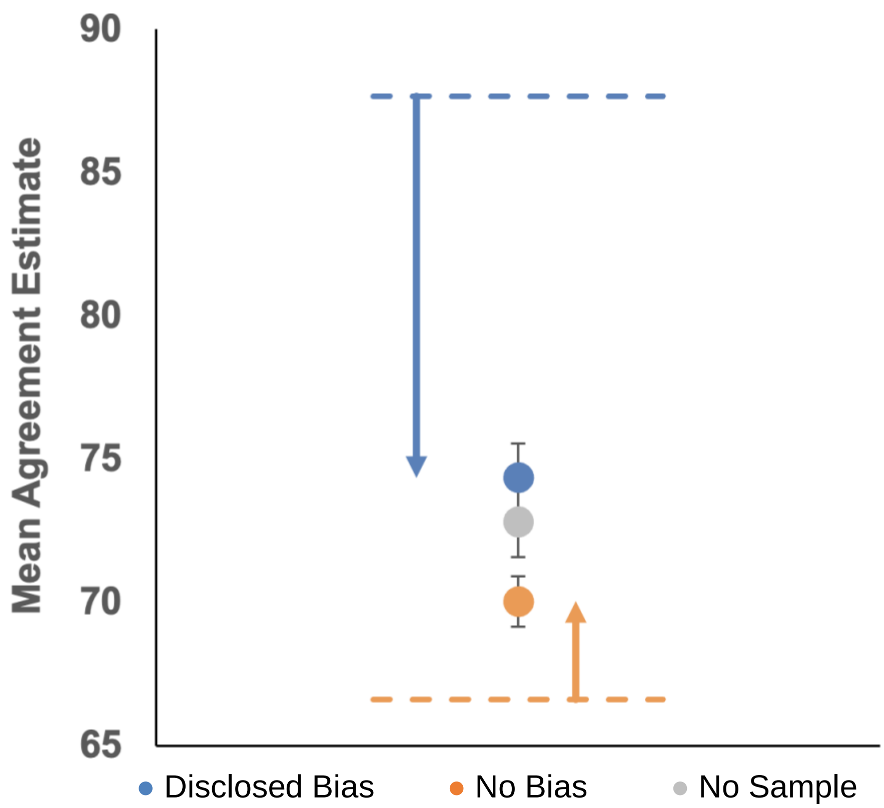

Figure 2. Experiment 1: Mean Agreement Estimates of the Average Partisan. Dashed lines represent the mean agreement ratings of survey respondents presented in biased (blue) and unbiased (orange) samples. Arrows depict the difference between sample means and participants' mean agreement rating estimates of the average partisan within a condition. Error bars represent +/- 1 SE.

Second, we assess whether adjustment in the Disclosed Bias condition was sufficient by comparing the mean estimate in this condition to that in the No Bias condition (and also the control, No Sample condition). While all participants overestimated the ideological extremity of the average partisan (p's < .001, d's > 0.45), this tendency was strongest for participants in the Disclosed Bias condition, suggesting that they failed to fully correct for sample bias, despite such bias being disclosed. One-way analysis of variance (ANOVA) revealed that estimated agreement ratings of the average partisan differed based on Sample Type (Disclosed Bias, No Bias, No Sample), $F(2, 277) = 3.26$, $p = .040$, $\eta_p^2 = .023$.[2] One-tailed independent samples *t*-tests indicated that estimates were more extreme when participants' viewed explicitly biased ($M = 74.33$, $SD = 12.25$) as opposed to unbiased samples ($M = 70.03$, $SD = 7.60$), $t(179) = 2.72$, $p = .004$, $d = 0.41$, 95% *CI* [0.11, 0.71]. However, while estimates in the No Bias condition were slightly more moderate than those in the No Sample condition ($M = 72.82$, $SD = 12.48$), $t(174) = 1.73$, $p = .043$, $d = 0.26$, 95% *CI* [0.04, 0.56], estimates did not reliably differ between Disclosed Biased and No Sample conditions, $t(201) = 0.87$, $p = .192$, $d = 0.12$, 95% *CI* [-0.15, 0.40]. Factorial ANOVAs exploring the effects of participant party affiliation and affiliation strength (along with Sample Type) showed no effects of these variables (*p's* > .344), suggesting that agreement estimates did not differ based on participants' political stance (see Supplementary Materials Part B).

Accuracy

Subjective and objective accuracy scores are illustrated in Figure 3. A one-way ANOVA revealed that subjective accuracy did not differ based on Sample Type, $F(2, 277) = 0.74$, $p = .480$, $\eta_p^2 = .005$. Participants shown biased samples did not expect their estimates to be any less accurate than did participants in the No Bias and No Sample conditions. However, *objective* accuracy reliably differed by Sample Type, $F(2, 277) = 59.01$, $p < .001$, $\eta_p^2 = .299$. Participants in the No Sample ($M = 3.62$, $SD = 2.31$) and Disclosed Bias ($M = 4.29$, $SD = 3.18$) conditions were similarly inaccurate in their agreement rating estimates of the average partisan ($p = .087$, *d*

[2] Across studies, the target party (Democrat or Republican) that participants were randomly assigned to evaluate exerted at most a minimal impact on their judgments. As such, all analyses including Target Party are exclusively reported in the supplementary materials (Part B).

= 0.24), and considerably less accurate than those in the No Bias condition (*M* = 8.25, *SD* = 3.44; *p's* < .001, *d's* > 1.20). Taken together, participants viewing representative samples of political attitudes (No Bias) slightly underestimated the accuracy of their estimates, *t*(76) = 3.07, *p* = .003, *d* = 0.35, 95% *CI* [0.12, 0.58], while those viewing no sample information or transparently biased sample information were considerably overconfident (*p's* < .001, *d*'s > 0.62).

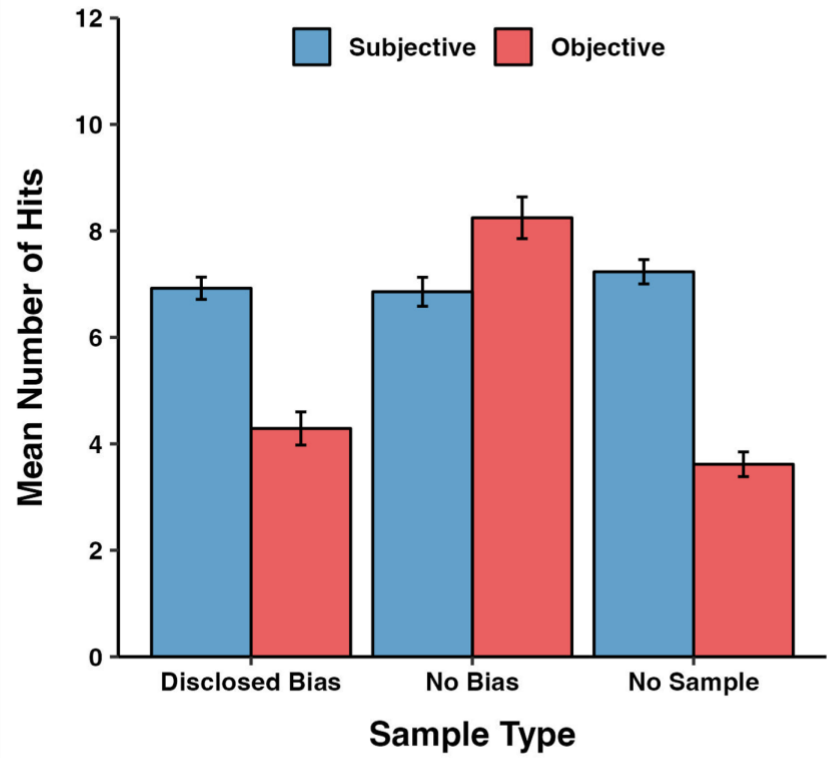

Figure 3. Experiment 1: Subjective and Objective Accuracy. Bars display participants' mean subjective and objective accuracy scores within the Sample Type condition for which participants were randomly assigned. Error bars represent +/- 1 SE.

Perceived Party Consensus

Perceptions of within-party consensus differed by Sample Type, *F*(2, 277) = 4.18, *p* = .016, η_p^2 = .029. Participants viewing biased samples (which tended to have less variance) perceived greater consensus (*M* = 68.22, *SD* = 19.33) among the attitudes of Democrats and Republicans compared to those viewing unbiased samples (*M* = 61.10, *SD* = 16.73), *t*(179) = 2.59, *p* = .010, *d* = 0.39, 95% *CI* [0.09, 0.69]. Perceptions of party consensus did not, however, differ between Disclosed Bias and No Sample conditions (M = 68.42, SD = 19.35), $t(201) = 0.08$, $p = .940$, $d = 0.01$, 95% CI [-0.27, 0.29].

Experiment 2

The results of Experiment 1 are consistent with the idea that participants shown biased samples deliberately adjust their inferences to compensate for sample bias. This would explain why their estimates fall farther from the mean of the samples they viewed compared to those shown unbiased samples. However, an alternative explanation is that participants in both conditions treated their samples as representative; on a Bayesian account, they may have adjusted from a prior in the direction of the sample mean they were shown. This would produce the same qualitative pattern of results as was observed in Experiment 1. As such, we lack definitive evidence that participants viewing biased samples deliberately corrected for known sample bias. Experiment 2 addresses this limitation by replacing the No Sample condition with another biased sample condition for which participants were *not* informed that the presented samples were biased (Undisclosed Bias condition). Greater belief correction in the Disclosed Bias compared to the Undisclosed Bias condition would suggest that participants deliberately adjust their estimates in response to known sample bias.

Methods

Participants

Three hundred and forty US residents were recruited from MTurk using the same recruitment criteria as Experiment 1. Those who participated in Experiment 1 were restricted from participating in Experiment 2. We excluded data from 45 participants who failed a post-task comprehension check, leaving data from 286 participants (56% Male; 165 Democrats, 64 Republicans, 55 Independents[3]) to be analyzed.

Materials and Measures

Experiment 2 included the same materials and measures as Experiment 1 with one exception: we replaced the No Sample condition featured in Experiment 1 with an Undisclosed Bias condition in Experiment 2. Participants assigned to the Undisclosed Bias condition viewed the same biased samples as those in the Disclosed Bias condition. However, unlike those in the Disclosed Bias condition, participants in the Undisclosed Bias condition were told (falsely) that ratings of sampled survey respondents were randomly selected from all Democrat [Republican] respondents. Thus, while the samples presented in the Undisclosed Bias condition matched those of the Disclosed Bias condition, instructions featured in the Undisclosed Bias condition mirrored those presented in the No Bias condition.

Design and Procedure

The design and procedure of Experiment 2 mirrored that of Experiment 1. Participants estimated the extent to which either the average Democrat or Republican agreed with 12 Democrat- or Republican-leaning statements, respectively. Based on random assignment, participants provided each estimate after viewing a representative sample of five respondents (No Bias condition), viewing an explicitly biased sample of five ideologically extreme respondents (Disclosed Bias condition), or viewing a biased sample of five extreme respondents falsely depicted as being representative (Undisclosed Bias condition). As in Experiment 1,

[3] Two participants did not report an affiliation.

participants responded to accuracy and perceived consensus questions following experimental trials[4] and provided demographic information prior to post-study debriefing.

Results

Mean estimates from the Disclosed Bias, No Bias, and Undisclosed Bias conditions are shown in Figure 4. Belief correction varied across Sample Type, $F(2, 283) = 72.70$, $p < .001$, $\eta_p^2 = .339$. In the critical comparison, belief correction was greater in the Disclosed Bias (M = -12.78, SD = 10.25) than in the Undisclosed Bias condition (M = -7.86, SD = 10.66), $t(190) = 3.26$, $p = .001$, $d = 0.47$, 95% *CI* [0.18, 0.76], suggesting deliberate adjustment in response to known sample bias. Belief correction was also greater in the Disclosed Bias compared to the No Bias condition (M = 3.73, SD = 7.90), $t(189) = 12.44$, $p < .001$, $d = 1.80$, 95% *CI* [1.46, 2.14]. As in Experiment 1, adjustment for known sample bias was insufficient. Critically, participants in the Disclosed Bias condition perceived the average partisan as agreeing more strongly with ideologically-congruent statements (M = 74.73, SD = 10.32) than those shown unbiased samples (M = 70.40, SD = 7.86), $t(189) = 3.25$, $p = .001$, $d = 0.47$, 95% *CI* [0.18, 0.76]. Furthermore, participants in the Disclosed Bias condition provided more moderate estimates than those in the Undisclosed Bias condition (M = 79.71, SD = 10.91), $t(190) = 3.25$, $p = .001$, $d = 0.47$, 95% *CI* [0.18, 0.76]. Therefore, viewing biased—as opposed to representative—samples of political attitudes increased partisan misperceptions leading people to more severely overestimate the ideological extremity of the average partisan,[5] while transparency surrounding the biases inherent in sample selection mitigated these misperceptions.

[4] Participants also completed four items designed to measure individual differences in numeracy (Cokely et al., 2012). Analyses featuring data from this measure can be viewed in the supplementary materials (Part B).
[5] Notably, regardless of the type of sample shown, participants tended to overestimate the ideological extremity of the average partisan (*p's* < .001, *d's* > 0.48).

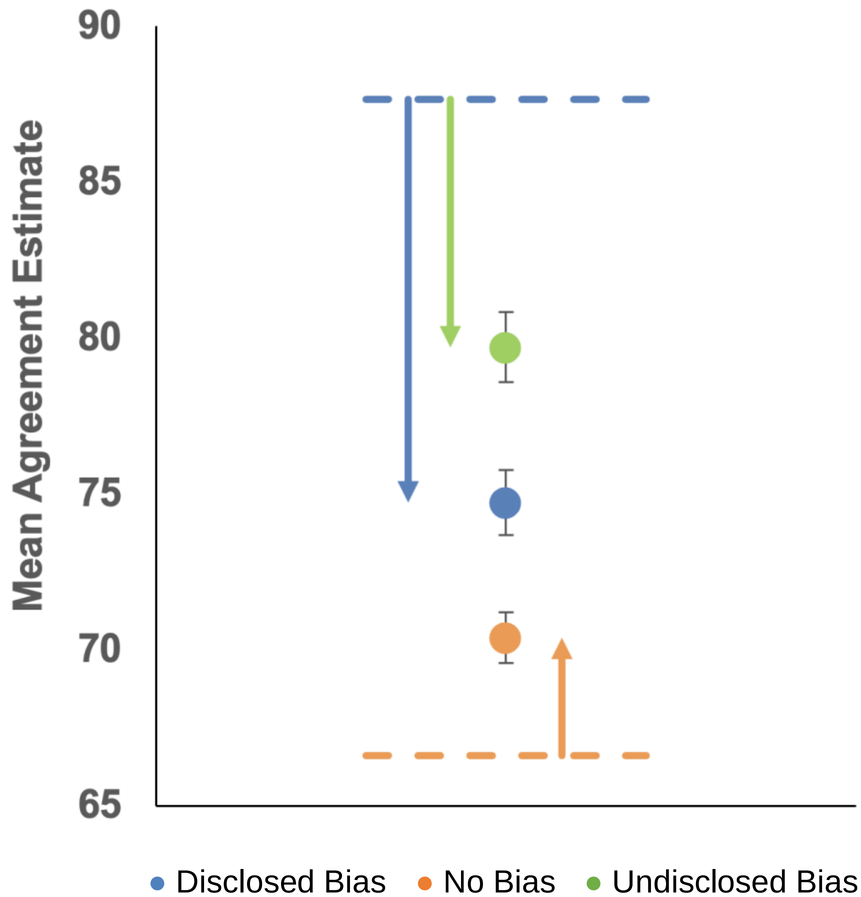

Figure 4. Experiment 2: Mean Agreement Rating Estimates of the Average Partisan. Dashed lines represent the mean agreement ratings of survey respondents presented in biased (blue/green) and unbiased (orange) samples. Arrows depict the difference between the relevant sample mean and participants' mean agreement rating estimates of the average partisan within a condition. Error bars represent +/- 1 SE.

Accuracy

As in Experiment 1, participants in the Disclosed Bias (M = 7.43, SD = 2.19) and No Bias conditions (M = 7.36, SD = 2.20) perceived their agreement rating estimates to be similarly accurate, $t(189) = 0.23$, $p = .823$, $d = 0.03$, 95% CI [-0.25, 0.32], despite those viewing unbiased samples being more accurate in their judgments, $t(189) = 6.43$, $p < .001$, $d = 0.93$, 95% CI [0.63, 1.23]. Thus, while the subjective accuracy of participants viewing unbiased samples was well calibrated, those viewing biased samples were again considerably overconfident, particularly when the biases inherent in sample selection were undisclosed (see Figure 5).

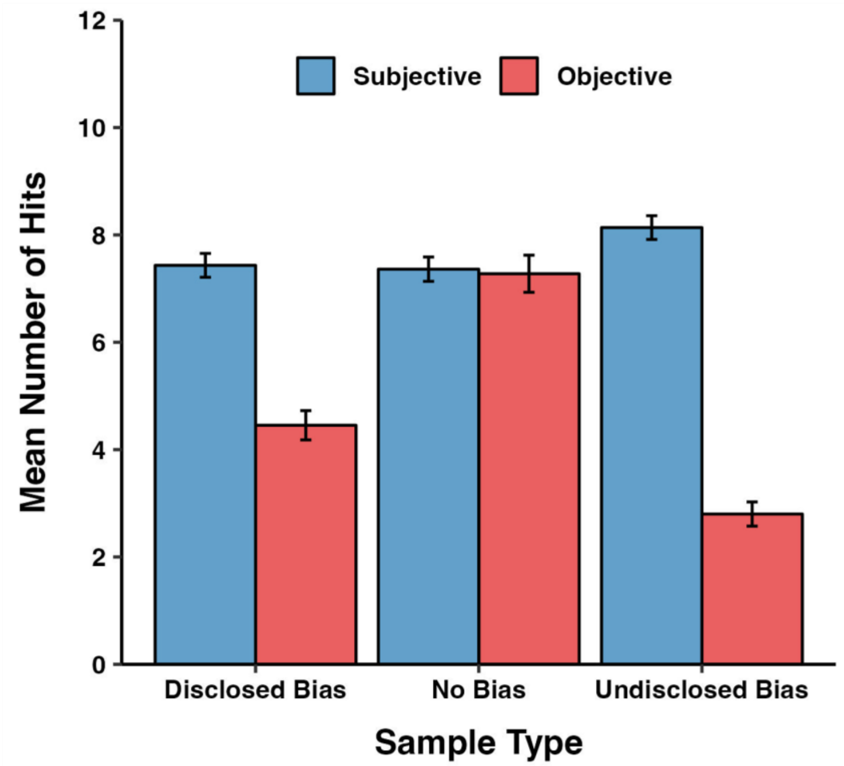

Figure 5. Experiment 2: Subjective and Objective Accuracy. Bars display participants' mean subjective and objective accuracy scores within the Sample Type condition for which participants were randomly assigned. Error bars represent +/- 1 SE.

Perceived Party Consensus

Perceptions of within-party consensus differed by Sample Type, $F(2, 282) = 4.05$, $p = .019$, $\eta_p^2 = .028$. Participants viewing biased (and more homogenous) samples perceived greater consensus among the attitudes of Democrats and Republicans compared to those shown unbiased samples. Specifically, participants in the Undisclosed Bias condition perceived greater consensus ($M = 72.73$, $SD = 15.88$) than those in the No Bias condition ($M = 65.41$, $SD = 20.83$), $t(186) = 2.71$, $p = .007$, $d = 0.40$, 95% *CI* [0.11, 0.68]. Participants in the Disclosed Bias condition also perceived greater consensus ($M = 69.20$, $SD = 15.81$) than those in the No Bias condition, however, this difference was small and not statistically significant, $t(189) = 1.42$, $p = .158$, $d = 0.20$, 95% *CI* [-0.08, 0.49].

General Discussion

People's tendency to overestimate the ideological extremity of the average partisan is well-established (Ahler, 2014; Westfall et al., 2015), as is the link between partisan misperceptions and political polarization (Lees & Cikara, 2021). However, less is known about *why* people overestimate the extremity of political in- and out-group members. One possibility is that the overrepresentation of politically engaged and ideologically extreme partisans on social media and other platforms systematically distorts peoples' perceptions of the *average* partisan. Despite the ubiquity of political expressions, the distribution of beliefs present in a population cannot be directly observed. Instead, this distribution must be inferred from a sample of available expressions drawn from the target population. To the extent that the attitudes of extreme partisans are commonly overrepresented and thus, more likely to be sampled, we may expect people to overestimate the ideological extremity of the average partisan. Consistent with this claim, we show that providing participants with biased—as opposed to representative—samples of political opinion increases the extent to which they overestimate the ideological extremity of Democrats and Republicans (i.e., the extent to which Democrats and Republicans agree with ideologically-congruent statements). Thus, we demonstrate that the political attitudes people encounter exert a direct and sizeable impact on how they perceive the political beliefs of others.

Ideally, when making inferences from biased samples people would adjust their inferences about a population in a manner that corrects for sample bias. While the specific processes responsible for sample bias may seldom be transparent, the present work created an idealized scenario in which participants were explicitly described the biases that shaped the samples they encountered. Within such scenarios, participants corrected for sample bias, albeit insufficiently. In Experiments 1 and 2, participants (correctly) estimated that the beliefs of the average partisan were more moderate than the beliefs presented in explicitly biased samples of ideologically extreme partisans, demonstrating adjustment for sample bias. Nevertheless, participants estimates were more extreme than the true sample mean and the estimates of those viewing unbiased samples, revealing that this adjustment tended to be insufficient. Notably, correction for sample bias was more pronounced when participants were informed of the biases inherent in sample selection, suggesting that such correction represented a deliberate attempt to correct for known sample bias.

Taken together, the present work reveals one mechanism (i.e., insufficient adjustment from biased samples) that can explain how biased samples of political expressions, such as those commonly encountered online (Brady et al., 2023), lead people to misperceive the political beliefs of others. First, hidden biases in sample selection, such as those that amplify the voices of extreme partisans, foster partisan misperceptions as individuals fail to realize that the political opinions they encounter do not represent the beliefs held within a target group. Second, even when aware of sample bias, people fail to adequately correct for it, resulting in false perceptions that are aligned with sample biases. Thus, even when transparent, biases in sample selection can facilitate false perceptions as people recognize the need to account for sample bias but fail to adjust their perceptions adequately. For instance, in the political domain, people may recognize that the beliefs of the average person tend to be more moderate than the viewpoints they frequently encounter, yet still fail to appreciate the extent to which these views are unrepresentative of those endorsed by the average person.

Misattributing the beliefs of the most extreme ideologues to the average partisan deepens political divides, as people increasingly come to view rival partisans as holding irreconcilable views. In the present work, participants’ failure to recognize and account for biased sampling methods worsened their tendency to overestimate the ideological extremity of Democrats and Republicans. These false perceptions have consequences. How people perceive the attitudes of others shapes their own political viewpoints, while overestimating the ideological extremity of opposing partisans hinders productive cross-party interactions and amplifies partisan animosities (Ahler, 2014; Lees & Cikara, 2021). Thus, exposure to biased samples, including those resulting from the overrepresentation of ideologically extreme voices (Hughes, 2019), can contribute to increasing partisan divides.

Conclusion

Scholars have noted the importance of increasing the transparency of content algorithms (e.g., those amplifying polarizing content), suggesting that increased transparency can mitigate social misperceptions by allowing individuals to adjust their inferences to account for biases in content selection (Brady et al., 2023). Mirroring such a scenario in which content selection biases are well-described, the present work demonstrates the potential benefits of algorithm transparency. Individuals adjust their inferences to account for known sample bias, with such adjustments mitigating partisan misperceptions. Nevertheless, this adjustment is often insufficient. Thus, while algorithm transparency may reduce social misperceptions, such misperceptions are likely to remain as a result of individuals exposure to biased samples. As we show, eliminating sample bias leads to more accurate inferences than making sample bias explicit. Nevertheless, the ability for polarizing content to capture attention and promote audience engagement (Brady et al., 2020) helps ensure its overrepresentation across platforms. Therefore, understanding how people interact with biased samples—transparent or not—to make social inferences is a worthwhile goal, as inaccurate perceptions not only facilitate inaccurate beliefs but also promote inter-group conflict.

References

Ahler, D. J. (2014). Self-fulfilling misperceptions of public polarization. *The Journal of Politics*, *76*(3), 607-620.

Ahler, D. J., & Sood, G. (2018). The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics*, *80*(3), 964-981.

Bayes, R., Druckman, J. N., Goods, A., & Molden, D. C. (2020). When and how different motives can drive motivated political reasoning. *Political Psychology*, *41*(5), 1031-1052.

Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The MAD model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science*, *15*(4), 978-1010.

Brady, W. J., Jackson, J. C., Lindström, B., & Crockett, M. J. (2023). Algorithm-mediated social learning in online social networks. *Trends in Cognitive Sciences*, *27*(10), 947-960.

Dorison, C. A., Minson, J. A., & Rogers, T. (2019). Selective exposure partly relies on faulty affective forecasts. *Cognition*, *188*, 98-107.

Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2022). (Mis)estimating affective polarization. *The Journal of Politics*, *84*(2), 1106-1117.

Hamill, R., Wilson, T. D., & Nisbett, R. E. (1980). Insensitivity to sample bias: Generalizing from atypical cases. *Journal of Personality and Social Psychology*, *39*(4), 578-589.

Hauser, D. J., Moss, A. J., Rosenzweig, C., Jaffe, S. N., Robinson, J., & Litman, L. (2023). Evaluating CloudResearch's approved group as a solution for problematic data quality on MTurk. *Behavior Research Methods*, *55*(8), 3953-3964.

Hong, S. (2013). Who benefits from Twitter? Social media and political competition in the U.S. House of Representatives. *Government Information Quarterly*, *30*(4), 464-472.

Hong, S., & Kim, S. H. (2016). Political polarization on Twitter: Implications for the use of social media in digital governments. *Government Information Quarterly*, *33*(4), 777-782.

Hughes, A. (2019, October 23). *A small group of prolific users account for a majority of political tweets sent by U.S. adults*. Pew Research Center. https://pewrsr.ch/3a31iDK.

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, *4*(3), 279-286.

Lees, J., & Cikara, M. (2021). Understanding and combating misperceived polarization. *Philosophical Transactions of the Royal Society B*, *376*(1822), 20200143.

Levendusky, M. S., & Malhotra, N. (2016). (Mis)perceptions of partisan polarization in the American public. *Public Opinion Quarterly*, *80*(S1), 378-391.

Moore-Berg, S. L., Ankori-Karlinsky, L. O., Hameiri, B., & Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proceedings of the National Academy of Sciences*, *117*(26), 14864-14872.

Rathje, S., Van Bavel, J. J., & Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, *118*(26), e2024292118.

Vlasceanu, M., Morais, M. J., & Coman, A. (2021). The effect of prediction error on belief update across the political spectrum. *Psychological Science*, *32*(6), 916-933.

Walker, A. C., Fugelsang, J. A., & Koehler, D. J. (2024). Partisan language in a polarized world: In-group language provides reputational benefits to speakers while polarizing audiences. Retrieved from: https://osf.io/preprints/psyarxiv/kxard

Westfall, J., Van Boven, L., Chambers, J. R., & Judd, C. M. (2015). Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science, 10*(2), 145-158.

Yudkin, D., Hawkins, S., & Dixon, T. (2019, June). *The perception gap: How false impressions are pulling Americans apart*. More in Common. https://perceptiongap.us