Extremely Partisan Samples Impact Perceptions of Political Group Beliefs

by

Alexandra van der Valk

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada, 2023

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Accurately inferring the beliefs of a partisan group (e.g. Democrats, Republicans) can be challenging when exposed to extremely partisan beliefs from that group. Across two studies (total N = 566), we tested whether people correct these inferences for sample bias when it was explicitly disclosed. Study 2 further assessed how much of this correction is deliberate. Participants read 12 statements that most members of a political party (Democrats or Republicans) generally agree with. They were shown how strongly five party members agreed with each statement. In the biased sample conditions, these five party members were selected from the top 10% most partisan members; this bias was either disclosed or undisclosed. In the unbiased sample condition, the five members were representatively sampled from the entire party. Then, participants estimated on average how much the entire party agreed with each statement, and the likelihood that party members of the same or opposing parties agreed with each other. Participants' mean estimates from the biased sample conditions were higher than the unbiased sample condition but lower than the samples viewed, indicating an (insufficient) attempt to correct for sample bias. Corrections were largest when sample bias was disclosed. Overall accuracy was highest when participants viewed unbiased samples, though across conditions there appeared a general tendency to overestimate strength of partisan beliefs. Parties were perceived as more homogeneous when participants viewed biased samples, regardless of whether bias was disclosed or not. While awareness of hyperpartisan bias helps correct judgments, it may not eliminate overestimation, overconfidence, or inflated perceptions of party homogeneity.

Acknowledgements

I am indebted to my supervisors, Derek Koehler and Jonathan Fugelsang. You are as intimidatingly brilliant as you are kind. I am forever grateful for the patience and understanding you've shown me as I bumbled through these past two years. If only all students were so lucky.

Thank you to my reader, Dr. Samuel Johnson, for his care in providing such valuable and insightful feedback. To my lab-mates for being so welcoming, especially Kaiden Stewart and Dr. Alexander Walker for their support and advice as I navigated this new and unfamiliar territory. And to my "Little Sib", Hamza Tariq, who taught me what true strength is.

To my parents, thank you for your support through my failures as much as my successes. We have each had a stressful few years – for many reasons, it's been the perfect storm. I wouldn't have finished this journey without you.

To my UW friends, I never imagined I'd meet such an amazing group of people who have become family to me. You made Waterloo a true home away from home. I miss you all dearly.

To my East Coast friends, mentors, and "Proud Mom" at home and away, I carry you all with me. Truly, "Nova Scotia Strong".

Table of Contents

List of Figures

List of Tables

Introduction

Political Polarization

Political polarization – the divergence of political attitudes often coupled with dislike for one's opposition – is rising (Kubin & von Sikorski, 2021; Wilson et al., 2020). This animosity can threaten democracy, prosocial intentions (Yap, 2023) and make people more tolerant of unethical ingroup action like politically-motivated violence (Hartman et al., 2022) and skirting protected rights (Kingzette et al., 2021). When polarization is prevalent among those in power, issues in cross-party dialogue and governmental action emerge (Hartman et al., 2022; Jones, 2001). Exposure to hyperpartisan misinformation can drive this divide (Gadarian et al., 2021; Jungkunz, 2021), and political content that aims to inflate perceptions of partisan beliefs is particularly ripe for worsening pre-existing mistrust towards political leaders and medical experts, fuelling hesitancy in public health guidance (Baumgaertner et al., 2018; Kennedy, 2019; Zimmerman et al., 2023). Repeated exposure to misleading information about political groups increases its believability (Pennycook et al., 2018) and can alter prototypical inferences about social groups (Ahler & Sood, 2018), ultimately exacerbating polarization and motivating people to spread of these views (Osmundsen et al., 2021). And, the more polarized people become, the more likely they will be exposed to such rhetoric (Rao et al., 2022).

Perceptions of Partisan Beliefs

Party members also frequently make inaccurate observations about their political opposition (Ahler & Sood, 2018; Moore-Berg et al., 2020; Yudkin et al., 2019). Partisanship is favoured over accuracy (Kahan, 2017; Van Bavel & Pereira, 2018), predicting belief in politically-polarized news (Anthony & Moulding, 2019) and overestimations of the accuracy of news aligning with one's political values (Armaly & Enders, 2023; Frenda et al., 2013). Partisan bias also drives a propensity to favour information that disparages partisan outgroups (Pereira et al., 2023) or highlights stereotypes that make them appear more threatening (Mernyk et al., 2022; Pasek et al., 2022). For example, party members have a tendency to support policies presented by their ingroup, but oppose the same policy when presented by the opposing party (Cohen, 2003; Wilson et al., 2020).

This in-group favoritism preference is bolstered by the magnitude of polarization. In one study, viewing polarized Democrats and Republicans led to judgments of policy support compatible with ingroup party values. When parties were presented as unpolarized, people appropriately considered each sides' arguments to guide their decisions (Druckman et al., 2013). Representatives lobbying for a particular cause could encourage these perceptions of polarization to drum up more support for their enterprise.

Polarized affect favouring one political party over another predicts the likelihood of making political misperceptions (Garrett et al., 2019), and such errors are largest among the most partisan members (Landry et al., 2023; Moore-Berg et al., 2020). Naturally, the spread of perceptions of polarization leads to inaccurate inferences about partisan groups that are challenging to correct. Even in the face of accurate evidence, alternate opinions can easily be sought elsewhere that better align with personal belief or motivation. Judgments of population values tend to be particularly uncalibrated. For example, a representative study of Americans found that regardless of personal political affiliation, the majority of participants underestimated the true proportion of national concern for climate change and support for mitigation action by nearly half (Sparkman et al., 2022). Those who underestimated the most were those who least supported such policies, estimating population belief in a manner that aligned with their own views.

Democrats and Republicans actually agree on various core values, yet they both poorly gauge their opposition's beliefs (Pasek et al., 2022) and overestimate the ideological gap between parties concerning political topics (Lees & Cikara, 2019; Moore-Berg et al., 2020), incorrectly believing that the opposition holds extreme political views (Ahler, 2014; Enders & Armaly, 2019; Graham et al., 2012; Westfall et al., 2015). This could in part stem from increased exposure to unrepresentative, hyperpartisan beliefs, making these beliefs appear more prevalent than they actually are (Lerman et al., 2016). Exposure to hyperpartisan information can lead to false perceptions of outgroup extremity and homogeneity that are difficult to attenuate, even in the face of contradictory evidence (Ahler & Sood, 2018). Assessments of the veracity of such misinformation can also be influenced by extremely biased claims, even when subjects know the information is false (Jost et al., 2020). Errors in inferences about partisan groups are an important determinant of outgroup animosity and subsequent polarization – more than actual group attitudes (Enders & Armaly, 2019; Wilson et al., 2020). When one side of the political spectrum views the other as more extreme, their own views become more extreme as well (Ahler, 2014), leading to an unwillingness to attend to evidence that contradicts their political worldview (Rollwage et al., 2019; Zmigrod, 2020).

Social Media

Misperceptions about one's political opponents may stem from several sources, but social media in particular is considered a key facilitator of polarization (Kubin & von Sikorski, 2021; Van Bavel, Rathje, et al., 2021). People rely on their social networks to select which issues require attending (Tokita et al., 2021), and with over half of the global population actively using social media (Statista, 2023), vulnerabilities are particularly salient. Our strengthening digital network makes it easy to seek content that supports or refutes virtually any topic. Mere access to the internet is associated with partisan hostility (Lelkes et al., 2017), and online, users are more likely to be exposed to distorted or unrepresentative content than in everyday life. For example, fake bot accounts imitating humans contribute significantly to the spread of distorted and highly partisan content (Badawy et al., 2018; Bastos & Mercea, 2019; Himelein-Wachowiak et al., 2021). Encountering extremely partisan content triggers more intensely negative feelings compared to accurate or less extreme information. These negative emotions subsequently foster polarisation (Weismueller et al., 2023) and can colour perceptions of political parties as a whole.

In tandem, social media amplifies divisive and outright false content (Rathje et al., 2022; S. van der Linden et al., 2021; Van Bavel, Harris, et al., 2021; Van Bavel, Rathje, et al., 2021), dedicating more airtime to extreme behaviour and conflict (Padgett et al., 2019). Sharing extreme opinions online is socially rewarded (Hong & Kim, 2016; Padgett et al., 2019), and people will repost content regardless of whether it aligns with their personal beliefs (Pennycook et al., 2021). Encountering outgroup extremism on social media can make people become more deeply rooted in their convictions and exacerbate polarization (Bail et al., 2018). As a result, social media has evolved to encourage controversial or divisive actions (Rathje et al., 2021), reinforcing group stereotypes and widening the partisan divide (Bail et al., 2018; Kubin & von Sikorski, 2021; Lorenz-Spreen et al., 2021).

Overview of Studies

The nature of political polarization is intricate, bolstered by exposure to distorted partisan beliefs, a relationship driven by the widely accessible social media. Underscoring the importance and complexity of mitigating these intertwined factors, there are clear links between partisan perceptions and political polarization that have important downstream consequences on interparty dialogue.

Correcting misconceptions can reduce partisan animosity, which is linked to polarization (Ruggeri et al., 2021). While any amount of attenuation in extreme perceptions is clearly desirable for reducing polarization, few studies investigate how engaging with extremely partisan beliefs impacts perceptions of political parties after debunking the underlying bias, or differentiate such effects when exposed to true information or a no-information control (Guay et al., 2023; Weismueller et al., 2023).

Comparing the magnitude of perception correction (if any) when participants are made aware of selection bias highlights the specific impact of extremity compared to the more classically researched correction when presented with unbiased information (Hartman et al., 2022). It is also unclear whether disclosing the bias underlying misinformation leads to more accurate judgments (Chan et al., 2017).

The first objective of Study 1, therefore, is to expose participants to biased samples of extreme members of left- and right-leaning political parties, testing whether these beliefs distort their perceptions of the views of members of that party – even when they are aware of the bias. We do this by first presenting participants with samples of extremely partisan beliefs, explicitly disclosing how these samples were selected, and then having them estimate true population belief. We anticipate that participants will provide estimates that are less extreme than the presented samples. Given the historical anchoring effect of presented beliefs, however, we anticipate that this correction will be insufficient when compared to participants given accurate information or participants given no additional materials (i.e., a valuable baseline of how the public generally perceives these parties). That is, that people shown biased samples overestimate party belief compared to those shown no samples. Explicit knowledge of selection bias should encourage participants to place less weight on these samples when making their evaluations and more weight on their prior expectations.

In Study 2, we build on these findings by examining whether results differ when participants are aware of the underlying selection bias versus unaware. We will present participants with the same biased samples, disclosing this bias to one study condition but not to another. We will also once again present a third condition with unbiased samples, as is typically done in the literature. Any differences in population estimates will also indicate whether participants are truly attending to the presented information.

Across these two studies, we also explore whether an awareness of selection bias impacts how much variability is perceived within and between partisan groups. We anticipate that skewed, low variance ("biased") distributions of party beliefs portrays parties as more cohesive than true distributions. Perceived and objective estimate accuracy will also be compared across sample types, with the expectation that those presented with extremely partisan samples will overestimate the accuracy of their judgments. Finally, we will examine whether individual differences – such as political identity – moderate these relationships.

While similar studies utilize fabricated manipulations, this study draws on existing data on the beliefs of Democrats and Republicans regarding a variety of topics. We also build on the typically single domain focus (e.g., climate change) of prior work to explore perceptions across a variety of topics, better assessing broad perceptions of these parties.

Study 1

Methods

Participants

Three hundred Mechanical Turk participants were recruited via CloudResearch to complete an online Qualtrics survey in exchange for $1.50USD. To maximize data quality (Hauser et al., 2022), only participants from CloudResearch's approved participants pool who had completed at least 100 HITS with a minimum approval rating of 95% were recruited. All participants provided informed consent and passed two brief pre-study attention checks and one post-instructions comprehension check. The project protocol was approved by a Research Ethics Board at the University of Waterloo (REB #44521).

Prior to analysis, one participant was excluded for insufficient completion, as were 19 who failed a post-survey attention check, leaving *N* = 280 with complete data. Of these, 48.2% identified as a Democrat, 23.6% as Republican, and 25.4% as Independent. The remaining 1.4% did not report affiliation. Participants were mostly white (74.6%), identified as male (59.6%), and were over the age of 18, though most (83.9%) indicated they were between the ages of 25-54.

Stimuli

Researchers at Princeton University (Vlasceanu et al., 2021) previously developed a set of politically polarizing statements regarding topics that a representative national sample of Democrats and Republicans disagreed on (e.g., immigration, healthcare, climate change). For example, "All cities in the US experience more extremely hot days compared to 50 years ago". These statements are listed in Appendix A. For each statement, the authors had 350 Democrats and 350 Republicans rate how much they agreed with each statement on a scale of 0 (*completely disagree*) to 100 (*completely agree*).

Of these statements, 12 were categorized as aligning better with Democrat attitudes (left-leaning) and 12 with Republicans' (right-leaning). Using the Princeton study's open data, we pulled samples of five agreement ratings for each item that either accurately reflect the beliefs of the larger study sample or distort them to make sample agreement appear higher than average. These sample beliefs were computed and shown to our participants as detailed below.

Sample Construction

Unbiased Sample. To get a sample of five agreement ratings that accurately reflect Democrat agreement, the set of agreement ratings (range: 0-100) for the first Democrat-leaning item was divided into quintiles. The median of each quintile was extracted, resulting in five median agreement ratings that comprise a representative sample of the distribution of Democrat beliefs for that item. This process was repeated for the remaining 11 Democrat-leaning items, resulting in 12 samples of unbiased, representative agreement ratings for Democrats that evaluated the Democrat-leaning item set.

The same was done within the 12 Republican items, resulting in 12 samples of agreement ratings that are representative of Republican beliefs regarding the Republican-leaning item set. The full set of constructed samples is available in Appendix G. We note that the individual samples were randomly presented in the same or opposite order to account for order effects.

Biased Sample. To create samples of biased party agreement, only participants with average agreement ratings at or above the 90th percentile were used (N=35 Democrats and N=35 Republicans). This 10% cut-off yielded sample distributions with less variability and that differed sufficiently from the respective unbiased distributions.

The sample selection process otherwise mimics that of the unbiased samples. For each Democrat-leaning item, the 35 Democrat agreement ratings were divided into quintiles, where the median of each quintile comprised the sample for that item. This process was repeated for the Republican-leaning items, resulting in 12 samples of biased agreement ratings for the Democrat-leaning item set, and 12 biased samples for the Republican item set.

To summarize, this approach yields 12 biased and 12 unbiased samples of agreement ratings provided by Democrats for each Democrat-leaning item, as well as 12 biased and 12 unbiased samples of agreement ratings provided by Republicans for Republican-leaning items. The biased samples depict agreement that is nearly 20-points higher on average (scale: 0-100) than the respective unbiased samples.

Study Instructions

In this 2x3 experimental design, participants were first briefed on the Princeton study and how they collected Democrat and Republican agreement with statements on a variety of issues. They were then randomly assigned to read either the 12 Democrat-leaning item set or the Republican-leaning item set, as well as to one of three experimental conditions determining what additional information would be provided prior to them making their estimates: the Disclosed Bias condition (*N* = 104), Disclosed No Bias condition (*N* = 77), or a No Sample control condition (*N* = 99). These conditions are detailed below.

Participants were first instructed to read each statement, then estimate on a scale of 0 (*completely disagree*) to 100 (*completely agree*) the average agreement rating given by all Democrats (or Republicans) who participated in the Princeton study. The presentation order of these statements was randomized.

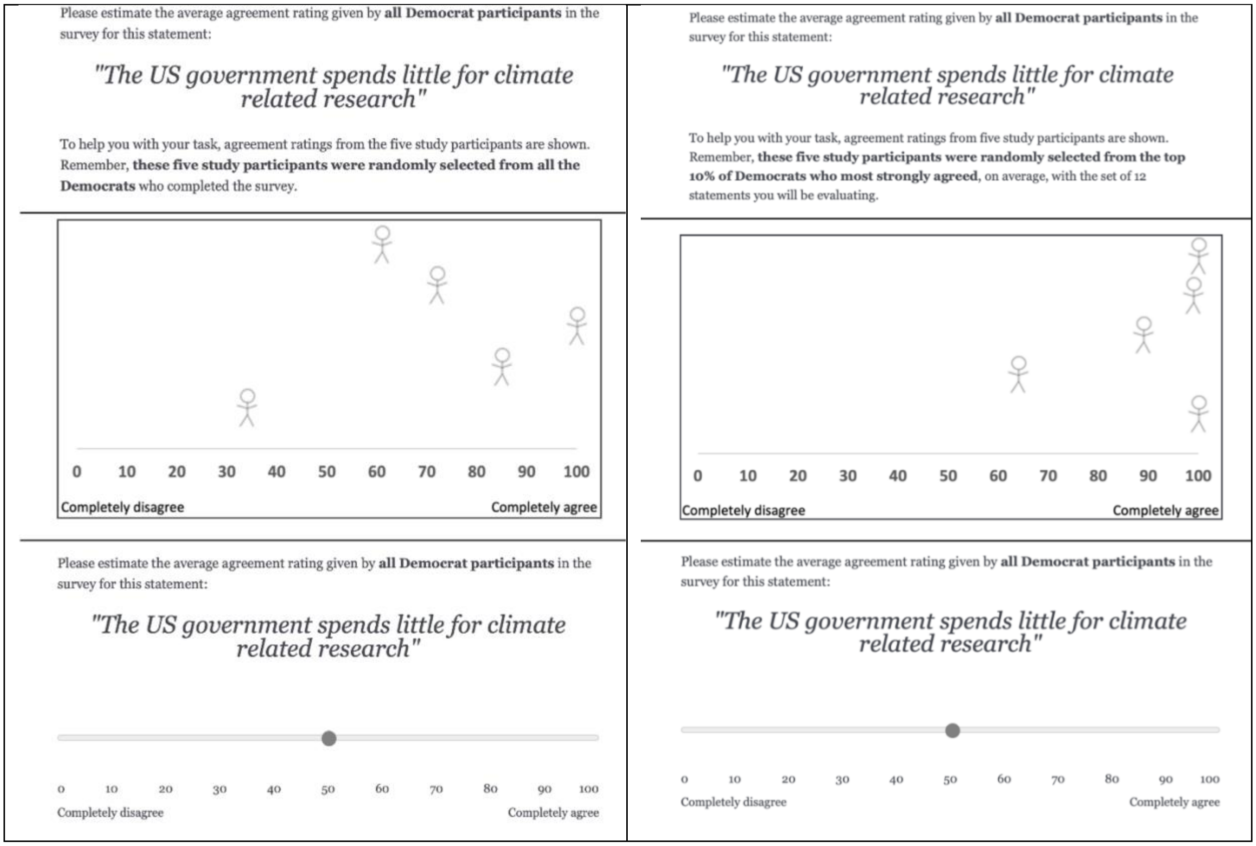
For the Disclosed Bias and Disclosed No Bias conditions, participants were told that they would be shown the agreement ratings of a subset of respondents for each statement, to help them with their evaluations. Those in the Disclosed No Bias condition were told that for each statement they would be shown five agreement ratings randomly selected from the set of all Democrats (or Republicans) in the Princeton study, samples that can be considered representative of the respective political population. Those in the Disclosed Bias condition were instructed similarly, but that the ratings shown were sourced from the top 10% most extremely partisan respondents, therefore being biased and unrepresentative of the respective party. That is, both the Disclosed Bias and Disclosed No Bias participants had explicit awareness of how these samples were selected. These participants also passed pre- and post-task mandatory attention checks assessing their understanding of the study task. The No Sample control condition was not shown – and had no knowledge of – samples of agreement ratings. An example of a Democrat statement is shown in Figure 1, with instructions for both the Disclosed No Bias and Disclosed Bias sample conditions. Full study materials for all sample type conditions are available in Appendices B and G.

After providing estimates of the average Democrat (or Republican) agreement with each of the 12 presented statements, participants rated the subjective accuracy of their estimates, global ratings of perceived consensus within and between parties, as well as demographic information, including political affiliation.

Figure 1

Example Item-level Stimuli for the Disclosed Bias and Disclosed No Bias Sample Types

Disclosed No Bias Condition	Disclosed Bias Condition



Please estimate the average agreement rating given by **all Democrat participants** in the survey for this statement:

"The US government spends little for climate related research"

To help you with your task, agreement ratings from the five study participants are shown. Remember, **these five study participants were randomly selected from all the Democrats** who completed the survey.

Please estimate the average agreement rating given by **all Democrat participants** in the survey for this statement:

"The US government spends little for climate related research"

Please estimate the average agreement rating given by **all Democrat participants** in the survey for this statement:

"The US government spends little for climate related research"

To help you with your task, agreement ratings from five study participants are shown. Remember, **these five study participants were randomly selected from the top 10% of Democrats who most strongly agreed**, on average, with the set of 12 statements you will be evaluating.

Please estimate the average agreement rating given by **all Democrat participants** in the survey for this statement:

"The US government spends little for climate related research"

Outcomes

Belief Correction. Each participant provided ratings of the average Democrat's (or Republican's) agreement with 12 statements on a scale of 0 (complete disagreement) to 100 (complete agreement). These 12 ratings were averaged to a single point estimate. Belief correction was calculated as the difference between a participant's average agreement rating and the average of the sample beliefs viewed. Individuals assigned to view no samples are excluded from correction analyses.

Accuracy. An accurate rating (a "hit") was defined as an estimate of party agreement that was within +/-10 points of the true average party agreement for the respective item, as sourced from Vlasceanu et al. For subjective accuracy, participants reported how many hits they thought they achieved across the 12 statements (0-12). Objective accuracy was measured as the true number of hits achieved.

Perceived Consensus. To assess perceptions of party consensus, participants estimated the global probability (0-100%) that two randomly selected Democrats (if viewed left-leaning items) or Republicans (if viewed right-leaning items) would report agreement ratings within 10 points of each other.

Perceptions of opposing party consensus were measured as the global probability (0-100%) that one randomly selected Democrat and one randomly selected Republican would report agreement ratings within 10 points of each other.

Participant Political Affiliation. Following the methodology of Vlasceanu et al. (2021), participants reported their political identity (Democrat, Republican, or Independent) and how strongly they identified with their chosen party (1 = Not at all, 5 = A great deal). To accommodate the fewer responses at the lower end of the scale, as has been found in previous work (Levay et al., 2016), affiliation strength was dichotomized by median split: Low strength (1-3) versus High strength (4-5). Sensitivity analyses indicated primary results did not vary using an alternative cutpoint.

Results

Any participants who failed the post-test attention checks were excluded from analysis. Of the 280 retained participants, *n* = 135 viewed Democrat-leaning statements and *n* = 145 viewed Republican-leaning statements. Within the three study conditions based on sample type, *n* = 104 were assigned to view a biased sample with a disclosed selection method (Disclosed Bias), *n* = 77 viewed a representative sample with a disclosed selection method (Disclosed No Bias), and an *n* = 99 control condition did not view any samples (No Sample).

Belief Correction

A univariate factorial analysis of variance (ANOVA) explored the impact of target party (Democrat item set vs Republican item set), sample type (Disclosed Bias, Disclosed No Bias), participant political affiliation (Democrat, Republican, Independent), and the strength of that affiliation (Low, High) on the magnitude of estimate correction, defined as the average difference across items between a participant's agreement rating and the average sample agreement viewed, which were both evaluated on a 100-point scale. Negative values indicate agreement ratings were lower than the samples shown, whereas positive values indicate agreement ratings were higher. Participants in the No Sample condition are excluded from these analyses.

Full model results are available in Table 1. Overall, the type of sample viewed had a substantial impact on the magnitude of belief correction, (*F*(1, 152) = 69.22, *MSE* = 7551.24, *p* < .001, η^2_p = 0.31), where the Disclosed No Bias participants adjusted their estimates of party agreement nearly four times less in magnitude (*M* = 3.54) than participants in the Disclosed Bias condition (*M* = -13.47). There were no main effects of target party (*F*(1, 152) = 3.35, *MSE* = 364.97, *p* = .069) or partisanship (party affiliation: *F*(2, 152) = 0.96, *MSE* = 104.99, *p* = .384; affiliation strength: *F*(1, 152) = 0.90, *MSE* = 98.46, *p* = .344). This indicates that belief adjustment was similar regardless of participants' own political stance.

Table 1.

Results of Factorial ANOVA predicting Belief Correction

Predictor Variables	df	Mean Square	*F*	*p*	Partial Eta Squared (η^2_p)
Sample Type	1	7551.24	69.22	**<.001**	.313
Target Party	1	364.97	3.35	.069	.022

Party Affiliation	2	104.99	.96	.384	.013
Affiliation Strength	1	98.46	.90	.344	.006
Party Affiliation * Sample Type	2	53.20	.49	.615	.006
Party Affiliation * Target Party	2	136.63	1.25	.289	.016
Party Affiliation * Affiliation Strength	2	39.58	.36	.696	.005
Sample Type * Target Party	1	12.69	.12	.734	.001
Sample Type * Affiliation Strength	1	2.65	.02	.876	.000
Target Party* Affiliation Strength	1	36.35	.33	.565	.002
Party Affiliation * Sample Type * Target Party	2	15.71	.14	.866	.002
Party Affiliation * Sample Type * Affiliation Strength	2	131.86	1.21	.301	.016
Party Affiliation * Target Party * Affiliation Strength	2	166.87	1.53	.220	.020
Sample Type * Target Party* Affiliation Strength	1	811.89	7.44	**.007**	.047
Party Affiliation * Sample Type * Target Party * Affiliation Strength	2	170.04	1.56	.214	.020
Error	152	109.10			
Total	176				

Note. **Bolded** *p*-values are significant at the .01 level or better.

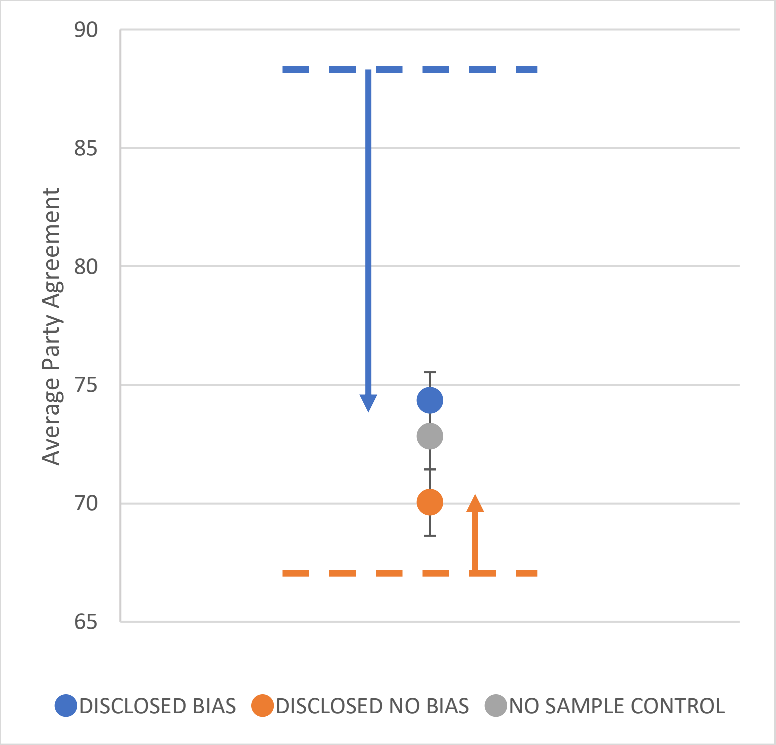
There were also no two- or three-way interactions between these variables, except for one three-way interaction between sample type, target party, and party affiliation strength, $F(1, 152) = 7.44$, $MSE = 811.89$, $p = .007$, $\eta^2_p = 0.05$). The two-way relationship between target party and affiliation strength was significant when exposed to the Disclosed No Bias ($F(1, 65) = 13.04$, $MSE = 613.71$, $p < .001$, $\eta^2_p = 0.17$) but not Disclosed Bias ($F(1, 87) = 1.58$, $MSE = 245.22$, $p = .212$) samples. It appears that people who viewed representative beliefs corrected their estimates of the Republican party more when they identified as weakly partisan ($M = 9.35$) as opposed to strongly partisan ($M = 0.66$). Given the limitations of the small cell sizes in these comparisons, however, we hesitate to extrapolate on this trend.

Regarding the point estimates of agreement, one-tailed independent t-tests indicate that the Disclosed No Bias participants estimated party agreement as significantly lower (*M* = 70.03) than the Disclosed Bias sample participants (*M* = 74.33), *t*(179) = 2.72, *p* = .004, *d* = 0.41. Notably, the Disclosed Bias judgments did not differ from the No Sample participants (*M* = 72.82), *t*(201) = 0.87, *p* = .192. The Disclosed No Bias participants, however, judged party agreement as significantly higher than the No Sample condition (*t*(174) = 1.73, *p* = .043, *d* = 0.26), though this difference was small.

Average estimates of party agreement pooled across target parties, and respective magnitudes of belief adjustment, are depicted in Figure 2. These results stratified by target party are available in Appendix C.

Figure 2

Pooled Average Perceived Agreement of Democrats and Republicans



Note. Error bars are +/- 1 SE. Blue dashed line is the sample mean shown to participants who viewed a Disclosed Bias sample. Orange dashed line is the sample mean shown to participants who viewed a Disclosed No Bias sample. Arrows depict the difference between the sample mean shown and the average estimate of party agreement.

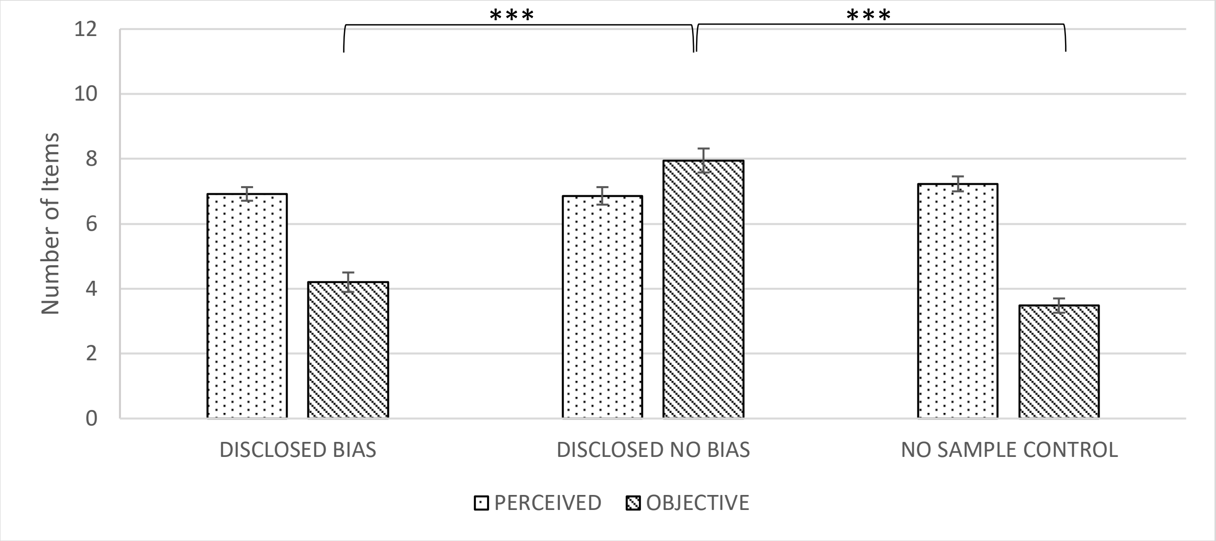
Accuracy

Perceived and objective hit rates (possible ranges: 0-12) are presented in Figure 3. A univariate factorial ANOVA found that perceived accuracy did not vary between the two target parties (*F*(1, 274) = 1.34, *MSE* = 6.78, *p* = .247) or between the three types of samples viewed; participants across all conditions anticipated achieving a similar number of hits, *F*(2, 274) = 1.27, *MSE* = 6.41, *p* = .282.

An ANOVA of objective hit rate, however, found a main effect of sample type, *F*(2, 274) = 50.07, *MSE* = 392.55, *p* <.001, η^2_p = 0.27. A Tukey post hoc test indicated that participants in the Disclosed Bias (*M* = 4.31) and No Sample (*M* = 3.52) conditions were similarly poorly calibrated (*p* = .144) and that were both significantly less accurate than the Disclosed No Bias participants (*M* = 7.71; both pairwise *p* < .001). There was also a main effect of target party, where objective accuracy was slightly higher when evaluating Democrat-leaning statements (*M* = 5.76) compared to Republican-leaning statements (*M* = 4.60), *F*(2, 274) = 11.37, *MSE* = 89.14, *p* <.001, though this effect was small (η^2_p = 0.04). There was no interaction between group condition and target party rated, *F*(2, 274) = 0.76, *MSE* = 5.95, *p* = .469. Follow-up independent samples t-tests of objective accuracy suggest that this better performance in evaluating Democrat beliefs is not an artifact of the greater representation of Democrats in the sample, where within study conditions the number of hits achieved was similar regardless of target party congruence (Disclosed Bias *t*(73) = 1.75, *p* = .084; Disclosed No Bias *t*(48) = 0.06, *p* = .478; No Sample *t*(74) = 0.19, *p* = .848.

Figure 3

Perceived versus Objective Hit Rates across Sample Types



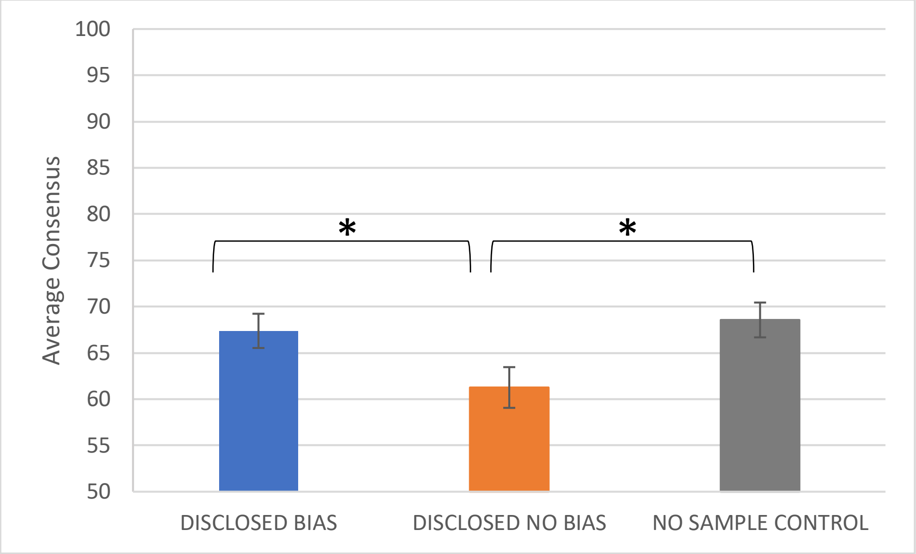
Note. Error bars are +/- 1 SE. Significance at the .05, .01, and .001 levels are denoted by *, **, and ***, respectively.

Perceived Consensus

Average ratings of perceived consensus between members of the same party and opposing parties are reported in Figures 4a and 4b. A univariate factorial ANOVA of perceptions of consensus among two members of the same party differentiated by sample type and target party found that consensus ratings were similar between the Democrat and Republican target parties, $F(1, 274) = 1.21$, $MSE = 414.98$, $p = .273$. Participants in the Disclosed Bias group ($M = 67.38$) rated consensus similarly to participants in the No Sample condition ($M = 68.56$), whereas individuals viewing a Disclosed No Bias sample ($M = 61.26$) judged party consensus as significantly lower than the other two study conditions, $F(2,274) = 3.52$, $MSE = 1211.17$, $p = 0.031$, $\eta^2_p = 0.03$ (Figure 4a).

Figure 4a

Average Perceived Consensus of Two Members of the Same Party

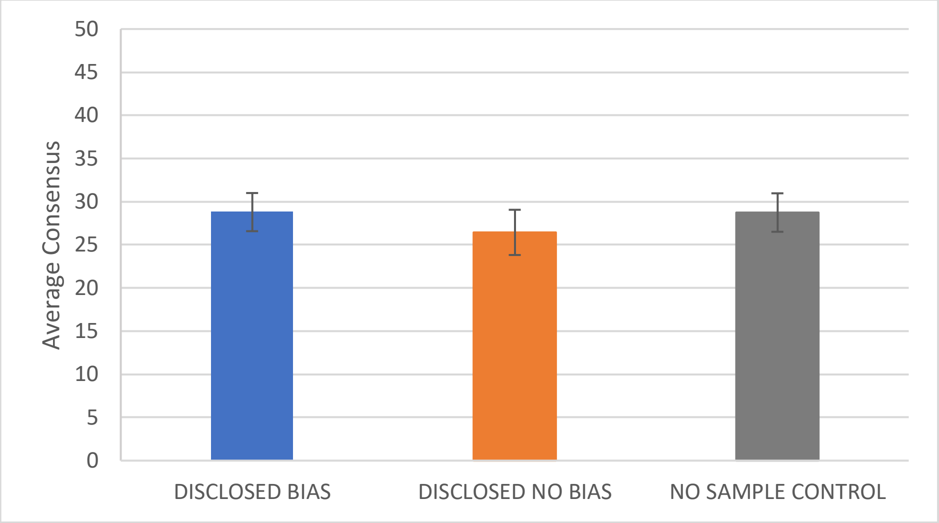


Note. Error bars are +/- 1 SE. Significance at the .05, .01, and .001 levels are denoted by *, **, and ***, respectively.

Interestingly, there were no differences between sample types regarding perceptions of consensus between two members of opposing parties, *F*(2,274) = 0.29, *MSE* = 141.65, *p* = .748, regardless of the target party shown, *F*(1, 274) = 0.29, *MSE* = 138.81, *p*=.594 (see Figure 4b).

Figure 4b

Average Perceived Consensus of Two Members of Opposing Parties



Discussion

The results of Study 1 indicate that participants made aware that they had viewed extreme political beliefs estimated party belief to be much lower than the sample beliefs they were shown, suggesting an attempt at correction to a level similar to participants who did not view any sample beliefs. The magnitude of this correction was larger than that of participants who saw representative samples of beliefs, who sensibly adjusted far less.

In terms of accuracy, viewing hyperpartisan beliefs did not impact individual perceptions of task performance. It did, however, significantly impair objective accuracy. Conversely, people who viewed unbiased agreement ratings most accurately judged party belief.

Finally, viewing representative agreement ratings attenuated perceptions of consensus among members of the same political party. Given the similarity in consensus ratings between people who viewed extremely partisan samples and those who did not view any samples, perhaps perceived consensus is generally high among the public until attenuated by viewing representative beliefs. Interestingly, participants viewed consensus between opposing parties to be similarly low, regardless of the samples viewed.

Study 2

Methods

Study 2 follows the same procedure as Study 1, but addresses some limitations. First, the first study's condition allocation was truly randomized; while beneficial, our cell sizes were diminished in some analyses. Second, the judgments of participants who viewed Disclosed Bias samples did not differ those who viewed no samples. This makes it difficult to interpret whether the Disclosed Bias participants deliberately corrected for the biased selection method, or if they naturally underestimated group belief due to the biased samples being disproportionately skewed to the latter end of the response scale. That is, did participants truly attend to the samples we showed them, or did they simply revert to personal belief? To address this, Study 2 uses a counterbalanced approach and replaces the No Sample condition with one where participants view the same extremely partisan samples as Study 1, but are told that the beliefs are representative (Undisclosed Bias). If participants truly understand the task, their estimates should remain skewed higher than those of participants aware of the selection bias.

Furthermore, we introduce additional measures in Study 2 to provide insights into possible mechanisms underlying our observed effects. The first is a measure of numeracy. As the sample beliefs shown to participants were numerical, individual differences in understanding and working with numbers could influence how they interact with these samples. We have added a measure of numeracy assessed using the 4-item, multiple-choice version of the Berlin Numeracy Test (Cokely et al., 2012). Participant scores reflected the number of correct responses (0-4), where 0 indicates poor numeracy and 4 indicates excellent numeracy. Scale items can be viewed in Appendix E.

Finally, we add two items assessing perceived party consensus. These measure agreement more broadly and act as logical checks to be correlated with the existing measures. The first item asks how much agreement there was across the 12 statements among the ratings of all Democrat respondents (for participants who viewed Democrat-leaning statements) or all Republican respondents (for participants who viewed Republican-leaning statements) on a 7-point Likert scale (1 = very little agreement, 7 = very strong agreement). This differs slightly from the original 100-point measure, which asks about perceived consensus between two party members. The second item had participants rate how much *disagreement* there was between the two groups of respondents (1 = very little disagreement, 7 = very strong disagreement) across the 12 items. The latter responses were reverse scored to reflect agreement.

Participants

Study 2 was conducted using the sample platforms and payment as Study 1, recruiting *N* = 340 American MTurk participants who did not participate in Study 1. Prior to analysis, 45 participants who failed a post-survey comprehension check were excluded, leaving *N* = 286 with complete data. Of these, 57.7% identified as a Democrat, 22.4% as Republican, and 19.2% as Independent. The remaining 0.2% did not report affiliation. Like Study 1, participants were mostly white (71%), identified as male (55.6%), and were over the age of 18, with most (82.5%) indicating they were between the ages of 25-54.

Participants were again assigned to view one of three sample types: Disclosed No Bias, Disclosed Bias, or Undisclosed Bias.

Study Instructions

The first two sample types replicate their counterparts in Study 1, while the third – Undisclosed Bias – is a hybrid of the two. These participants were shown extremely partisan samples but were instructed that they were unbiased.

These participants were informed of the deception used post-study and re-consented to having their data included for analysis, per the local ethics committee guidelines (all re-consented participants assented). Full instructions for Study 2 are available in Appendix D.

Results

For these *N* = 286 participants, *n* = 142 viewed Democrat-leaning statements and *n* = 144 viewed Republican-leaning statements. Across sample types, *n* = 97 were assigned to the Disclosed Bias sample condition, *n* = 94 to the Disclosed No Bias condition, and *n* = 95 to the Undisclosed Bias condition.

Belief Correction

A univariate factorial ANOVA of the impact of sample type (Disclosed Bias, Undisclosed Bias, Disclosed No Bias), target party (Democrat, Republican), party affiliation (Democrat, Republican, Independent), and affiliation strength (Low, High) on belief correction found that for the Disclosed Bias and Disclosed No Bias sample types, estimates of party agreement and the magnitudes of belief adjustment were similar to those elicited in Study 1. The difference between estimated party agreement and the sample mean viewed varied across sample types, *F*(2, 246) = 52.30, *MSE* = 4930.27, *p* < .001, $\eta^2{}_p$ = 0.30. Tukey post hoc tests revealed that the magnitude of this correction was largest in the Disclosed Bias condition (*M* = -13.80), exceeding that of participants who viewed Undisclosed Bias (*M* = -8.20; *p* = .001) and Disclosed No Bias (*M* = 3.59; *p* < .001) samples. There were no main effects of target party (*F*(1, 246) = 2.03, *MSE* = 191.53, *p* = .155), partisan affiliation, *F*(2, 246) = 1.56, *MSE* = 147.32, *p* = .212), or affiliation strength (*F*(1, 246) = 1.70, *MSE* = 160.59, *p* = .193), and there were no two- or three-way interactions between these variables. Full model results are reported in Table 2. Average estimates of party agreement are depicted in Figure 5 (pooled across target parties) and Appendix F (stratified by target party).

Two-tailed independent t-tests showed that all three study conditions provided significantly different point estimates of party agreement based on the sample they viewed. The Undisclosed Bias estimates were largest (M = 79.71), differing from the Disclosed Bias (M = 74.73; $t(190) = 3.25$, $p = .001$, $d = 0.47$). The Disclosed Bias condition, in turn, differed from the Disclosed No Bias estimates (M = 70.40; $t(189) = 3.25$, $p = .001$, $d = 0.47$).

One-sample t-tests indicated that like Study 1, the Disclosed Bias point estimates (M = 74.73) did not differ from those expected had participants viewed no samples at all (i.e., the No Sample control participants of Study 1 (M = 72.82)), $t(96) = 1.82$, $p = .072$). Only the Disclosed No Bias (M = 70.40) and Undisclosed Bias (M = 79.71) estimates differed significantly from the control estimates (Disclosed No Bias: $t(93) = 2.99$, $p = .004$, $d = 0.31$; Undisclosed Bias: $t(94) = 6.16$, $p < .001$, $d = 0.63$).

Table 2.

Results of Factorial ANOVA predicting Belief Correction

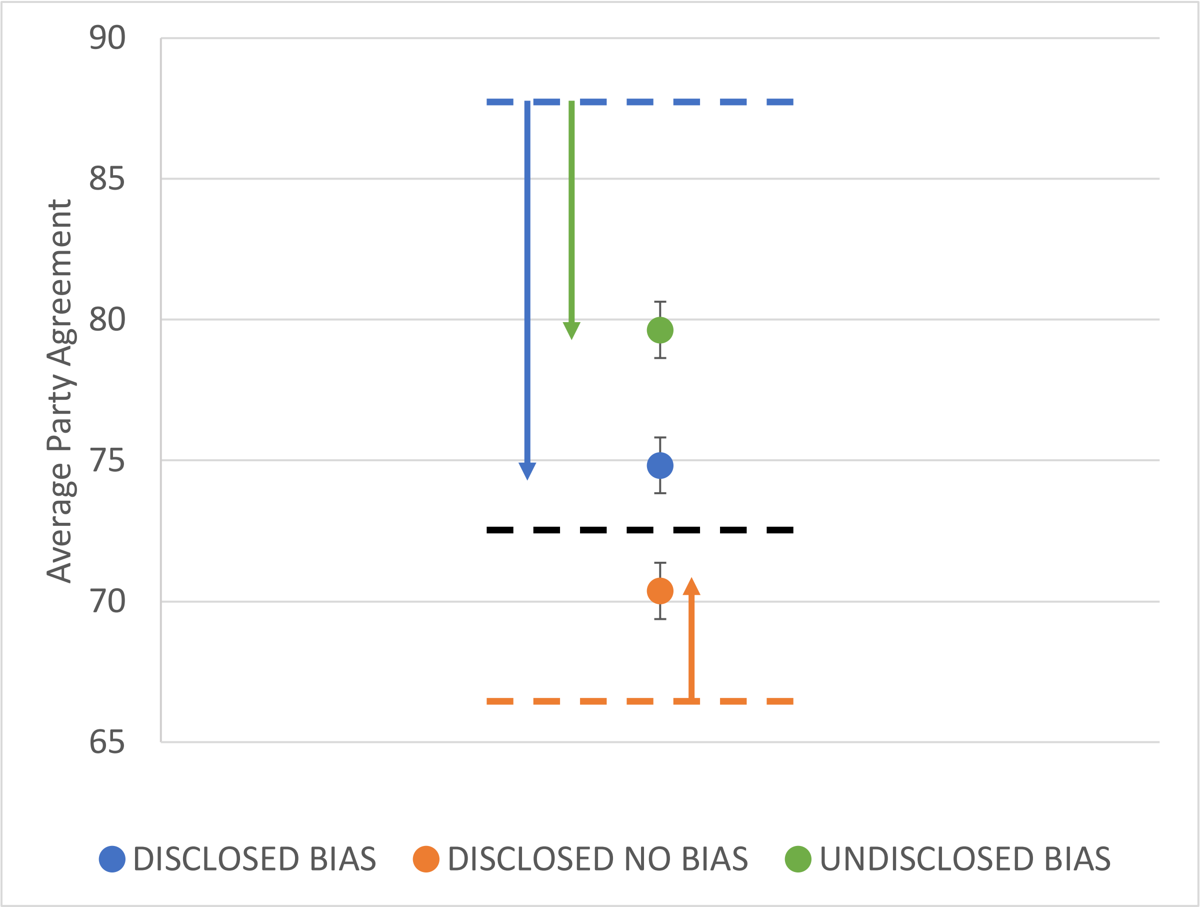
Predictor Variables	df	Mean Square	*F*	*p*	Partial Eta Squared (η^2_p)
Sample Type	2	4930.27	52.30	**<.001**	.298
Target Party	1	191.53	2.03	.155	.008
Affiliation Strength	1	160.59	1.70	.193	.007
Party Affiliation	2	147.32	1.56	.212	.013
Sample Type * Target Party	2	21.19	.23	.799	.002
Sample Type * Affiliation Strength	2	52.17	.55	.576	.004
Party Affiliation * Sample Type	4	66.09	.70	.592	.011
Target Party * Affiliation Strength	1	0.88	.01	.923	.000
Party Affiliation * Target Party	2	146.11	1.55	.214	.012

Party Affiliation * Affiliation Strength	2	79.01	.84	.434	.007
Sample Type * Target Party * Affiliation Strength	2	4.82	.05	.950	.000
Party Affiliation * Sample Type * Target Party	4	84.65	.90	.466	.014
Sample Type * Party Affiliation * Sample Type * Affiliation Strength	4	45.37	.48	.750	.008
Party Affiliation * Target Party * Affiliation Strength	2	22.25	.24	.790	.002
Party Affiliation * Sample Type * Target Party * Affiliation Strength	4	91.53	.97	.424	.016
Error	246	94.27			
Total	282				

Note. **Bolded** *p*-values are significant at the <.001 level.

Figure 5

Pooled Average Perceived Agreement of Democrats and Republicans



Note. Error bars are +/- 1 SE. Blue dashed line is the sample mean shown to participants who viewed a Disclosed Bias sample. Orange dashed line is the sample mean shown to participants who viewed a Disclosed No Bias sample. Black dashed line is the average estimate of group agreement provided by participants in the Study 1 No Sample (control) group. Arrows depict the difference between the sample mean shown and the average estimate of party agreement.

Accuracy

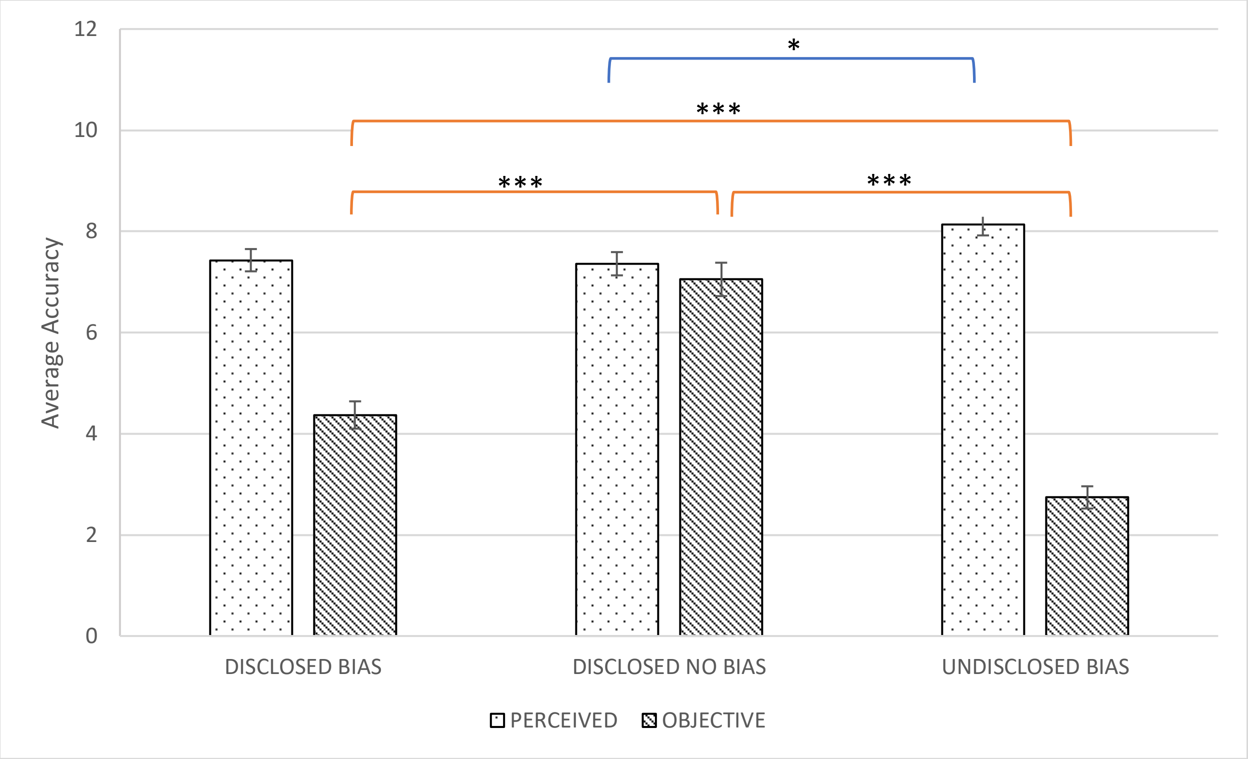
A univariate factorial ANOVA explored the impact of target party (Democrat, Republican) and sample type (Disclosed Bias, Disclosed No Bias, Undisclosed Bias) on perceived and objective hit rates (possible ranges: 0-12). Building on the results of Study 1, perceived accuracy did not vary between the Democrat versus Republican target parties, *F*(1,

280) = 0.91, *MSE* = 4.33, p = .342. There was a significant – but small – main effect of sample type on perceived accuracy (*F*(2, 280) = 3.64, *MSE* = 17.38, p = 0.028, η^2_p = 0.03) where participants in the new Undisclosed Bias condition (*M* = 8.13) believed they were more accurate than participants in the Disclosed No Bias condition (*M* = 7.36), p = .041 (see Figure 6). There was no sample type by list interaction, *F*(2, 280) = 0.25, *MSE* = 1.19, p = .780.

Considering true accuracy, we replicated the main effect of sample type, *F*(2, 280) = 62.57, *MSE* = 442.81, p <.001, η^2_p = 0.31. Tukey post hoc tests indicated that all sample types differed significantly from each other, all p < .001, where participants who viewed Disclosed No Bias samples gave the most accurate ratings of party belief (*M* = 7.05) compared to those who viewed Disclosed Bias (*M* = 4.39) or Undisclosed Bias (*M* = 2.76) samples. There was no variability in accuracy among the two target parties (*F*(1, 280) = 0.004, *MSE* = 0.03, p = .948). A small sample type by target party interaction (*F*(2, 280) = 3.42, *MSE* = 24.17, p = .034, η^2_p = 0.02) indicated that people in the Undisclosed Bias condition were slightly more accurate when estimating Republican (*M* = 3.35) than Democrat (*M* = 2.16) beliefs, but otherwise there were no other differences among sample type conditions.

Figure 6

Perceived versus Objective Hit Rates across Sample Types



Note. Error bars are +/- 1 SE. Significance at the .05, .01, and .001 levels are denoted by *, **, and ***, respectively.

Perceived Consensus

Spearman's rank order correlations (r_s) showed that the new 7-point measure of perceived consensus among all members of the same party correlated strongly with the 100-point measure of perceived consensus among two members of the same party, $r_s(285) = 0.519$, $p <$ .001. Overall, a factorial ANOVA of perceptions of consensus among members of the same party found main effects of sample type among both scales, (0-100 scale: $F(2, 279) = 3.94$, $MSE =$ 1236.30, $p = 0.021$, $\eta^2_p = 0.03$; 1-7 scale: $F(2, 280) = 13.11$, $MSE = 9.39$, $p <.001$, $\eta^2_p = 0.09$; see Figures 7a and 7b). For both scales, this effect was similar across target parties, (0-100 scale: $F(1, 279) = 0.36$, $MSE = 113.93$, $p = .547$; 1-7 scale: $F(1, 279) = 0.63$, $MSE = 0.45$, $p = .429$).

Tukey post hoc tests indicated that the Disclosed No Bias condition perceived less consensus compared to participants in the Undisclosed Bias condition (0-100 scale: *p* = .016; 1-7 scale: *p* < .001). Notably, perceptions of consensus were similarly high among the Disclosed Bias and Undisclosed Bias conditions across both scales (0-100 scale: *p* = .353, 1-7 scale: *p* = .190).

Figure 7a

Average Perceived Consensus of Two Members of the Same Party (0-100 scale)

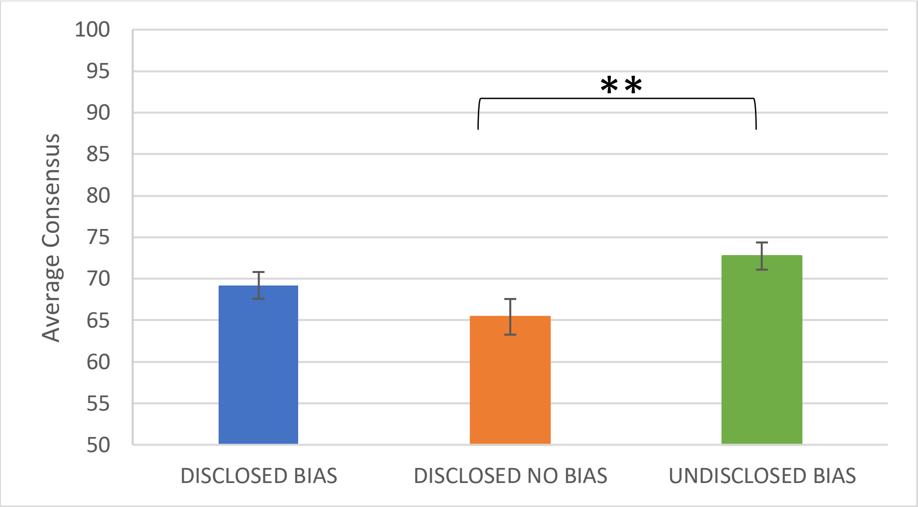
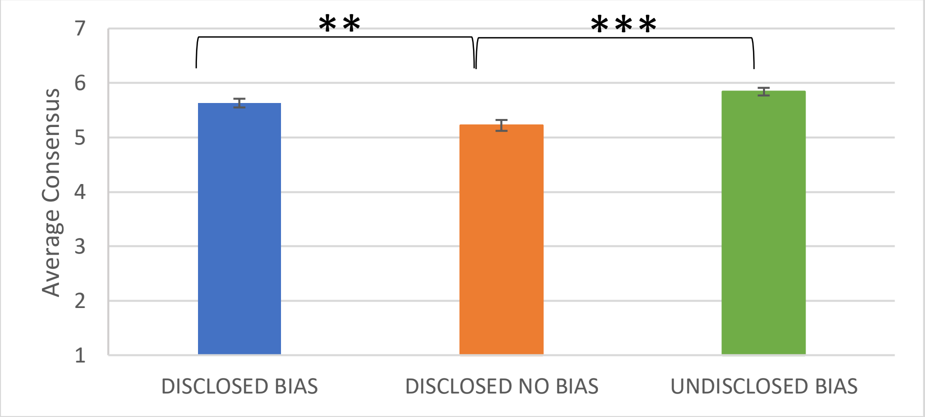


Figure 7b

Average Perceived Consensus of Two Members of the Same Party (1-7 scale)



Note. Error bars are +/- 1 SE. Significance at the .05, .01, and .001 levels are denoted by *, **, and ***, respectively.

Like Study 1, perceived consensus between members of opposing parties did not vary by sample type when rated on a 0-100 or 1-7 scale (0-100 scale: *F*(2, 280) = 0.36, *MSE* = 210.27, *p* = .701; 1-7 scale: *F*(2, 279) = 0.87, *MSE* = 2.15, *p* = .420; See Figures 8a and 8b).

Figure 8a

Average Perceived Consensus of Two Members of Opposing Parties (0-100 scale)

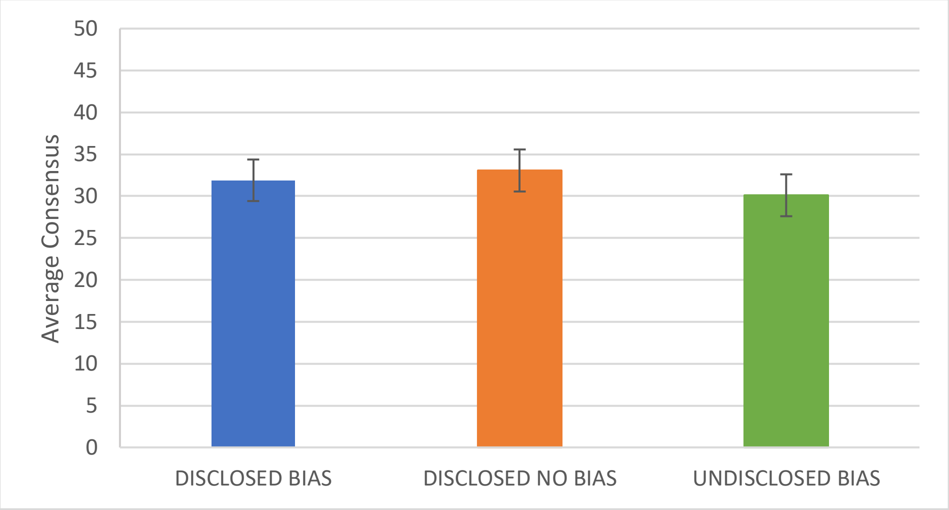
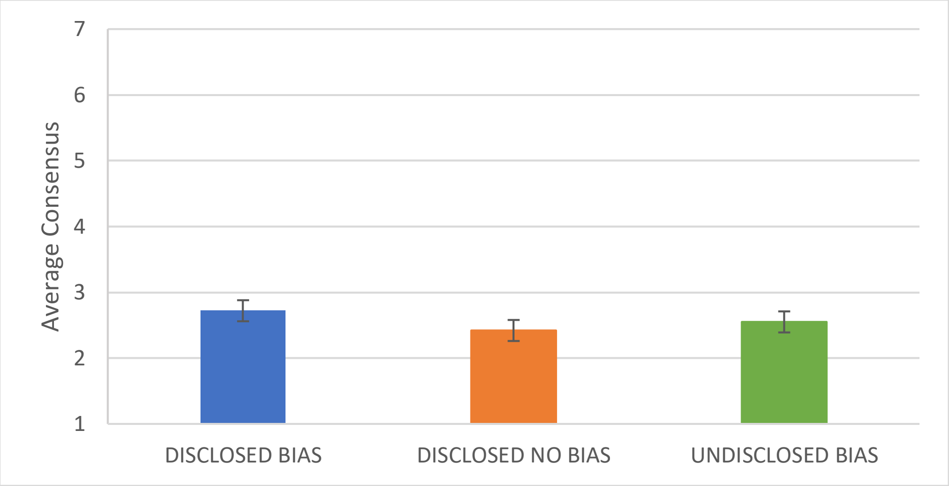


Figure 8b

Average Perceived Consensus of Two Members of Opposing Parties (1-7 scale)



Numeracy

A simple one-way ANOVA showed that performance on the Berlin Numeracy Test was generally moderate (*M*= 2.05, *SD* = 1.24) and that no condition was more or less numerate than another, *F*(2, 281) = 0.25, *MSE* = 0.39, *p* = .776. Spearman's rank order correlations (r_s) found no association between numeracy and participants' estimates of party belief (r_s(284) = .047, *p* = .426) or how much they corrected their estimates (r_s(284) = .036, *p* = .541). Overall, numeracy was positively associated with accuracy (r_s(284) = .129, *p* = .030), though this relationship is considered weak to negligible (Akoglu, 2018).

Discussion

In Study 2, we introduced a new experimental condition to help differentiate the effect of explicit awareness of selection bias versus anchoring to presented beliefs. Replicating Study 1, participants who were exposed to extremely partisan agreement ratings gave significantly lower estimates of party agreement compared to the biased beliefs they viewed. This correction was more pronounced when participants were aware of the bias compared to unaware. Participants who viewed extreme beliefs but viewed them as representative also provided estimates much lower than the samples they viewed, however, participants aware of the bias reduced these views much more, indicating an explicit attempt at correction. On the other hand, those who viewed representative ratings of agreement estimated closely to the presented beliefs, slightly overestimating party agreement. This overestimation, however, was consistent with the general public's perceptions of party agreement. Like Study 1, political partisanship did not predict estimate correction, and the patterns of results remained the same.

The same is true for perceived accuracy, where all group conditions viewed their performance similarly. The only significant difference was between participants unaware of partisan bias estimating that they achieved marginally more hits than those in the Disclosed No Bias group – a less than 1-point difference on the 12-point scale. Objective accuracy results replicated those of Study 1, where participants who saw biased scores – whether they were aware of this bias or not – tended to be overconfident and least accurate, whereas participants who saw unbiased, accurate sample beliefs were far better calibrated.

Furthermore, the new and existing measures of consensus were well-correlated, indicating good understanding of the task. Viewing unrepresentative agreement ratings elevated perceptions of consensus among members of the same party, though this effect was small. Consistent with the prior study, there was no difference in perceived consensus between members of opposing parties, regardless of the target party shown.

In summary, the new experimental condition reaffirmed the effect of selection bias on estimates of party agreement. Additionally, judgment accuracy was reduced by exposure to biased samples of population agreement – regardless of whether participants were aware of this underlying bias or not. Biased samples of beliefs also differentially impacted perceptions of political group consensus, though only among perceptions within parties.

General Discussion

Political polarization has become a pervasive threat to informed decision making. Hyperpartisan content in particular can exacerbate polarization and misinformed prototypical inferences of partisan groups (Ahler & Sood, 2018), motivating people to spread derogatory content and political misinformation (Osmundsen et al., 2021). Debunking these misperceptions has been shown to reduce partisan animosity (Moore-Berg et al., 2020). In a similar vein, the current studies aimed to investigate selection bias as a possible contributor to this political polarization. In particular, whether presenting extremely partisan content impacts perceptions of Democrats and Republican beliefs when this selection bias is explicitly salient. Broadly, we found that when sample bias is explicitly disclosed, exposure to extremely partisan beliefs can still lead to overestimation, overconfidence, and inflated perceptions of homogeneity among members of the same political party. The magnitude and direction of these effects were largely replicated between studies.

Belief Correction

It is generally challenging to revert the beliefs of people who have encountered unrepresentative information to a pre-exposure baseline (Kan et al., 2021; Lewandowsky et al., 2012). People anchor judgments to the initial values presented to them, often adjusting insufficiently away from those anchor values, regardless of whether they are rewarded for accuracy (Tversky & Kahneman, 1974). With this in mind, we hypothesized that the biasing effect of seeing strong Democrat and Republicans agreement with various topics would lead to significant – but insufficient – correction of perceptions of population agreement. While our findings support this hypothesis, the magnitude of debiasing was nevertheless large, supporting prior literature on the benefits of detailed debunking messaging, bias disclosure, and source salience (Chan et al., 2017; Jost et al., 2020; Siebert & Siebert, 2023; Stapel et al., 1998; Traberg et al., 2022). But, akin to previous work (Chan et al., 2017; Ecker et al., 2022; Lewandowsky et al., 2012), the distorted samples remained influential, regardless of whether the underlying bias was disclosed. Like prior research has indicated (Ahler & Sood, 2018), these effects were not driven by participants' individual numeracy skills.

One may argue that group differences in belief correction results can be interpreted differently. On one hand, correction was insufficient enough that participants viewing these scores still significantly overestimated true agreement. On the other, however, point estimates were no different from the participants who did not view external beliefs. That is, they provided estimates no different than would be expected in the general population, which may already hold inflated views of partisan beliefs. This supports previous research reporting that even under neutral conditions, people will correct for presented information (Petty & Wegener, 1993). When given additional context meant to debunk misleading information, people are generally good at updating their beliefs to a level similar to one had they never encountered the biasing information in the first place (Kan et al., 2021). We could conclude that disclosing selection bias leads people to change their judgments sufficiently enough to return to their "baseline" judgment – partially refuting the aforementioned work on the persistent impacts of misleading content – especially given that participants viewing the same biased samples unaware of their source performed far worse. Yet, when compared to participants exposed to accurate beliefs, both participants exposed to skewed agreement beliefs and those not shown any sample beliefs gave significantly inflated group estimates whether they were aware of the bias or not. While participants viewing representative information also overestimated true population agreement, their estimates were far closer anchored – on average, less than a 4-point difference from true belief – much closer than the other group conditions, and far more accurate. This may be a timely indicator that the general public already holds significant misperceptions of Democrat and Republican beliefs. Disclosing the bias underlying selected hyperpartisan content may be a good first step to debiasing, but the results of this study underscore the human tendency to rely on the information they're presented, underscoring the importance of disseminating factual content.

As hypothesized, participants who saw biased information were less accurate in their estimations compared to those who saw unbiased information, distorting the ability to accurately perceive the distribution of opinions within political parties. What was interesting to find, however, was a differential effect of bias awareness on perceived accuracy. Participants unaware of selection bias believed their task performance was significantly better than those who were aware of it; they were also the least accurate, though both groups performed far worse than those exposed to representative information. You could also frame these results to say participants in the Disclosed Bias group had the largest magnitude of belief correction – improving accuracy the most. While any degree of belief correction in the right direction is advantageous, their final judgments remained the least accurate overall. Parsing improvement from overall performance provide two valuable indicators of intervention efficacy.

Overall, it appears that exposure to extremely partisan beliefs – whether selection bias is disclosed or not – leads to perceptions of strong task performance, but in reality far less accurate judgments, à la Dunning-Kruger (Kruger & Dunning, 1999). While associated with judgment stability, perceived accuracy and objective accuracy are often poorly calibrated (Keren, 1991; Moore & Schatz, 2017). Previous research has found that confidence can influence susceptibility to misleading information (Auslander et al., 2017; Lyons et al., 2021). While we cannot draw conclusions on the causality of this relationship in the present investigation, it nonetheless highlights the importance of affect in decision making. With false and biased information becoming more widespread – little of which is explicitly contextualized or debunked (Guay et al., 2023) – so, too, could the general public's confidence in their perceptions of truth. Higher confidence increases the likelihood of seeking biased information (Kaanders et al., 2022), which in turn boosts confidence in one's beliefs (Jiwa et al., 2023).

Finally, the effect of extremely partisan beliefs on belief correction was not driven by participants' individual numeracy skills, as supported by prior research (Ahler & Sood, 2018).

Perceived Consensus

Regarding consensus, our results also support our hypothesis that viewing unrepresentatively high agreement ratings can boost people's perceptions of cohesion between party members, as has been found previously (Cook et al., 2017; Kohl et al., 2016; Lewandowsky et al., 2013; S. L. van der Linden, Clarke, et al., 2015). Interestingly, this effect occurred when assessing members of the same party but not opposing parties, where perceived consensus was similar low across group conditions. We found no evidence that the congruence between participants' own political affiliation and the party they evaluated affected perceptions of party consensus, nor did consensus judgments vary across target parties. Though partisan identity is generally found to modify perceptions of scientific consensus (Bromme et al., 2022; Nagler et al., 2020; Nicolo et al., 2023), the relationship between partisanship and misleading information is not consistent (Mazepus et al., 2023). This is a potential indication that our task led to intervention-driven judgments of consensus rather than politically-motivated.

Furthermore, ratings of within-party consensus for the disclosed and undisclosed bias conditions were just as high as the no sample control group; only the unbiased condition estimated party consensus was lower. This supports prior research that people generally hold inflated perceptions of party consensus (Ahler & Sood, 2018) which are at least partially mitigable by exposure to accurate data. It is also possible that being given sample beliefs for only one party made participants feel more knowledgeable judging that party's consensus, where being presented with such "one-sided" evidence makes judgments of that party more salient compared to judgments of the opposing party (Brenner et al., 1996), leading participants to revert to a common prior when asked to draw comparisons between them.

The finding that interparty consensus ratings were similarly low for participants who saw samples of beliefs versus no samples also supports that participants hold a common perception of diverging consensus that were unaffected by viewing sample beliefs of only one party, a hypothesis worth further investigation. Follow-up studies that present participants with both parties' beliefs and that collect confidence in judgments of consensus would also help tease apart these precipitates.

Limitations and Future Directions

Despite the insights gained from this study, there are important limitations that must be considered for refining future research. First, the primary stimuli used in this study were validated in US Census-matched political subpopulations. While advantageous in its national representativeness, it limits the generalizability of our results to other countries, though the framework of this study is easily transferable to other relevant datasets. While our study did not seek to emphasize partisan identity, the strength of partisan perceptions can vary internationally (Ruggeri et al., 2021). In a recent global study of cooperation, national identity, and policy support during the COVID-19 pandemic, Azevedo et al. (2023) found that compared to the other world regions sampled, Europe exhibits lower collective narcissism and national identity, domains that could influence participants' baseline perceptions of political group beliefs and their willingness to adjust them. Future work could collect and control for participants' baseline beliefs while exploring these effects across different regions, especially those where the divergence between political identities is more (or less) distinct.

Given that few results differed when evaluating Democrats or Republicans, we cannot conclude whether these results would differ if we addressed the Republican underrepresentation, though future work can very simply counterbalance aspects of partisan recruitment to investigate this. As is typical for studies utilizing MTurk, our participants were younger and more liberal, akin to distributions of partisan identity found by Levay et al. (2016). Prior research, however, indicates that MTurk samples hold ideological and motivational characteristics similar to the general population, supporting their validity in political research; any differences – which are small – also diminish with increasing conservatism (Clifford et al., 2015).

Furthermore, strength of partisanship did not impact judgments, nor did results differ when evaluating the Democrat or Republican target party. This is to be expected, given that like other studies, most of our sample (approximately two-thirds) did not identify as strongly partisan (Levay et al., 2016). Strong political convictions undermine belief change (Zwicker et al., 2020), so a follow-up study targeting a more strongly partisan sample could elicit smaller effect sizes than ours. However, it is also possible that our studies' effects were unaffected by partisan strength. Instead, they may indicate that our instructions to evaluate distributions of beliefs encouraged participants to attend more to the statistical nature of the task – as we intended – and less on the content of the statements read, decreasing the likelihood that participants relied on pre-existing political attitudes to guide their estimates. Such analytical processing can shroud partisan bias (Bago et al., 2020).

Alternatively, it is likely that participants hold stronger than average opinions on some political issues but weak opinions on others. Domain-specific salience would be difficult to clearly observe among the diversity of our stimuli. As we did not collect baseline beliefs in this study, we cannot conclude whether prior beliefs had item-specific or global impacts on judgments. Generally, people favour their party values over personal belief (Van Bavel & Pereira, 2018), though these values are predicted by personal belief (Vandeweerdt, 2022). Future work should collect baseline item or domain-specific beliefs to explore whether participants react differently as a function of their own views - for instance, as a function of the congruence between their own beliefs and the presented information – aside from their partisan identity. Using our 100-point scale as an example, a participant with a personal agreement of 20 on a given statement may respond to sample agreement ratings in the 90s differently than a participant with a baseline view of 80 viewing the same scores. In the face of disconfirming evidence, polarization strengthens, however this research is mixed (Balietti et al., 2021; Kubin & von Sikorski, 2021), with some suggesting that the impact of evidence on polarization and belief update may vary according to baseline personal belief (Bago et al., 2020; Dalege & van der Does, 2022; Sunstein et al., 2016) and perceived shifts in party beliefs (Busch, 2016).

Conclusion

This study offers insights into the complex interaction of exposure to false beliefs and awareness of this bias with inferences about political parties. Across two exploratory studies, we replicated findings that even when aware of sample bias, exposure to highly partisan beliefs can lead to persistent overestimation, overconfidence, and inflated perceptions of party homogeneity.

A recent review called on more interdisciplinary research testing interventions aimed at reducing partisan animosity (Hartman et al., 2022) as a means of reducing political polarization.

Our study lends supports for this recommendation, where explicitly debunking the source and selection method of biased data is largely beneficial in debiasing judgments, in line with related works (Gretton et al., 2021; Lewandowsky & van der Linden, 2021; S. van der Linden et al., 2017) and a method already implemented on social media (e.g., Twitter attaching warnings and additional context to disputed posts).

This debiasing may, in turn, reduce partisan animosity, though further research is needed. There are already, however, promising works suggesting that correcting misguided intergroup perceptions can reduce polarization, partisan violence, and reactance to societal norms (Braley et al., 2022; Mernyk et al., 2022). Furthermore, that targeting misinformed perceptions of the attitude prevalence among group members can reduce partisan animosity (Lees & Cikara, 2019; Ruggeri et al., 2021; Voelkel et al., 2021). Disseminating correct information where possible appears to be the ideal approach when judgment accuracy is the priority. Yet, it appears that even when given accurate evidence, there remains a general tendency to overestimate these judgments. Furthermore, the beneficial effects of exposure to corrective information – while significant – can be short-lived (Nyhan, 2021). Future research should explore these effects over time, targeting the relationship between presented claims and perceptions of partisan groups to improve longitudinal accuracy and debiasing in tandem. These and the discussed insights from these studies can aid addressing the challenges posed by growing political polarization in the digital age.

References

Ahler, D. J. (2014). Self-Fulfilling Misperceptions of Public Polarization. *The Journal of Politics, 76*(3), 607–620. https://doi.org/10.1017/S0022381614000085

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. https://doi.org/10.1086/697253

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Anthony, A., & Moulding, R. (2019). Breaking the news: Belief in fake news and conspiracist beliefs. *Australian Journal of Psychology, 71*(2), 154–162. https://doi.org/10.1111/ajpy.12233

Armaly, M. T., & Enders, A. M. (2023). Filling in the Gaps: False Memories and Partisan Bias. *Political Psychology*, *44*(2), 281–299. https://doi.org/10.1111/pops.12841

Azevedo, F., Pavlović, T., Rêgo, G. G., Ay, F. C., Gjoneska, B., Etienne, T. W., Ross, R. M., Schönegger, P., Riaño-Moreno, J. C., Cichocka, A., Capraro, V., Cian, L., Longoni, C., Chan, H. F., Van Bavel, J. J., Sjåstad, H., Nezlek, J. B., Alfano, M., Gelfand, M. J., … Sampaio, W. M. (2023). Social and moral psychology of COVID-19 across 69 countries. *Scientific Data*, *10*(1), 272. https://doi.org/10.1038/s41597-023-02080-8

Badawy, A., Ferrara, E., & Lerman, K. (2018). Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 258–265. https://doi.org/10.1109/ASONAM.2018.8508646

Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, *149*(8), 1608–1613. https://doi.org/10.1037/xge0000729

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Balietti, S., Getoor, L., Goldstein, D. G., & Watts, D. J. (2021). Reducing opinion polarization: Effects of exposure to similar people with differing political views. *Proceedings of the National Academy of Sciences*, *118*(52), e2112552118. https://doi.org/10.1073/pnas.2112552118

Bastos, M. T., & Mercea, D. (2019). The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review*, *37*(1), 38–54. https://doi.org/10.1177/0894439317734157

Baumgaertner, B., Carlisle, J. E., & Justwan, F. (2018). The influence of political ideology and trust on willingness to vaccinate. *PLOS ONE*, *13*(1), e0191728. https://doi.org/10.1371/journal.pone.0191728

Braley, A., Lenz, G., Adjodah, D., Rahnama, H., & Pentland, A. (2022). *The Subversion Dilemma: Why Voters Who Cherish Democracy Participate in Democratic Backsliding* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-1766479/v1

Brenner, L. A., Koehler, D. J., & Tversky, A. (1996). On the evaluation of one-sided evidence. *Journal of Behavioral Decision Making*, *9*(1), 59–70. https://doi.org/10.1002/(SICI)1099-0771(199603)9:1<59::AID-BDM216>3.0.CO;2-V

Bromme, R., Mede, N. G., Thomm, E., Kremer, B., & Ziegler, R. (2022). An anchor in troubled times: Trust in science before and within the COVID-19 pandemic. *PLOS ONE*, *17*(2), e0262823. https://doi.org/10.1371/journal.pone.0262823

Busch, K. B. (2016). Estimating parties' left-right positions: Determinants of voters' perceptions' proximity to party ideology. *Electoral Studies*, *41*, 159–178. https://doi.org/10.1016/j.electstud.2016.01.003

Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science*, *28*(11), 1531–1546. https://doi.org/10.1177/0956797617714579

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics*, *2*(4), 205316801562207. https://doi.org/10.1177/2053168015622072

Cohen, G. L. (2003). Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs. *Journal of Personality and Social Psychology*, *85*(5), 808–822. https://doi.org/10.1037/0022-3514.85.5.808

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, *7*(1), 25–47. https://doi.org/10.1017/S1930297500001819

Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*, *12*(5), e0175799. https://doi.org/10.1371/journal.pone.0175799

Dalege, J., & van der Does, T. (2022). Using a cognitive network model of moral and social beliefs to explain belief change. *Science Advances*, *8*(33), eabm0137. https://doi.org/10.1126/sciadv.abm0137

Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How Elite Partisan Polarization Affects Public Opinion Formation. *American Political Science Review*, *107*(1), 57–79. https://doi.org/10.1017/S0003055412000500

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y

Enders, A. M., & Armaly, M. T. (2019). The Differential Effects of Actual and Perceived Polarization. *Political Behavior*, *41*(3), 815–839. https://doi.org/10.1007/s11109-018-9476-2

Frenda, S. J., Knowles, E. D., Saletan, W., & Loftus, E. F. (2013). False memories of fabricated political events. *Journal of Experimental Social Psychology*, *49*(2), 280–286. https://doi.org/10.1016/j.jesp.2012.10.013

Gadarian, S. K., Goodman, S. W., & Pepinsky, T. B. (2021). Partisanship, health behavior, and policy attitudes in the early stages of the COVID-19 pandemic. *PLOS ONE*, *16*(4), e0249596. https://doi.org/10.1371/journal.pone.0249596

Garrett, R. K., Long, J. A., & Jeong, M. S. (2019). From Partisan Media to Misperception: Affective Polarization as Mediator. *Journal of Communication*, *69*(5), 490–512. https://doi.org/10.1093/joc/jqz028

Graham, J., Nosek, B. A., & Haidt, J. (2012). The Moral Stereotypes of Liberals and Conservatives: Exaggeration of Differences across the Political Spectrum. *PLoS ONE*, *7*(12), e50092. https://doi.org/10.1371/journal.pone.0050092

Gretton, J. D., Meyers, E. A., Walker, A. C., Fugelsang, J. A., & Koehler, D. J. (2021). A brief forewarning intervention overcomes negative effects of salient changes in COVID-19 guidance. *Judgment and Decision Making*, *16*(6), 1549–1574. https://doi.org/10.1017/S1930297500008548

Guay, B., Berinsky, A. J., Pennycook, G., & Rand, D. (2023). How to think about whether misinformation interventions work. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-023-01667-w

Hartman, R., Blakey, W., Womick, J., Bail, C., Finkel, E. J., Han, H., Sarrouf, J., Schroeder, J., Sheeran, P., Van Bavel, J. J., Willer, R., & Gray, K. (2022). Interventions to reduce partisan animosity. *Nature Human Behaviour*, *6*(9), 1194–1205. https://doi.org/10.1038/s41562-022-01442-3

Hauser, D. J., Moss, A. J., Rosenzweig, C., Jaffe, S. N., Robinson, J., & Litman, L. (2022). Evaluating CloudResearch's Approved Group as a solution for problematic data quality on MTurk. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01999-x

Himelein-Wachowiak, M., Giorgi, S., Devoto, A., Rahman, M., Ungar, L., Schwartz, H. A., Epstein, D. H., Leggio, L., & Curtis, B. (2021). Bots and Misinformation Spread on Social Media: Implications for COVID-19. *Journal of Medical Internet Research*, *23*(5), e26933. https://doi.org/10.2196/26933

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. https://doi.org/10.1016/j.giq.2016.04.007

Jiwa, M., Cooper, P. S., Chong, T. T.-J., & Bode, S. (2023). Hedonism as a motive for information search: Biased information-seeking leads to biased beliefs. *Scientific Reports*, *13*(1), 2086. https://doi.org/10.1038/s41598-023-29429-8

Jones, D. R. (2001). Party Polarization and Legislative Gridlock. *Political Research Quarterly*, *54*(1), 125. https://doi.org/10.2307/449211

Jost, P. J., Pünder, J., & Schulze-Lohoff, I. (2020). Fake news—Does perception matter more than the truth? *Journal of Behavioral and Experimental Economics*, *85*, 101513. https://doi.org/10.1016/j.socec.2020.101513

Jungkunz, S. (2021). Political Polarization During the COVID-19 Pandemic. *Frontiers in Political Science*, *3*, 622512. https://doi.org/10.3389/fpos.2021.622512

Kahan, D. M. (2017). Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2973067

Kan, I. P., Pizzonia, K. L., Drummey, A. B., & Mikkelsen, E. J. V. (2021). Exploring factors that mitigate the continued influence of misinformation. *Cognitive Research: Principles and Implications*, *6*(1), 76. https://doi.org/10.1186/s41235-021-00335-9

Kennedy, J. (2019). Populist politics and vaccine hesitancy in Western Europe: An analysis of national-level data. *European Journal of Public Health*, *29*(3), 512–516. https://doi.org/10.1093/eurpub/ckz004

Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica*, *77*(3), 217–273. https://doi.org/10.1016/0001-6918(91)90036-Y

Kingzette, J., Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2021). How Affective Polarization Undermines Support for Democratic Norms. *Public Opinion Quarterly*, *85*(2), 663–677. https://doi.org/10.1093/poq/nfab029

Kohl, P. A., Kim, S. Y., Peng, Y., Akin, H., Koh, E. J., Howell, A., & Dunwoody, S. (2016). The influence of weight-of-evidence strategies on audience perceptions of (un)certainty when media cover contested science. *Public Understanding of Science*, *25*(8), 976–991. https://doi.org/10.1177/0963662515615087

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, *45*(3), 188–206. https://doi.org/10.1080/23808985.2021.1976070

Landry, A. P., Schooler, J. W., Willer, R., & Seli, P. (2023). Reducing Explicit Blatant Dehumanization by Correcting Exaggerated Meta-Perceptions. *Social Psychological and Personality Science*, *14*(4), 407–418. https://doi.org/10.1177/19485506221099146

Lees, J., & Cikara, M. (2019). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, *4*(3), 279–286. https://doi.org/10.1038/s41562-019-0766-4

Lelkes, Y., Sood, G., & Iyengar, S. (2017). The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect: EFFECT OF BROADBAND INTERNET ACCESS ON PARTISAN AFFECT. *American Journal of Political Science*, *61*(1), 5–20. https://doi.org/10.1111/ajps.12237

Lerman, K., Yan, X., & Wu, X.-Z. (2016). The “Majority Illusion” in Social Networks. *PLOS ONE*, *11*(2), e0147617. https://doi.org/10.1371/journal.pone.0147617

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The Demographic and Political Composition of Mechanical Turk Samples. *SAGE Open*, *6*(1), 215824401663643. https://doi.org/10.1177/2158244016636433

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lewandowsky, S., Gignac, G. E., & Vaughan, S. (2013). The pivotal role of perceived scientific consensus in acceptance of science. *Nature Climate Change*, *3*(4), 399–404. https://doi.org/10.1038/nclimate1720

Lewandowsky, S., & van der Linden, S. (2021). Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology*, *32*(2), 348–384. https://doi.org/10.1080/10463283.2021.1876983

Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2021). *A Systematic Review of Worldwide Causal and Correlational Evidence on Digital Media and Democracy* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/p3z9v

Mazepus, H., Osmudsen, M., Bang-Petersen, M., Toshkov, D., & Dimitrova, A. (2023). Information battleground: Conflict perceptions motivate the belief in and sharing of misinformation about the adversary. *PLOS ONE*, *18*(3), e0282308. https://doi.org/10.1371/journal.pone.0282308

Mernyk, J. S., Pink, S. L., Druckman, J. N., & Willer, R. (2022). Correcting inaccurate metaperceptions reduces Americans' support for partisan violence. *Proceedings of the National Academy of Sciences*, *119*(16), e2116851119. https://doi.org/10.1073/pnas.2116851119

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. https://doi.org/10.1073/pnas.2001263117

Nagler, R. H., Vogel, R. I., Gollust, S. E., Rothman, A. J., Fowler, E. F., & Yzer, M. C. (2020). Public perceptions of conflicting information surrounding COVID-19: Results from a nationally representative survey of U.S. adults. *PLOS ONE*, *15*(10), e0240776. https://doi.org/10.1371/journal.pone.0240776

Nicolo, M., Kawaguchi, E., Ghanem-Uzqueda, A., Soto, D., Deva, S., Shanker, K., Lee, R., Gilliland, F., Klausner, J. D., Baezconde-Garbanati, L., Kovacs, A., Van Orman, S., Hu, H., & Unger, J. B. (2023). Trust in science and scientists among university students, staff, and faculty of a large, diverse university in Los Angeles during the COVID-19 pandemic, the Trojan Pandemic Response Initiative. *BMC Public Health*, *23*(1), 601. https://doi.org/10.1186/s12889-023-15533-x

Nyhan, B. (2021). Why the backfire effect does not explain the durability of political misperceptions. *Proceedings of the National Academy of Sciences*, *118*(15), e1912440117. https://doi.org/10.1073/pnas.1912440117

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. *American Political Science Review*, *115*(3), 999–1015. https://doi.org/10.1017/S0003055421000290

Padgett, J., Dunaway, J. L., & Darr, J. P. (2019). As Seen on TV? How Gatekeeping Makes the U.S. House Seem More Extreme. *Journal of Communication*, *69*(6), 696–719. https://doi.org/10.1093/joc/jqz039

Pasek, M. H., Ankori-Karlinsky, L.-O., Levy-Vene, A., & Moore-Berg, S. L. (2022). Misperceptions about out-partisans' democratic values may erode democracy. *Scientific Reports*, *12*(1), 16284. https://doi.org/10.1038/s41598-022-19616-4

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Pereira, A., Harris, E., & Van Bavel, J. J. (2023). Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news. *Group Processes & Intergroup Relations*, *26*(1), 24–47. https://doi.org/10.1177/13684302211030004

Petty, R. E., & Wegener, D. T. (1993). Flexible Correction Processes in Social Judgment: Correcting for Context-Induced Contrast. *Journal of Experimental Social Psychology*, *29*(2), 137–165. https://doi.org/10.1006/jesp.1993.1007

Rao, A., Morstatter, F., & Lerman, K. (2022). Partisan asymmetries in exposure to misinformation. *Scientific Reports*, *12*(1), 15671. https://doi.org/10.1038/s41598-022-19837-7

Rathje, S., Robertson, C., Brady, W. J., & Van Bavel, J. J. (2022). *People think that social media platforms do (but should not) amplify divisive content* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/gmun4

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. https://doi.org/10.1073/pnas.2024292118

Rollwage, M., Zmigrod, L., de-Wit, L., Dolan, R. J., & Fleming, S. M. (2019). What Underlies Political Polarization? A Manifesto for Computational Political Psychology. *Trends in Cognitive Sciences, 23*(10), 820–822. https://doi.org/10.1016/j.tics.2019.07.006

Ruggeri, K., Većkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., Barea-Arroyo, P., Berge, M. L., Bjørndal, L. D., Bursalıoğlu, A., Bühler, V., Čadek, M., Çetinçelik, M., Clay, G., Cortijos-Bernabeu, A., Damnjanović, K., Dugue, T. M., Esberg, M., Esteban-Serna, C., … Folke, T. (2021). The general fault in our fault lines. *Nature Human Behaviour, 5*(10), 1369–1380. https://doi.org/10.1038/s41562-021-01092-x

Siebert, J., & Siebert, J. U. (2023). Effective mitigation of the belief perseverance bias after the retraction of misinformation: Awareness training and counter-speech. *PLOS ONE, 18*(3), e0282202. https://doi.org/10.1371/journal.pone.0282202

Sparkman, G., Geiger, N., & Weber, E. U. (2022). Americans experience a false social reality by underestimating popular climate policy support by nearly half. *Nature Communications*, *13*(1), 4779. https://doi.org/10.1038/s41467-022-32412-y

Stapel, D. A., Martin, L. L., & Schwarz, N. (1998). The Smell of Bias: What Instigates Correction Processes in Social Judgments? *Personality and Social Psychology Bulletin*, *24*(8), 797–806. https://doi.org/10.1177/0146167298248002

Statista. (2023). *Number of social media users worldwide from 2017-2027*. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Sunstein, C. R., Bobadilla-Suarez, S., Lazarro, S., & Sharot, T. (2016). How people update beliefs about climate change: Good news and bad news. *Cornell Law Rev.*, *102*, 1431.

Tokita, C. K., Guess, A. M., & Tarnita, C. E. (2021). Polarized information ecosystems can reorganize social networks via information cascades. *Proceedings of the National Academy of Sciences*, *118*(50), e2102147118. https://doi.org/10.1073/pnas.2102147118

Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2022). Psychological Inoculation against Misinformation: Current Evidence and Future Directions. *The ANNALS of the American Academy of Political and Social Science*, *700*(1), 136–151. https://doi.org/10.1177/00027162221087936

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & Tucker, J. A. (2021). Political Psychology in the Digital (mis)Information age: A Model of News Belief and

Sharing. *Social Issues and Policy Review*, *15*(1), 84–113. https://doi.org/10.1111/sipr.12077

Van Bavel, J. J., & Pereira, A. (2018). The Partisan Brain: An Identity-Based Model of Political Belief. *Trends in Cognitive Sciences*, *22*(3), 213–224. https://doi.org/10.1016/j.tics.2018.01.004

Van Bavel, J. J., Rathje, S., Harris, E., Robertson, C., & Sternisko, A. (2021). How social media shapes polarization. *Trends in Cognitive Sciences*, *25*(11), 913–916. https://doi.org/10.1016/j.tics.2021.07.013

van der Linden, S. L., Clarke, C. E., & Maibach, E. W. (2015). Highlighting consensus among medical scientists increases public support for vaccines: Evidence from a randomized experiment. *BMC Public Health*, *15*(1), 1207. https://doi.org/10.1186/s12889-015-2541-4

van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the Public against Misinformation about Climate Change. *Global Challenges*, *1*(2), 1600008. https://doi.org/10.1002/gch2.201600008

van der Linden, S., Roozenbeek, J., Maertens, R., Basol, M., Kácha, O., Rathje, S., & Traberg, C. S. (2021). How Can Psychological Science Help Counter the Spread of Fake News? *The Spanish Journal of Psychology*, *24*, e25. https://doi.org/10.1017/SJP.2021.23

Vandeweerdt, C. (2022). Someone like you: False consensus in perceptions of Democrats and Republicans. *Journal of Elections, Public Opinion and Parties*, *32*(3), 739–749. https://doi.org/10.1080/17457289.2021.1942891

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. https://doi.org/10.1177/0956797621995208

Voelkel, J. G., Chu, J., Stagnaro, M., Mernyk, J. S., Redekopp, C., Pink, S. L., Druckman, J., Rand, D. G., & Willer, R. (2021). *Interventions Reducing Affective Polarization Do Not Necessarily Improve Anti-Democratic Attitudes* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/7evmp

Weismueller, J., Gruner, R. L., Harrigan, P., Coussement, K., & Wang, S. (2023). Information sharing and political polarisation on social media: The role of falsehood and partisanship. *Information Systems Journal*, isj.12453. https://doi.org/10.1111/isj.12453

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. https://doi.org/10.1177/1745691615569849

Wilson, A. E., Parker, V. A., & Feinberg, M. (2020). Polarization in the contemporary political and media landscape. *Current Opinion in Behavioral Sciences*, *34*, 223–228. https://doi.org/10.1016/j.cobeha.2020.07.005

Yap, J. F. C. (2023). Response: Political polarization and its impact on COVID-19 vaccine acceptance. *Journal of Public Health*, fdad045. https://doi.org/10.1093/pubmed/fdad045

Yudkin, D., Hawkins, S., & Dixon, T. (2019). *The Perception Gap: How False Impressions are Pulling Americans Apart* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/r3h5q

Zimmerman, T., Shiroma, K., Fleischmann, K. R., Xie, B., Jia, C., Verma, N., & Lee, M. K. (2023). Misinformation and COVID-19 vaccine hesitancy. *Vaccine*, *41*(1), 136–144. https://doi.org/10.1016/j.vaccine.2022.11.014

Zmigrod, L. (2020). The role of cognitive rigidity in political ideologies: Theory, evidence, and future directions. *Current Opinion in Behavioral Sciences*, *34*, 34–39. https://doi.org/10.1016/j.cobeha.2019.10.016

Zwicker, M. V., van Prooijen, J.-W., & Krouwel, A. P. M. (2020). Persistent beliefs: Political extremism predicts ideological stability over time. *Group Processes & Intergroup Relations*, *23*(8), 1137–1149. https://doi.org/10.1177/1368430220917753

Appendix A: Stimuli of Vlasceanu et al. (2021)

Democrat-leaning Statements:

1. The US has loose gun laws.
2. The US government spends little for climate related research.
3. Obamacare has successfully decreased the number of uninsured Americans.
4. Embryonic stem cell therapy is a successful modern treatment method.
5. Colleges and Universities are having a positive effect on young generations' futures.
6. The Affordable Care Act saved the US a huge amount of money.
7. All cities in the US experience more extremely hot days compared to 50 years ago.
8. Children in the US are at high risk of witnessing gun violence.
9. The US allocates too much of the spending budget to Defense and Military.
10. Children raised by same-sex parents are just as adjusted as children raised by opposite-sex parents.
11. The amount government assistance to poor families in the US is not high enough.
12. Immigrant households in the US rarely access welfare programs.

Republican-leaning Statements:

1. The US is at great risk of illegal drug activity.
2. African American women get more abortions than Caucasian women.
3. Police use of force in the US is not causing that many deaths.
4. A large proportion of immigrants in the US are not in the workforce.
5. The amount of US corporate income taxes paid yearly is high.
6. A large number of undocumented workers are working illegally in the US.
7. Currently, foreign-born terrorists are a big threat to Americans in the US.
8. A large percentage of abortions in the US are paid for with public funds.
9. In the US, men and women are, on average, paid equally for the same job.
10. The US justice system is fair to racial minorities.
11. Government regulations have large costs for the US economy.
12. Small businesses owned by immigrants in the US do not provide that many jobs.

Appendix B: Study 1 Condition Instructions

Last year, researchers at Princeton University conducted a survey in which people were asked to rate how strongly they agreed with statements on a variety of issues. Survey respondents rated their agreement with each statement on a scale from 0 (completely disagree) to 100 (completely agree).

Survey respondents were also asked to identify themselves as Democrats or Republicans. Democrats tended to agree more than Republicans with some of the statements included in the survey (Democrat/left-leaning statements); Republicans tended to agree more than Democrats with other statements (Republican/right-leaning statements).

We will show you 12 Democrat[Republican] / left[right]-leaning statements. For each statement, your task is to estimate the average of the agreement ratings given by all Democrats[Republicans] who participated in the Princeton survey. Remember, these agreement ratings were given on a scale from 0 (completely disagree) to 100 (completely agree).

Disclosed No Bias Sample:

To help you with your task, for each statement, you will be shown agreement ratings of five survey respondents.

These five survey respondents were randomly selected from the set of all Democrat[Republican] participants in the Princeton survey. In short, you will be shown the agreement ratings of five survey respondents who were selected in an unbiased manner and who can therefore be considered representative of Democrats[Republicans] as a whole who completed the survey. Approximately 350 Democrats[Republicans] completed the survey, and the agreement ratings of the five Democrats[Republicans] you will be shown were selected at random from this group.

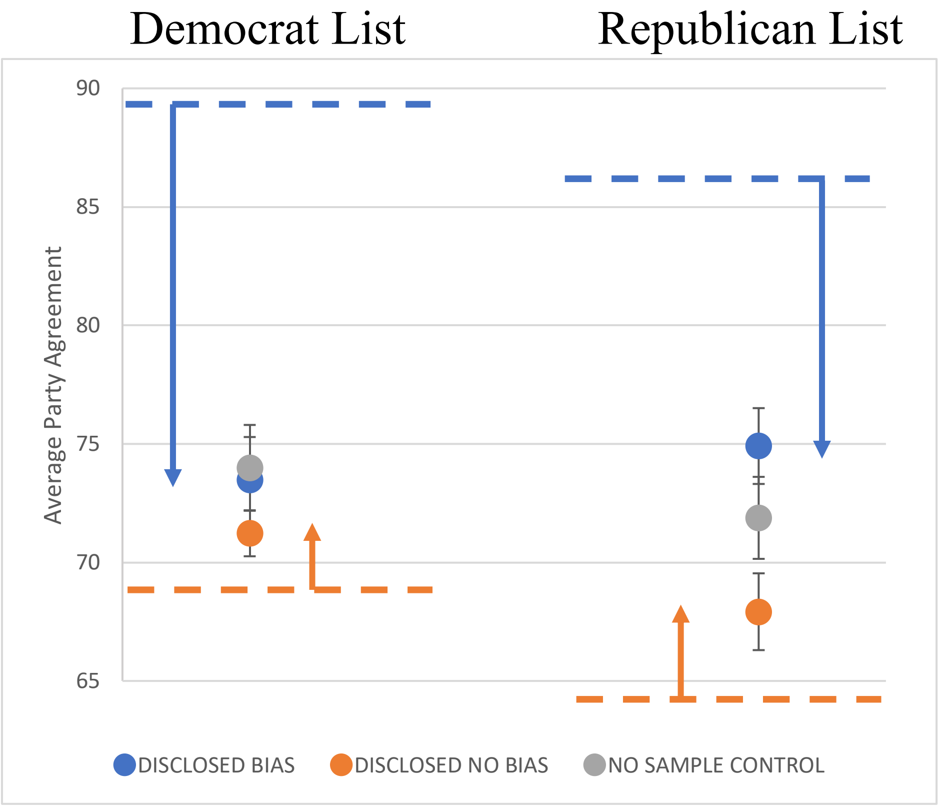
Disclosed Bias Sample:

To help you with your task, for each statement, you will be shown agreement ratings of five survey respondents.

These five survey respondents were randomly selected from the set of Democrats[Republicans] who agreed most strongly with the 12 Democrat[Republican] /left[right]-leaning statements. The average agreement rating across the 12 statements placed these survey respondents in the top 10% of Democrat[Republican] participants in the Princeton survey. In short, you will be shown the agreement ratings of five survey respondents who were selected in a biased manner and who cannot therefore be considered representative of Democrats[Republicans] as a whole who completed the survey. Approximately 350 Democrats[Republicans] completed the survey. Of those, the 35 Democrats[Republicans] who gave the highest average agreement ratings across the 12 statements (placing them in the top 10% of Democrats[Republicans]) were identified. The agreement ratings of the five Democrats[Republicans] you will be shown were selected at random from this smaller, selected group.

No Sample Control:
Approximately 350 Democrats completed the survey.

Appendix C: Study 1 – Average Perceived Agreement stratified by Target Party



Note. Error bars are +/- 1 SE. Blue dashed line is the sample mean shown to participants in the Disclosed Biased group. Orange dashed line is the sample mean shown to participants in the Disclosed No Bias group. Arrows depict the difference between the sample mean shown and the average estimate of group agreement.

Appendix D: Study 2 Condition Instructions

Disclosed No Bias Sample:
Same as Study 1

Disclosed Bias Sample:
Same as Study 1

Undisclosed Bias Sample:
…
To help you with your task, for each statement, you will be shown agreement ratings of five survey respondents.

These five survey respondents were randomly selected from the set of all Democrat[Republican] participants in the Princeton survey. In short, you will be shown the agreement ratings of five survey respondents who were selected in an unbiased manner and who can therefore be considered representative of Democrats[Republicans] as a whole who completed the survey. Approximately 350 Democrats[Republicans] completed the survey, and the agreement ratings of the five Democrats[Republicans] you will be shown were selected at random from this group.

Note: This condition uses the Disclosed Bias Condition stimuli with the Disclosed No Bias Condition instructions

Appendix E: Berlin Numeracy Test Stimuli (Multiple Choice Version)

Instructions: Finally, please answer the questions below. Do not use a calculator but feel free to use scratch paper for notes.

1) Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5).
   a) 5 out of 50 throws
   b) 25 out of 50 throws
   c) 30 out of 50 throws **
   d) None of the above

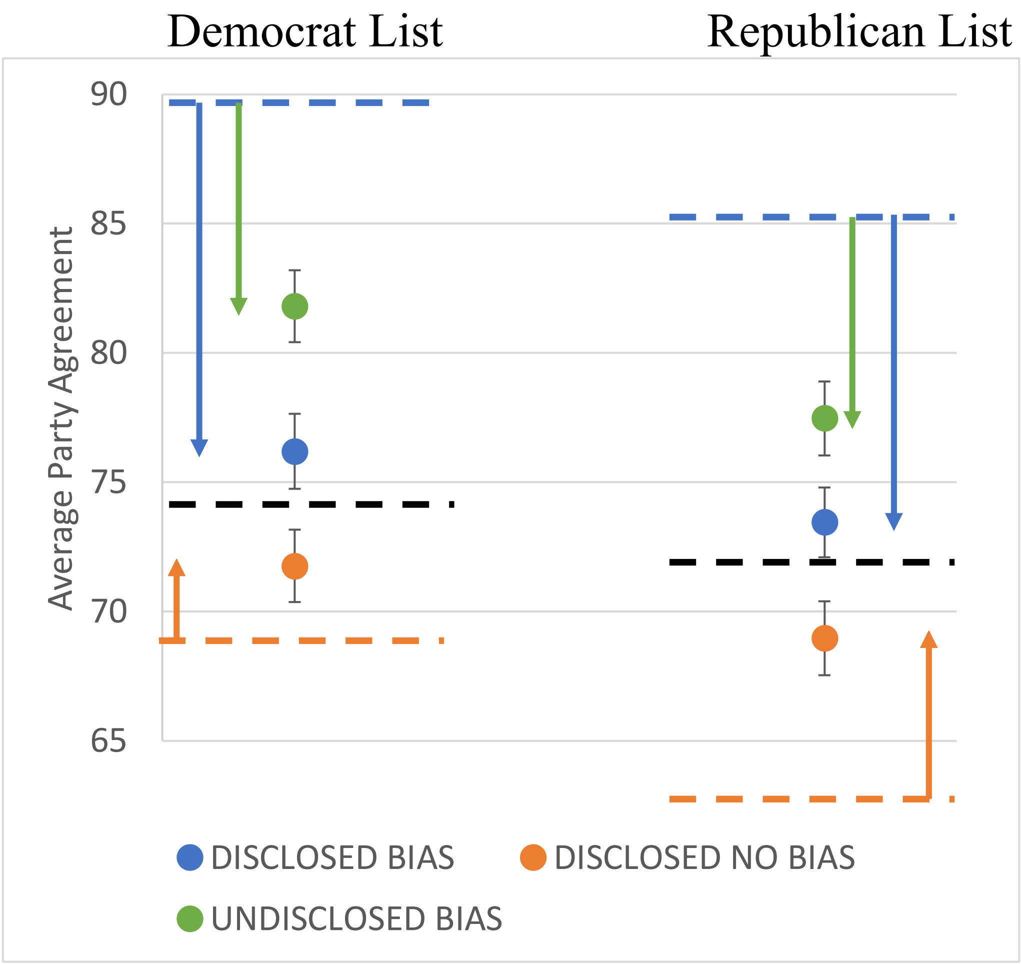
2) Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent.
   a) 10%
   b) 25% **
   c) 40%
   d) None of the above

3) Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws, about how many times would the die show the number 6?
   a) 20 out of 70 throws **
   b) 23 out of 70 throws
   c) 35 out of 70 throws
   d) None of the above

4) In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?
   a) 4 %
   b) 20 %
   c) 50 % **
   d) None of the above

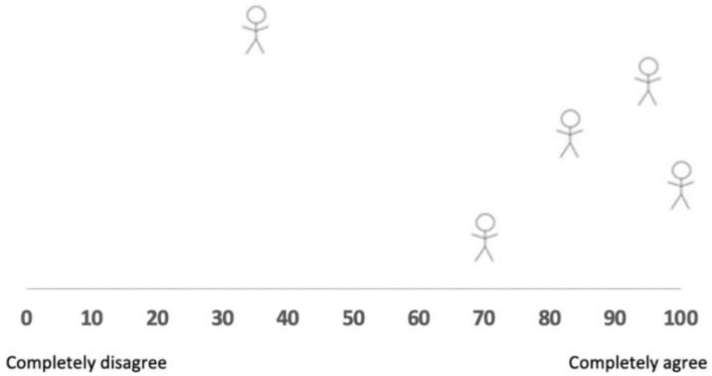
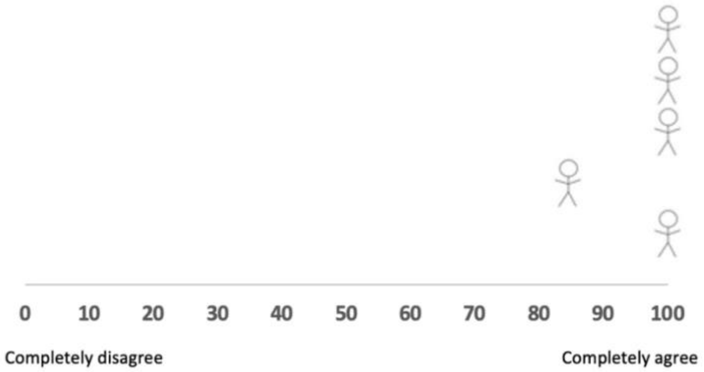
[Scoring = Count total number of correct (**) answers.]

Appendix F: Study 2 – Average Perceived Agreement stratified by Target Party

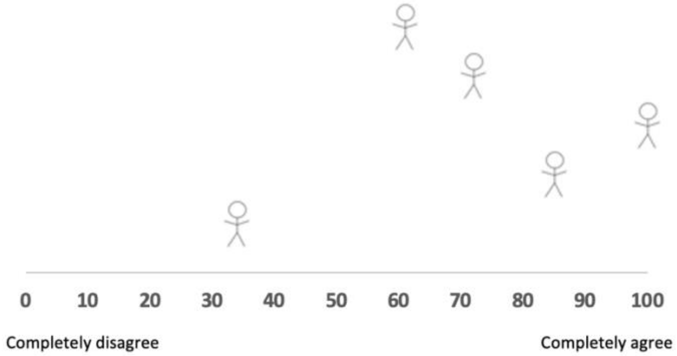


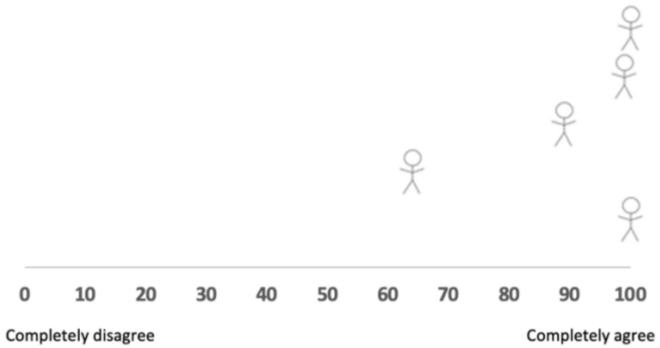
Note. Error bars are +/- 1 SE. Blue dashed line is the sample mean shown to participants in the Disclosed Bias group. Orange dashed line is the sample mean shown to participants in the Disclosed No Bias group. Black dashed line is the average estimate of group agreement provided by participants in the Study 1 No Sample (control) group. Arrows depict the difference between the sample mean shown and the average estimate of group agreement.

Appendix G: Item-level Stimuli shown to the Disclosed No Bias and Disclosed Bias Samples

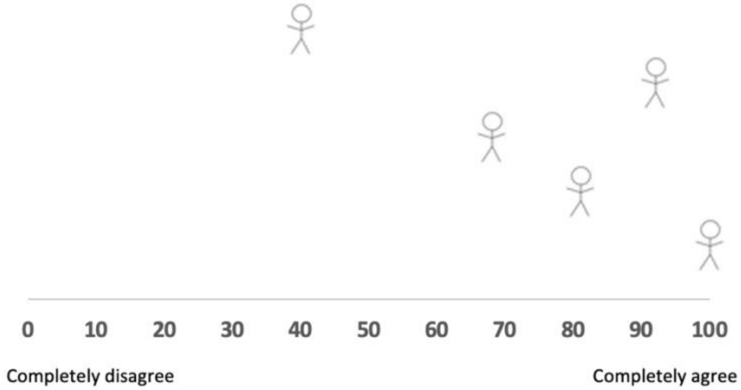
Item	Disclosed No Bias Condition (Representative Agreement Ratings)	Disclosed Bias Condition (Extremely Partisan Agreement Ratings)
	Democrat-leaning Item Set	
1 - "The US has loose gun laws"		

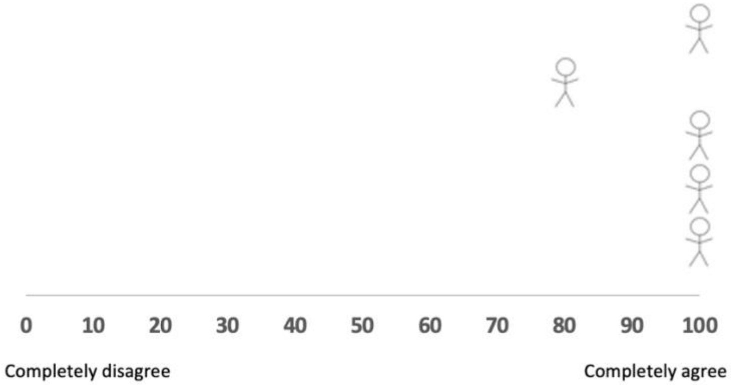
2 -
"The US government spends little for climate related research"



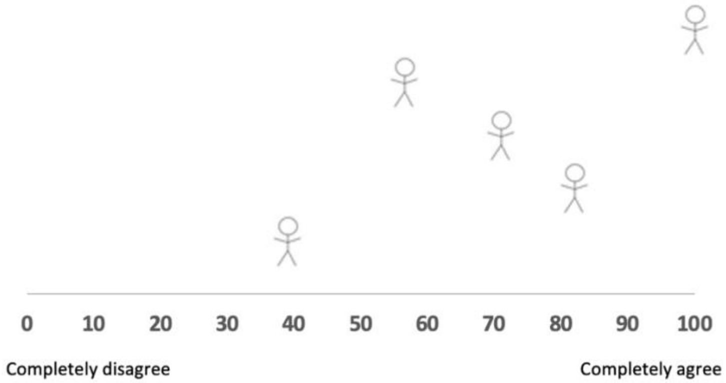


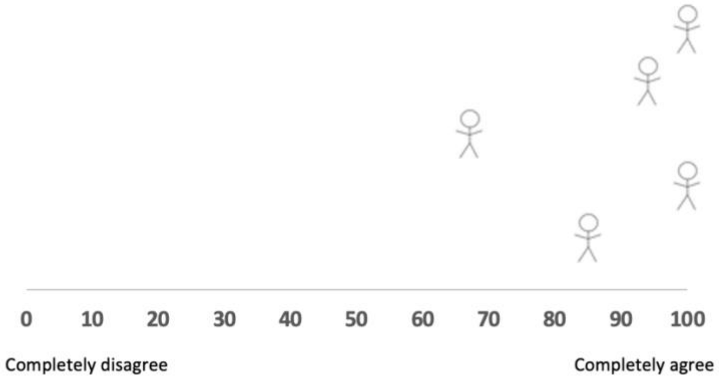
3 -
"Obamacare has successfully decreased the number of uninsured Americans"



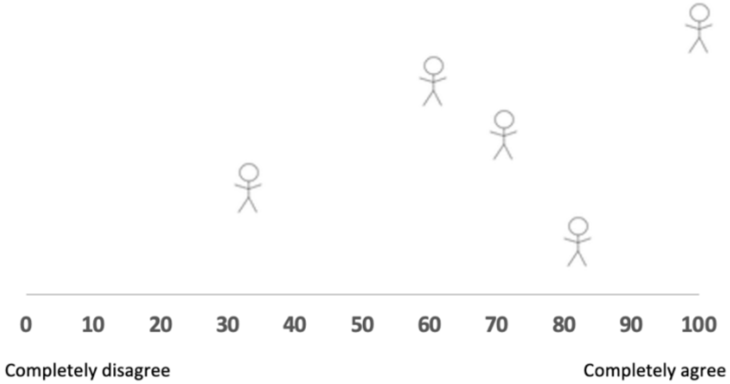


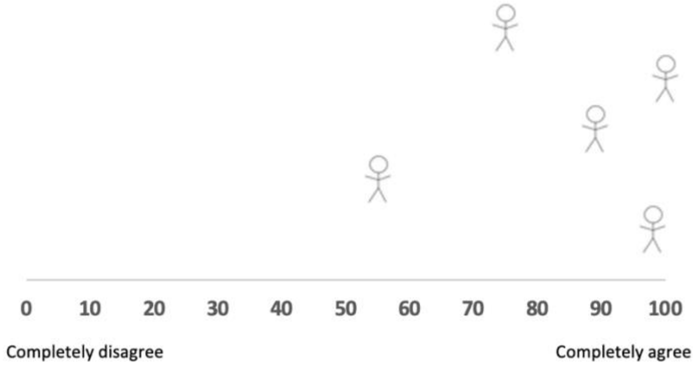
4 -
"Embryonic stem cell therapy is a successful modern treatment method"



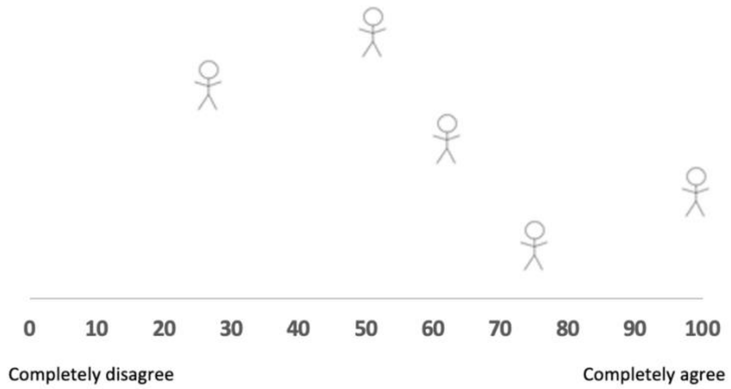


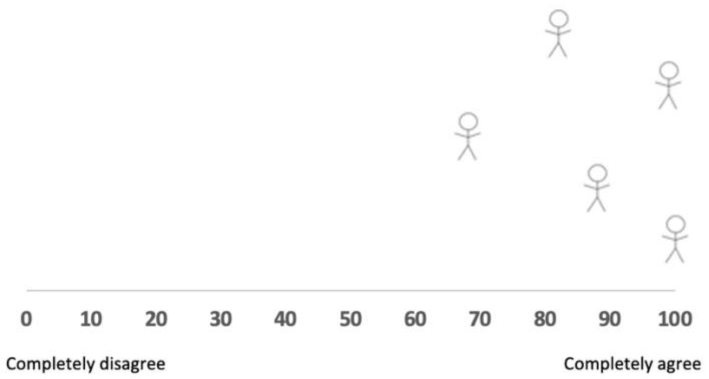
5 -
"Colleges and Universities are having a positive effect on young generations' futures"



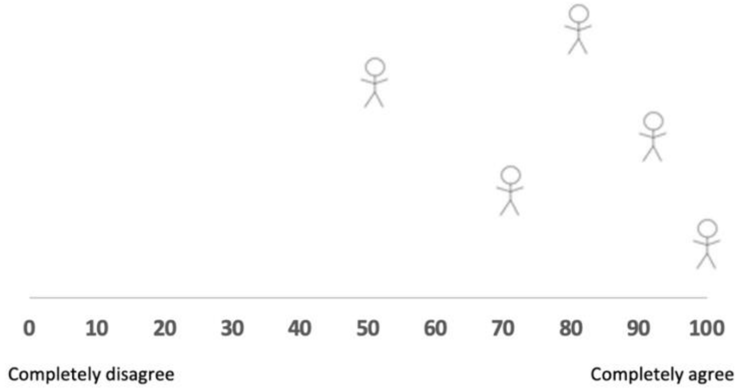


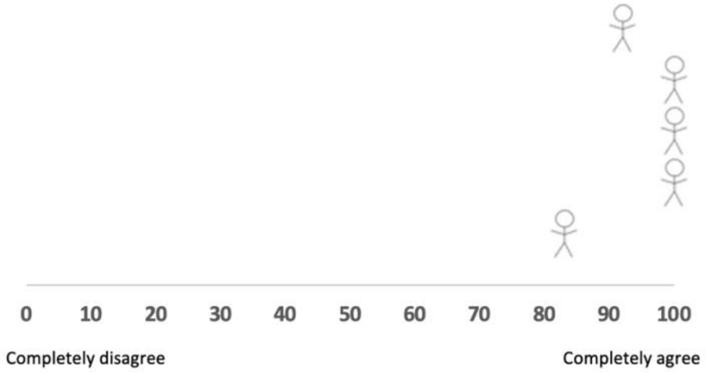
6 -
"The Affordable Care Act saved the US a huge amount of money"



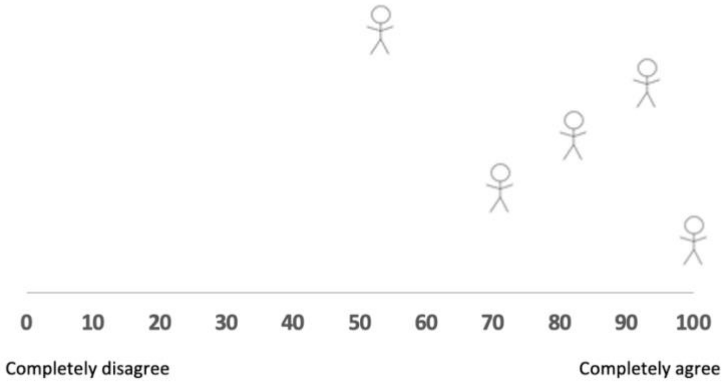


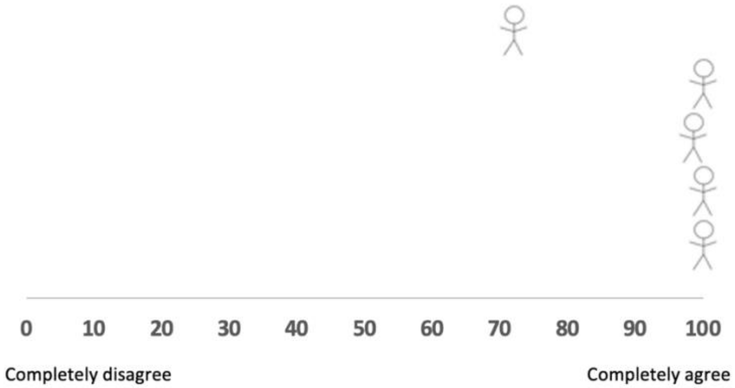
7 -
"All cities in the US experience more extremely hot days compared to 50 years ago"



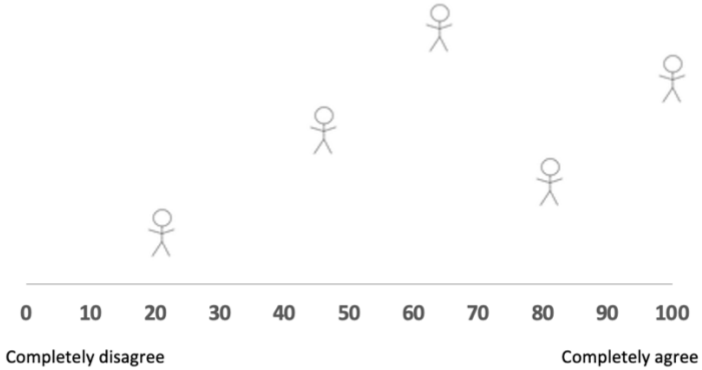


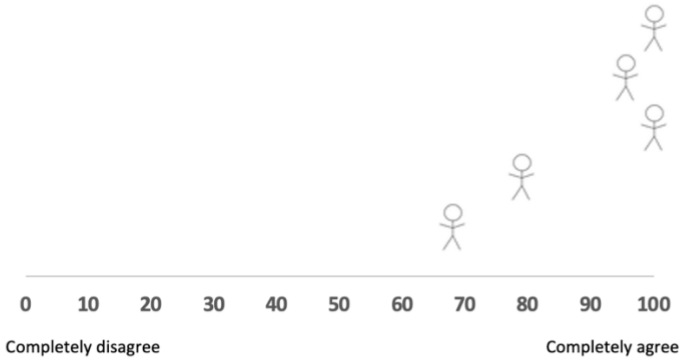
8 -
"Children in the US are at high risk of witnessing gun violence"



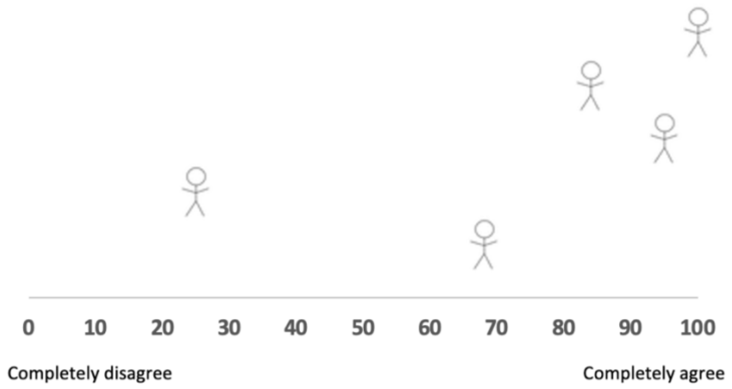


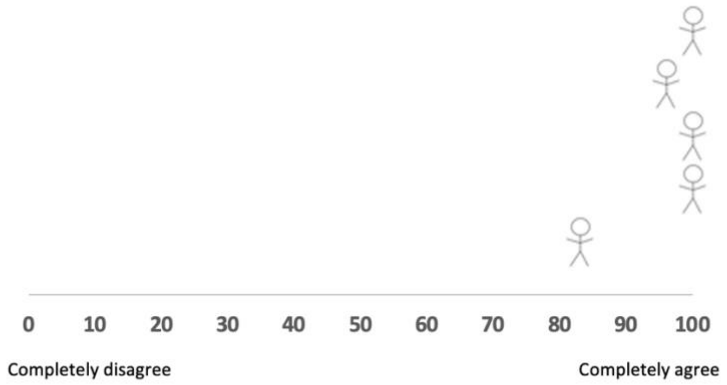
9 -
"The US allocates too much of the spending budget to Defense and Military"



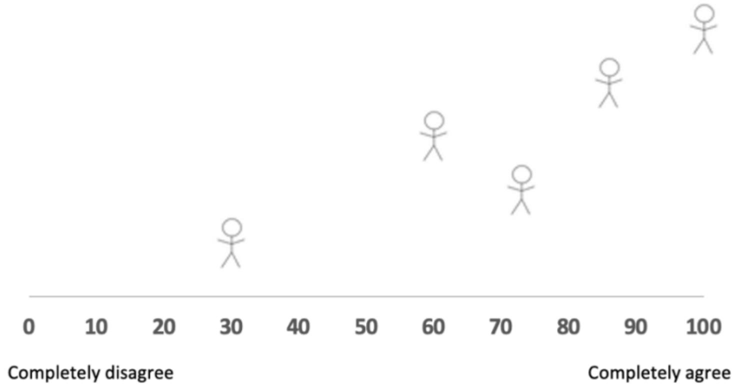


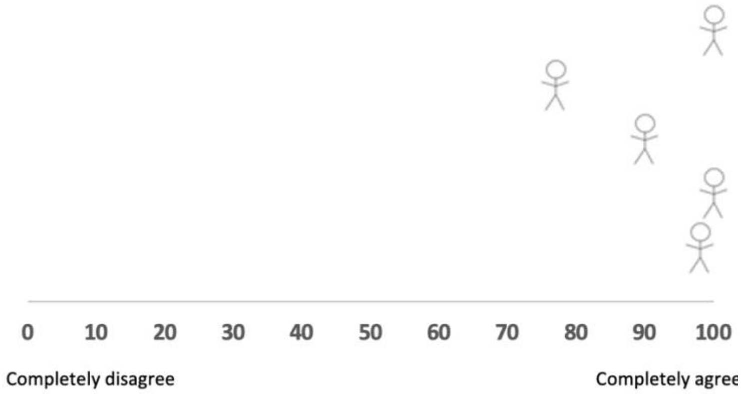
10 -
"Children raised by same-sex parents are just as adjusted as children raised by opposite-sex parents"



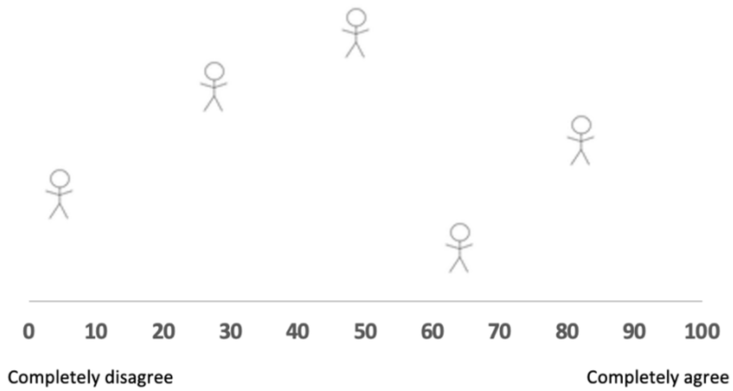


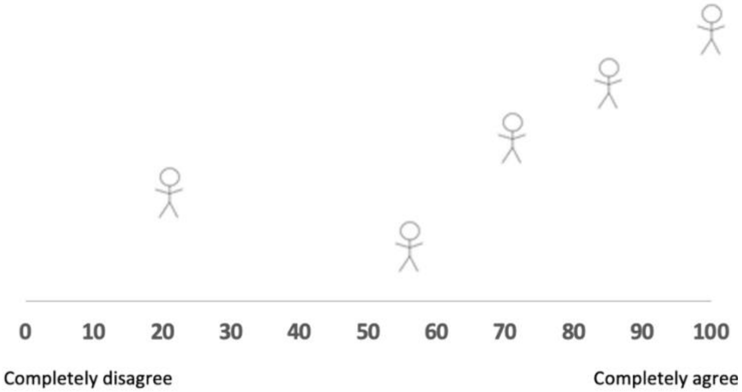
11 -
"The amount of government assistance to poor families in the US is not high enough"





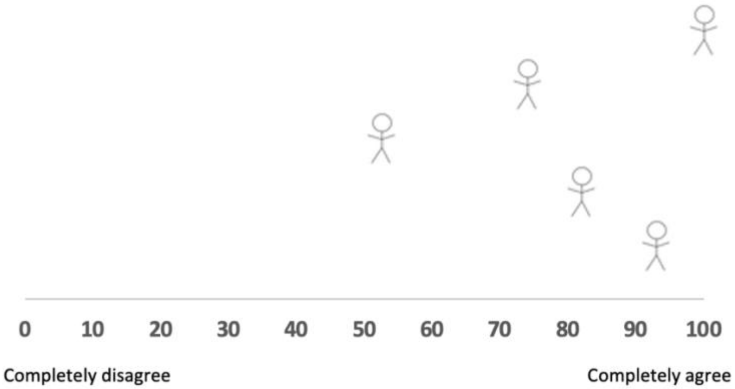
12 -
"Immigrant households in the US rarely access welfare programs"

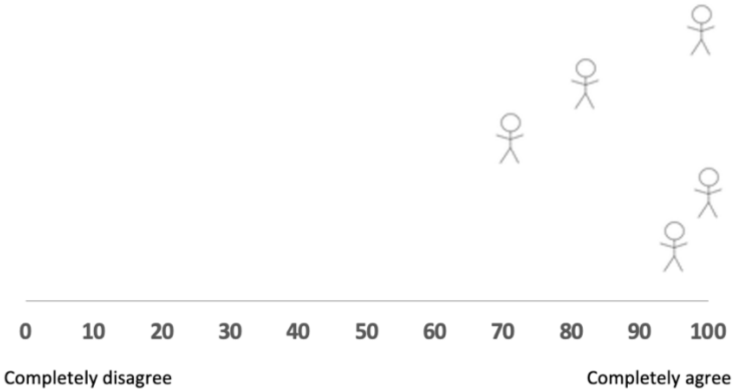




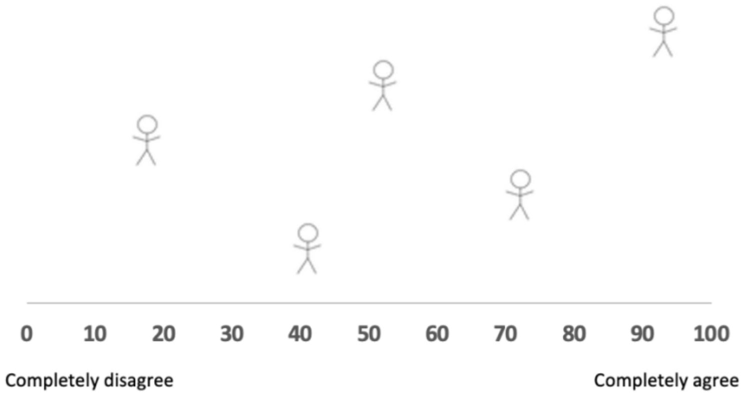
Republican-leaning Item Set

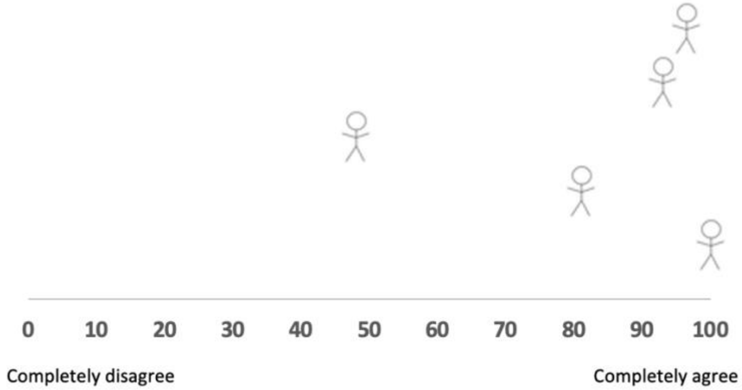
1 -
"The US is at great risk of illegal drug activity"



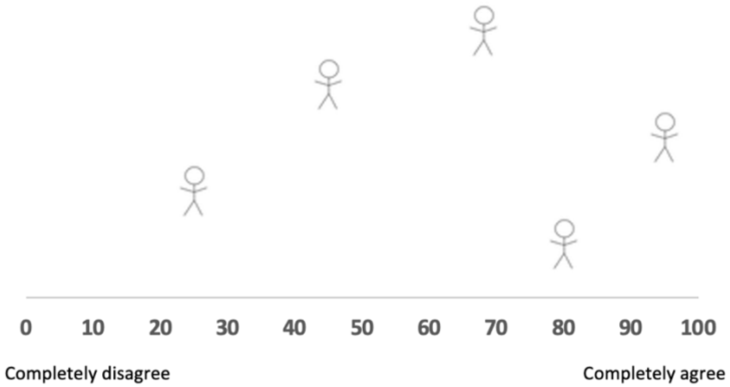


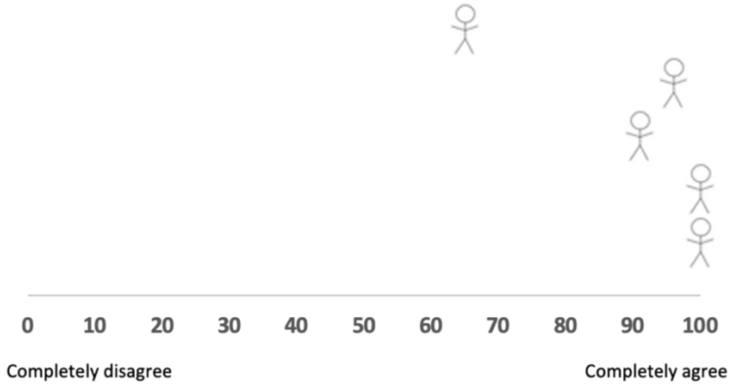
2 -
"African American women get more abortions than Caucasian women"



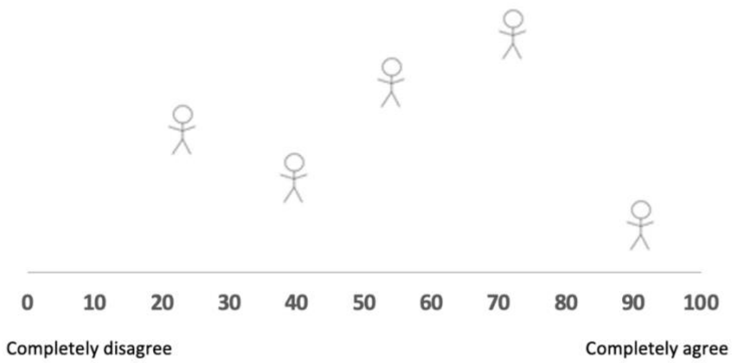


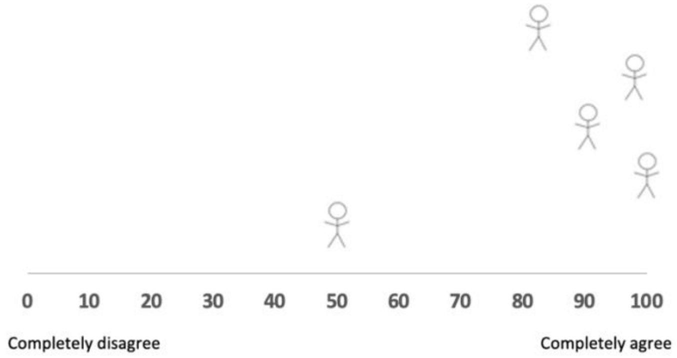
3 -
"Police use of force in the US is not causing that many deaths"



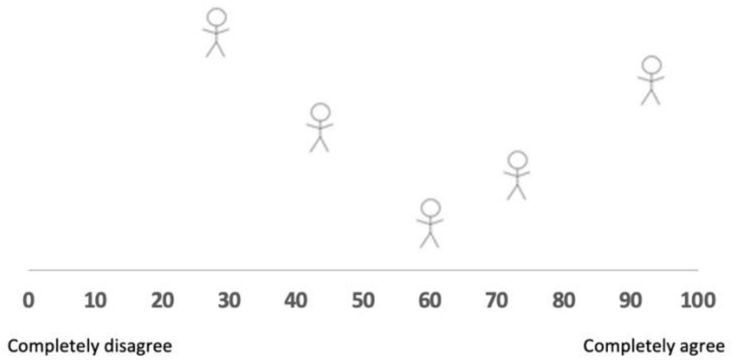


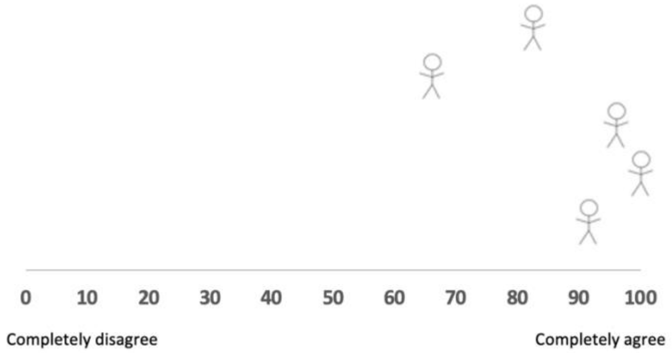
4 -
"A
large proportion of
immigrants in the
US are not in the
workforce"



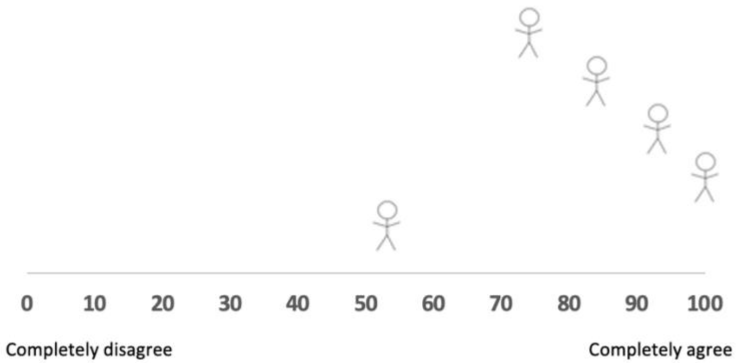


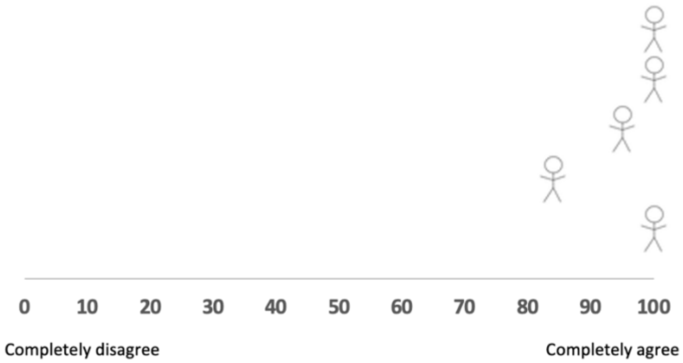
5 -
"The amount of
US corporate
income taxes paid
yearly is high"



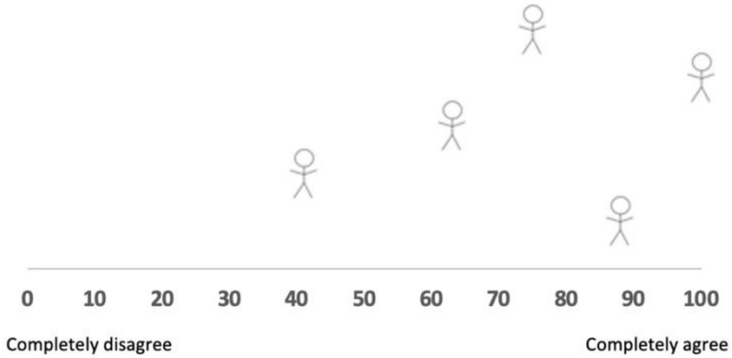


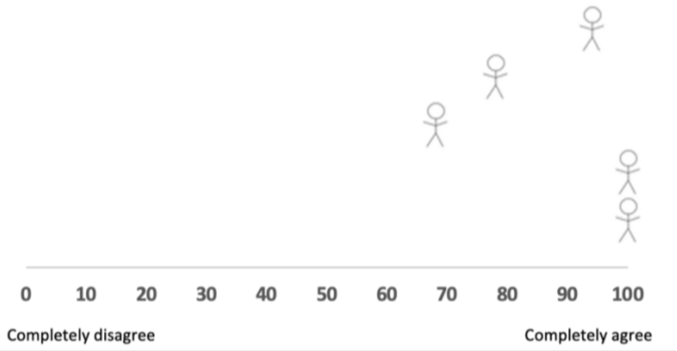
6 -
"A large number of undocumented workers are working illegally in the US"



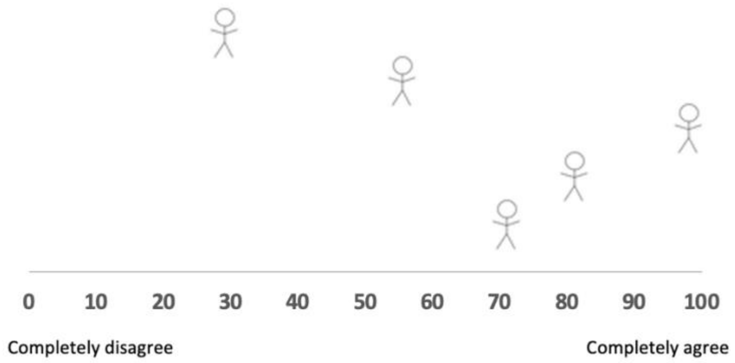


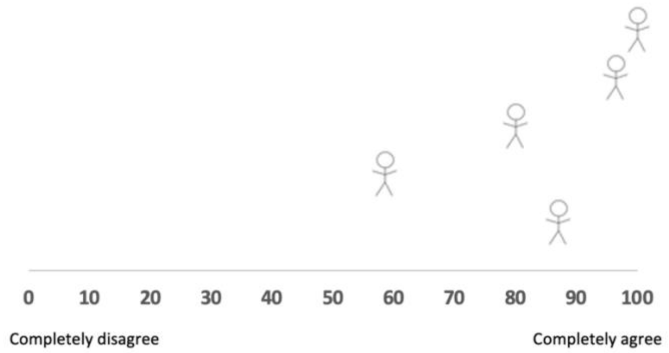
7 -
"Currently, foreign-born terrorists are a big threat to Americans in the US"



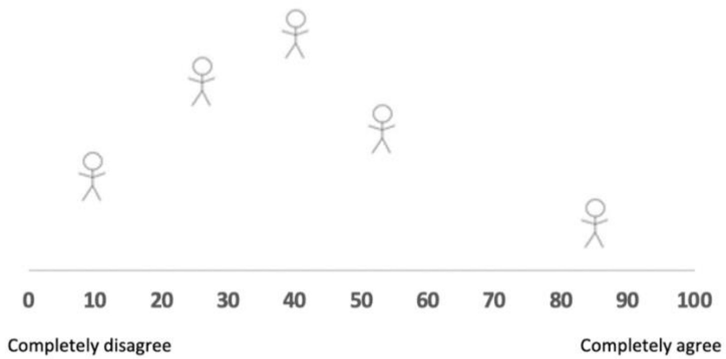


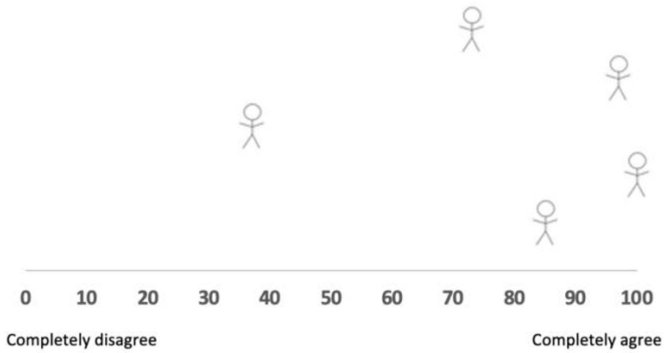
8 -
"A large percentage of abortions in the US are paid for with public funds"



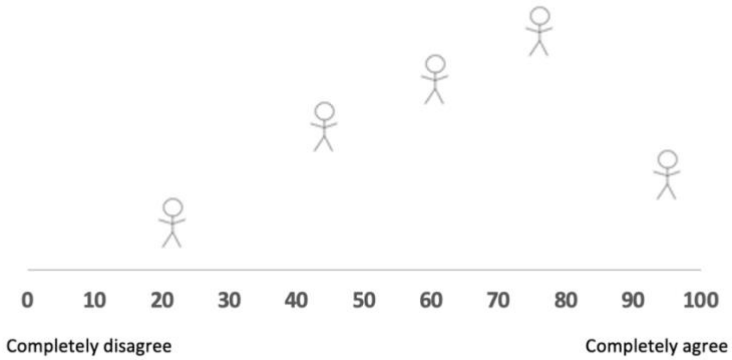


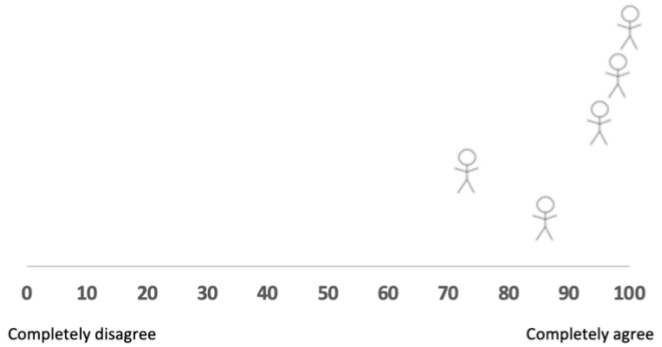
9 -
"In the US, men and women are, on average, paid equally for the same job"



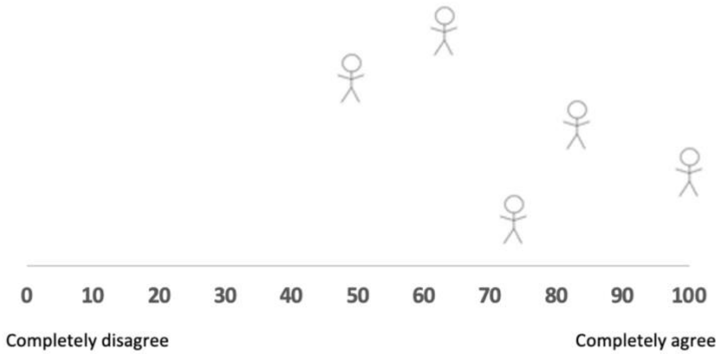


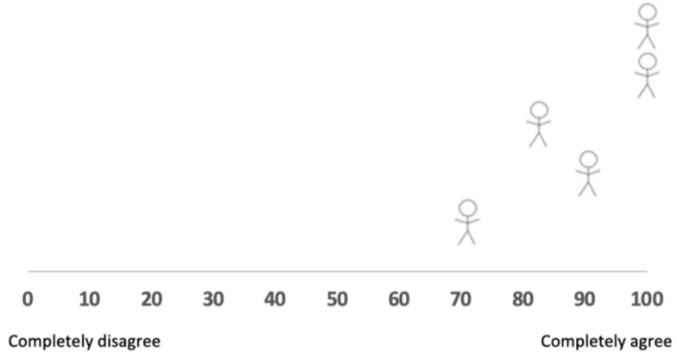
10 -
"The US justice system is fair to racial minorities"



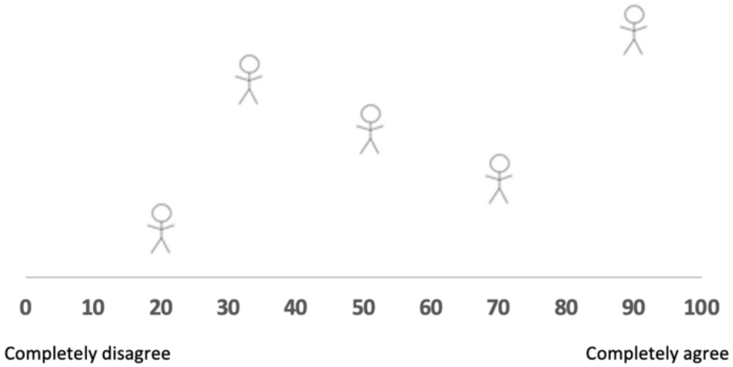


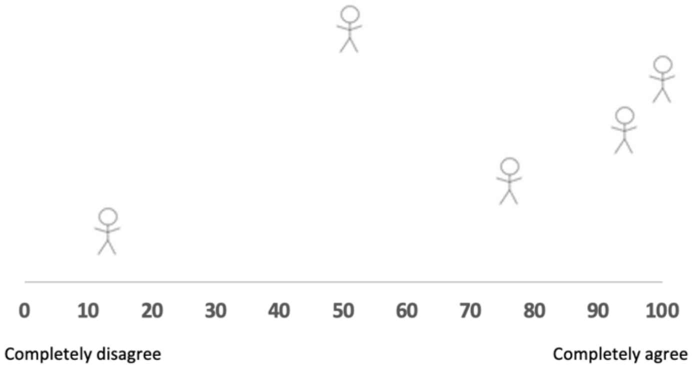
11 -
"Government regulations have large costs for the US economy"





12 -
"Small businesses owned by immigrants in the US do not provide that many jobs"





Note. Individual samples were randomly presented in the same or opposite order as the figures shown, to account for order effects. Data used to compute these samples were sourced from Vlasceanu et al. (2021).