Supplementary Information for

The Wisdom of Partisan Crowds

Joshua Becker, Ethan Porter, and Damon Centola

Correspondence to Joshua Becker
Email: joshua.aaron.becker@gmail.com

This PDF file includes:

Supplementary text
Figs. S1 through S4

Other supplementary materials for this manuscript include the following:

Supplementary Datasets
R Code for Statistical Replication

Contents

1. Supplementary Methods

1.1. Pre-Registration and Data Collection

The research design and hypotheses for Experiment 1 were pre-registered with the Open Science Framework on November 9th, 2017 and can be accessed at: https://osf.io/fk7rz/wiki/home/ . Data collection for Experiment 1 ran from November 10th to November 12th, 2017. Data collection for Experiment 2 ran from January 8th to January 24th, 2019.

All tests presented in this paper were pre-registered, except the test comparing average pairwise difference as a supplementary method for measuring the robustness of our findings on polarization. All statistical tests use two-tailed Wilcoxon signed-rank test (for paired data or single vector tests) or rank-sum test (for unpaired data) except where otherwise stated. R code for this analysis is provided in the supplementary materials.

1.2. Replication Materials

Datasets for pilot questions, Experiment 1, and Experiment 2 as well as R code for data cleaning, analysis, and figure generation can be found at the following two web locations:

- https://github.com/joshua-a-becker/wisdom-of-partisan-crowds
- https://dataverse.harvard.edu/dataverse/joshuabecker

1.3. Measuring Change in Polarization

To measure polarization as the distance in the mean belief of Republicans and the mean belief of Democrats, we first paired each trial according to the trial number, i.e. the chronological order in which trials were conducted, according to our pre-registered analysis plan. We then measured the distance in the mean at each round as the absolute value of the arithmetic difference between the two means. We calculated this distance separately for each of the 4 questions. For each question, we then calculated the change in this distance from Round 1 to Round 3. We then averaged this value across all four questions to measure the change in polarization for each trial.

To measure polarization as average pairwise distance, we first calculated the distance (absolute value of the arithmetic difference) between every Republican and every Democrat in the two paired trials for a single question. We then calculated the mean of this value to produce the average pairwise difference for a single question for a single pair of trials. We then calculated the change in this value from Round 1 to Round 3. We then averaged this change across all four questions to measure the change in polarization for each of the 12 trials.

1.4. Question Selection

Questions for Experiment 1 were chosen from a set of 25 questions that were pre-tested to identify which questions showed the greatest partisan bias. These 25 questions were chosen to be difficult to answer quickly via web search. The necessity to avoid easily accessible answers precluded the use of many questions which appeared previously in literature on partisan bias. All questions showed a substantial variance in answers, suggesting that subjects did not find the true value in

	Experiment 1					Experiment 2			
	Mean		Median			Mean		Median	
	Repub.	Dem.	Repub.	Dem.		Repub.	Dem.	Repub.	Dem.
California	45.2	42.5	45	43	Immigration	28.5	6.09	14	5
Election	223	205	235	210	Military	27.2	33.2	22	30
Taxes	22.7	21	22	20	Soldiers	20614	19802	3000	4000
Unemployment	9.06	7.79	6	5	Unemployment	6.72	1.34	3.2	-2

Table S1. Non-normalized mean and median of initial beliefs.

the time frame allotted. Questions were selected via manual inspection of the F-statistic comparing responses by both parties as well as the difference in mean, with a subjective assessment to determine which were most likely to show partisan bias in our full experimental sample. The four chosen questions were indicated in the pre-registration prior to data collection. Questions for Experiment 2 were chosen subjectively based on previous literature and their connection to policy controversies.

1.5. Question Order

We used four question sets, which varied only by the order in which questions were presented. The four sets were as follows: (1,2,3,4), (2,1,4,3), (3,4,1,2), (4,3,2,1) where each number indicates a question, and the sets indicate the order in which they were presented. This process ensured that each question occupied each position once. We followed this pre-registered procedure for both Experiment 1 and Experiment 2.

2. Supplementary Analysis

2.1. Initial Partisan Bias

We test the initial difference between Republican and Democrat beliefs using a non-parametric Wilcoxon rank-sum test, which tests a difference in "location" that is comparable to a difference in median. While our Figures report the mean belief at each round of the experiment, they show normalized beliefs as measured in our statistical tests. To demonstrate more directly the initial difference in belief between Republicans and Democrats, Table S1 shows the non-normalized mean and median response prior to social influence. This table shows mean and median beliefs with extreme response (e.g., due to typographical error) removed, as described both in the main text and in section 2.6 and 2.7 below.

2.2. Estimation Bias and Partisan Bias

Despite the prevalance of bias due to motivated reasononing, we find that Democrats and Republicans share an overall tendency to under- or overestimate, meaning that the bias of the average belief tends to be in the same direction for both parties. This can be explained as a motivated partisan bias added on top of a general estimation bias. The supplementary R code includes data for our 25 pilot questions and a script to measure the shared estimation bias. The

supplemetary R code also includes a script to download data provided by Bullock et. al (21) and measure the shared estimation bias.

2.3. Change in Error and Polarization for Control Groups

In Experiment 1, polarization increased for the control groups even as error decreased. This result can be explained by the observation that the magnitude of change in the mean belief for Republicans and Democrats differs for individual questions, allowing the difference between the two groups to increase. However, we note that neither the change in error of the mean nor the change in polarization of the mean is statistically significant for the control groups. The control group data was collected solely for comparison against the social groups, and does not provide a sufficient sample size for inference on the dynamics of revision by isolated individuals.

2.4. Analysis of the Change in Error of the Median

We found no significant effect of social influence on the change in the error of the median. We emphasize that our theoretical model describes the movement of the mean (30) with no prediction for the median, and so this analysis is strictly exploratory.

In Experiment 1, we found that social influence significantly reduced the error of the median belief ($P<0.001$), but this difference was not significantly different from control groups ($P>0.25$). Overall, the error of the median decreased in 83% of social trials and 63% of control trials. In Experiment 2, social influence did reduce the error of the median, but the effects were not significant ($P>0.41$) and the social and control group were not significantly different ($P>0.30$). The error for the median decreased in 58% of social trials and decreased in 88% of control trials.

Our results can be explained by the observation that the network dynamics of social influence determine mean beliefs (30). Thus as the belief distribution collapses toward the mean, the median will be drawn toward the mean as well. If by chance the median is initially more accurate than the mean, then social influence will decrease the accuracy of the median, despite the fact that the mean belief and the mean individual both become more accurate. Therefore, the effect of social influence on the median depends on the chance relationship between the median and the mean.

2.5. Relationship Between Speed of Response, Round, and Accuracy

The following analysis is based on time data collected for Experiment 2.

For both social and control conditions, the mean time to answer drops over 50% from Round 1 (μ=26 seconds) to Round 3 (μ=14 seconds).

Interestingly, time to answer shows no systematic relationship to accuracy with a negligible positive correlation in Round 1 (ρ=0.02, P>0.28) and a negligible negative correlation in Round 2 (ρ =-0.004, P>0.75) and Round 3 (ρ=-0.02, P>0.19).

2.6. Experiment 1: Alternative Normalization Procedure

To test the robustness of our results, we repeated the analysis in the main text with an alternative normalization procedure. In the main text, we logarithm-transformed all responses and true values prior to analysis to control for extreme answers (e.g., due to entry error) and normalize responses across questions with true values of very different magnitude. Here, we present the results of the analysis in which we control for outliers by removing responses greater than one order of magnitude above the true value. This process resulted in the removal 42 out of 4,024 provided estimates. We then normalize responses by dividing by the standard deviation of all remaining responses.

As shown in the following results, we find that this alternative normalization procedure provides qualitatively identical outcomes to the analysis presented in the main text. The following section indicates the results of the same statistical tests as presented in the main text, but using this alternative normalization procedure. R code for this analysis is provided as a supplementary file.

Baseline Differences in Belief. Responses at Round 1 are significantly different between Republicans and Democrats ($P<0.001$ for all questions except race in California, for which $P<0.05$).

Change in Error of Mean in Social Condition. Error at Round 3 was significantly lower than error at Round 1 for both the 12 Republican trials ($P<0.03$) as well as the 12 Democrat trials ($P<0.001$).

Difference in Change in Error of Mean, Control and Social Conditions. Error in the control condition decreased slightly, but the change was not significant ($P>0.94$) and the change in the social condition was significantly greater ($P<0.01$).

Change in Standard Deviation of Responses. Standard deviation decreased significantly in the social condition ($P<0.001$) but increased slightly in the control condition ($P<0.55$) and the two conditions were significantly different ($P<0.001$).

Change in Average Individual Error. Average individual error decreased in the social condition ($P<0.001$) but increased slightly in the control condition ($P<0.39$) and the two conditions were significantly different ($P<0.001$).

Change in Polarization – Difference of Mean. The mean belief of Republicans and Democrats became more similar for social conditions, though the result is statistically weaker than that presented in the main text ($P<0.39$), but the mean belief became less similar in control conditions ($P<0.41$). Crucially, the two conditions are significantly different ($P<0.01$).

Change in Polarization – Average pairwise Difference. The average belief of a randomly selected Republican and Democrat became more similar in social conditions ($P<0.001$) but became less similar in control conditions ($P=0.25$). The two conditions are significantly different ($P<0.01$).

2.7. Experiment 2: Analysis Including Extreme Values

The analysis presented in the main text normalizes responses by dividing them by the true value, such that responses represent percent of the true value and errors represent percent error. However, this analysis is sensitive to extreme values, which subjects occasionally enter for example due to entry error. We therefore presented our main results with the following extreme values removed:

Military spending (percent): answers outside the range 0 to 100.
Soldier deaths (number above 0): answers about 1 million and answers below 0
Immigration (percent change): answers above 1000 or below -1000
Unemployment (percent change): answers above 1000 (no answers fell below -1000)

This process removed fewer than 1% of all responses.

To ensure that our results were not determined by this data cleaning process, we replicated our analysis including all responses, which we describe below. We do not report these results in the main text because the magnitude of these results are not plausible: for example, the mean of the error of the mean in the social condition was 42,500% at Round 1 (due to absurdly large values generating apparently large errors in the average) and decreased to 63% in Round 3 (presumably due to simple typographic error corrrection) and was accompanied by a decrease in standard deviation of 250,000%. However, as we report below, the general characteristics of the results were the same.

Baseline Differences in Belief. Responses at Round 1 are significantly different between Republicans and Democrats ($P<0.001$ for all questions).

Change in Error of Mean in Social Condition. Error at Round 3 was significantly lower than error at Round 1 for both the 12 Republican trials ($P<0.001$) as well as the 12 Democrat trials ($P<0.001$).

Difference in Change in Error of Mean, Control and Social Conditions. Error in the control condition increased, but the change was not significant ($P>0.74$) and the change in the social condition was significantly greater ($P<0.02$).

Change in Standard Deviation of Responses. Standard deviation decreased significantly in the social condition ($P<0.001$) but increased in the control condition ($P>0.74$) and the two conditions were significantly different ($P<0.02$).

Change in Average Individual Error. Average individual error decreased in the social condition ($P<0.001$) but increased in the control condition, though the change is not statistically significant ($P>0.94$) and the two conditions were significantly different ($P<0.02$).

Change in Polarization – Difference of Mean. The mean belief of Republicans and Democrats became more similar for social conditions, and the result is statistically stroner than that presented in the main text ($P<0.001$), and the mean beliefs became more similar in control conditions but the result is not statistically significant ($P>0.63$). Crucially, the social and control conditions are significantly different ($P<0.02$).

Change in Polarization – Average pairwise Difference. The average belief of a randomly selected Republican and Democrat became more similar in social conditions ($P<0.001$) but became less similar in control conditions but the result is not statistically significant ($P>0.63$). The two conditions are significantly different ($P<0.02$).

3. Simulation Study

3.1. Initial Hypothesis Generation

We simulate the DeGroot model (30) with the additional assumption that people fall into two types. Here, type A has a positive bias (toward overestimation), and type B has a negative bias (toward underestimation). We assume that the same mechanism which generates bias (e.g., political partisanship) also moderates response to social influence. Thus, the response to social influence decreases as bias increases. For type A (B), this means people with extreme overestimates (underestimates) pay relatively less attention to social information.

The model follows that described by Becker, Brackbill, and Centola (15). To generate a noisy correlation between belief and response to social influence, we determine the response to social influence as

$$\alpha=S(R*B+N)$$

where α indicates the amount of weight placed on an agent's own belief relative to social information; S indicates the sigmoid function; R indicates belief at Round 1, and is drawn from a normal distribution with mean 0 and variance 1; B indicates partisan bias, and is either +1 or -1 for each of the two classes; and N generates noise, and is a random normal variable with mean zero and variance ε. Noise is either low ($\varepsilon=1$) or high ($\varepsilon=5$) with comparable results for no noise ($\varepsilon=0$). Because the expected error of belief (R) is zero, α is proportional to belief extremity, and this function means that agents with more extreme beliefs place more weight on their own belief—i.e., make less use of social information.

R code for this simulation is provided as a supplementary file. Results from this simulation are shown in Figure S4.

3.2. Potential Failures of the Wisdom of Parisan Crowds

To identify conditions under which the wisdom of crowds might fail, we conduct empirically calibrated computational experiments in which an individal's centrality in their social network is correlated to their belief extremity. We first measure the direction of partisan bias for each political party for each of the 8 questions used in the experiment described in the main text. For example, since Republicans provided higher average estimates of for the change in immigration than Democrats, Republicans with high value beliefs are considered "more extreme" than Republicans with low value beliefs; the opposite is true for Democrats, for whom individuals with low value beliefs are considered more extreme than individuals with high value beliefs.

We use these empirically calibrated parameters to set initial conditions for the same simulation described above in section 3.1, varying the correlation between centrality and belief extremity. Because the simulation is sensitive to extreme (and unlikely) response values (due e.g. to entry error), we use only the inner 99% of the estimations (i.e., exclude the 0.5% highest and 0.5% lowest responses). To generate the sample population for each run of the simulation, we sample 100 individuals (with replacement) from the empirical response distribution. We then measure their extremity rank, assigning each simulated individual a value from 1 to 100 based on their belief. To assign individuals positions in a social network, we generate a vector of values from 1 to 100, experimentally varying the correlation between this new vector and the extremity ranking vector. We then assign individuals a location in the social network corresponding to the node with that rank of centrality.

Centralized networks are generated according to the barabasi.game algorithm in the igraph package of the R statistical computing platform, with the power parameter set to 2 and m set to 2, generating highly centralized networks with an average of 4 connections per node.

R code for this simulation is provided as a supplementary file. Results from this simulation are shown in Figure S5.

4. Supplementary Figures

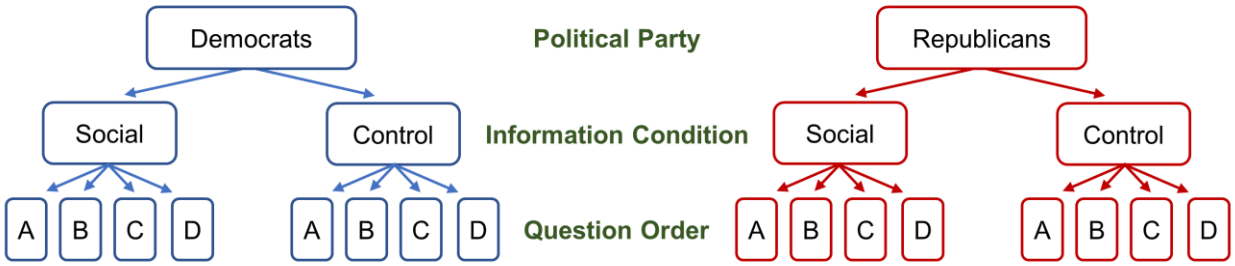

Figure S1. Schematic indicating randomization procedure. Randomization was conducted separately for Republicans and Democrats. Subjects were randomized to either the social or a control condition group. Within each group, subjects were randomized to one of four question orders.

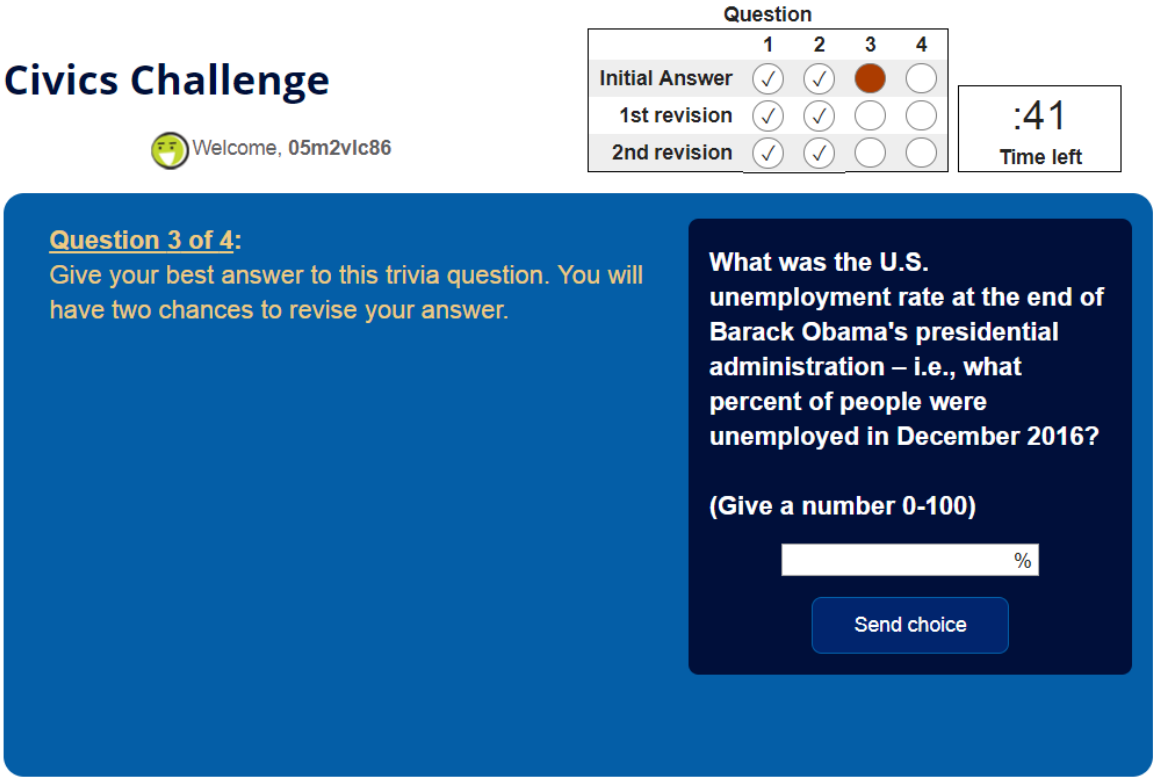

Figure S2. A screenshot of the experimental interface for Experiment 1.

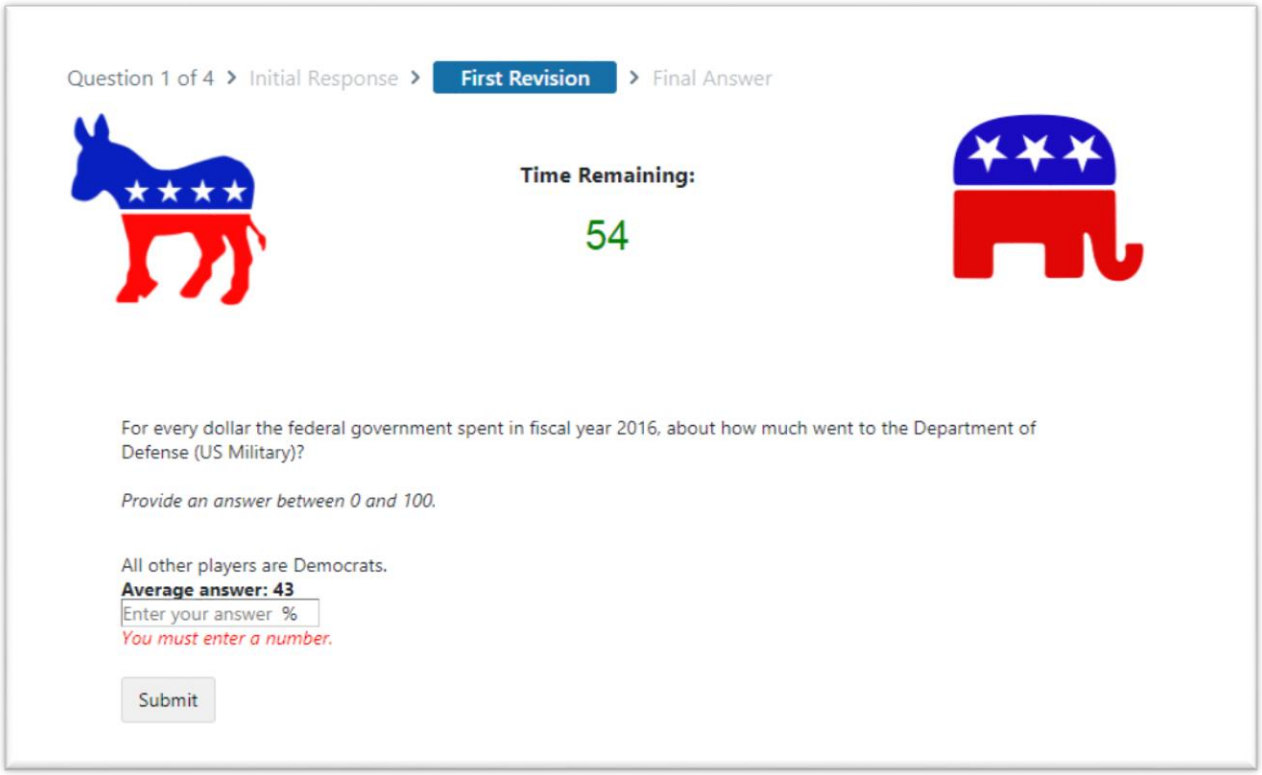

Figure S3. A screenshot of the experimental interface for Experiment 2. In addition to the visual primes shown above, the experiment was called the "Politics Challenge," and subjects were required to confirm their party identification immediately prior to the start of the experiment.

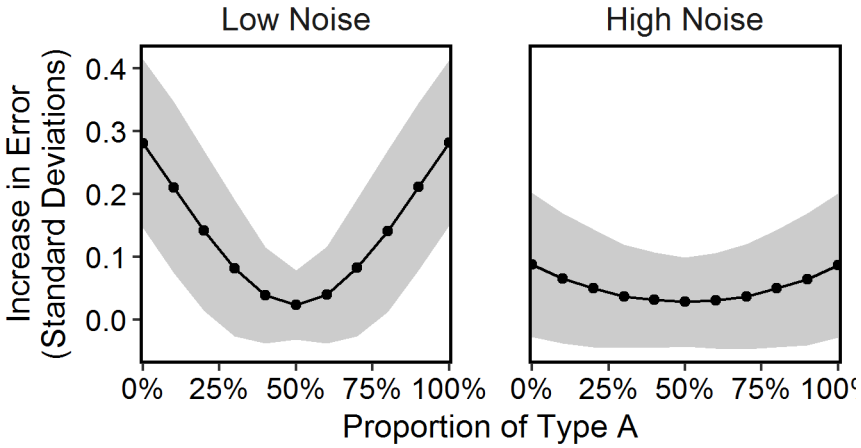

Figure S4. Simulating law of group polarization. Each point represents the mean change in the absolute value of the error of the mean belief for 10,000 simulated replications of a theoretical model which assumes that partisan bias correlates with response to social influence. Grey area represents standard deviation. Under this assumption, the effect of social influence on the average belief in a group depends on the proportion of members from each of the two partisan groups. When groups are composed of equal numbers of members from each group (Proportion of type A=50%) the bias of group A cancels out the bias of group B, and the only change in error is due to random fluctuation. However, when one group is more prevalent then the bias of the larger group will determine overall group beliefs, and the error will increase. This theoretical model indicates that the experimental design most likely to detect the effect of partisan bias on group beliefs is one in which subjects are arranged into homogeneous social networks.

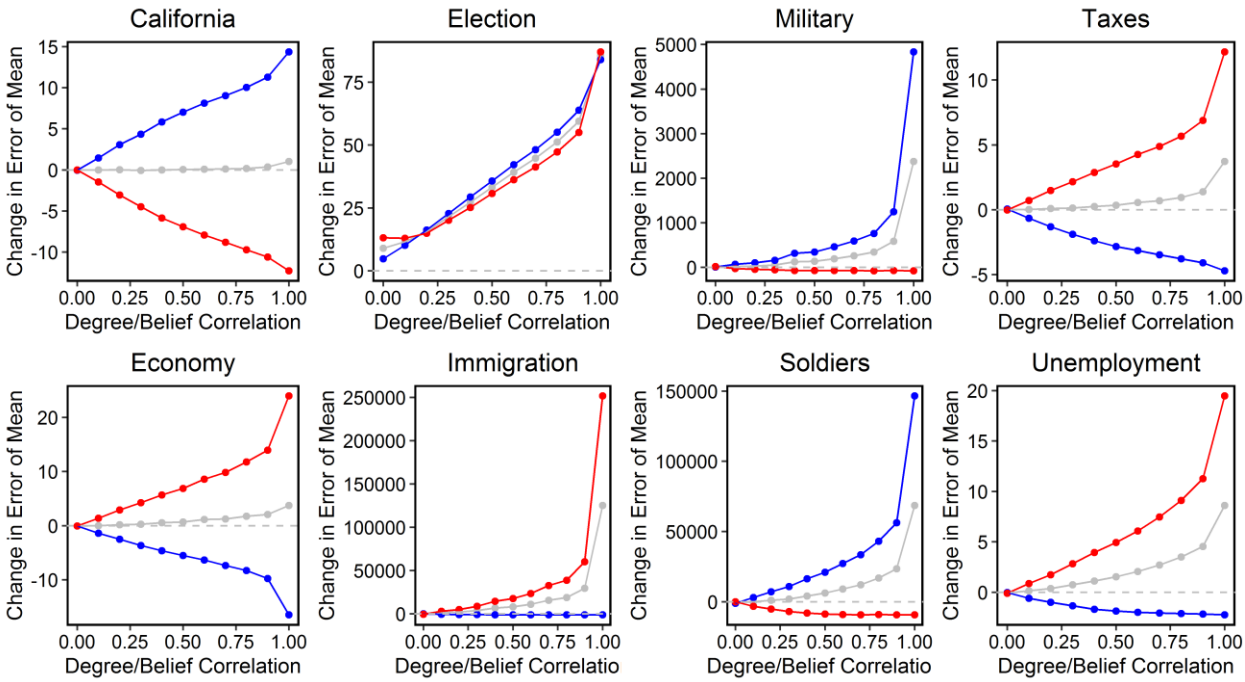

Figure S5. Simulating effect of centrality/extremity correlation. Each point represents the mean change in the error of mean for 5,000 simulations. The x-axis indicates the correlation between an individual's belief partisanship (see text in section 3.2 above) and network centrality. Each panel shows outcomes for simulations applies to a different empirical belief distribution. The first row indicates simulated outcomes for data from Experiment 1, and the second row indicates simulated outcomes for data from Experiment 2.