Combining survey questions with a Bayesian bootstrap method yields accurate election forecasts

Henrik Olsson[1,a], Wändi Bruine de Bruin[b], Mirta Galesic[a], & Drazen Prelec[c]

[a]Santa Fe Institute, Santa Fe, NM, 87501; [b]Sol Price School of Public Policy, Dornsife Department of Psychology, Schaeffer Center for Health Policy and Economics, and Center for Economic and Social Research, University of Southern California, Los Angeles, CA 90089; [d]Sloan School of Management, Department of Economics, and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02142.

Abstract

We present a new Bayesian bootstrap method for election forecasts that combines traditional polling questions about people's own intentions with their expectations about how others will vote. It treats each participant's election winner expectation as an optimal Bayesian forecast given private and background evidence available to that individual. It then infers the independent evidence and aggregates it across participants. The bootstrap forecast outperforms forecasts based on own intentions questions posed on large national samples before the 2018 and 2020 U.S. elections. The bootstrap forecast puts most weight on people's expectations about how their social contacts will vote, which might incorporate information about voters who are difficult to reach or who hide their true intentions. Beyond election polling, the new method is expected to improve the validity of other social science surveys.

[1] Corresponding author: Henrik Olsson, Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501. Email: olsson@santafe.edu

In the 2020 U.S. presidential election, most polls overestimated Biden's advantage in national vote shares, on average predicting that his lead over Trump will be around 8 percentage points on the day of the election (*1*) when the actual result was 4.5 percentage points. Indeed, the bulk of the coverage of election polling after the 2020 election asked where polling went wrong (*2–5*), with few answers provided (*6*). How can such a large average polling error be explained and how can we improve poll-based forecasts to avoid such errors in the future?

To counteract problems with traditional polls that ask about participants' *own voting intentions*, researchers have been exploring *wisdom-of-crowds polling questions* (*7–10*). These questions ask respondents to act either as lay forecasters, predicting the *election winner*, or as lay ethnographic informants, estimating the voting expectations of individuals in their *social circle*. Previous research has shown that wisdom-of-crowds methods can produce more accurate forecasts than own intentions (*8*). Most notably, social-circle questions have outperformed own intentions questions in five national elections in four countries: the 2017 French presidential election, the 2017 Dutch parliamentary election, the 2018 Swedish parliamentary election and the 2018 U.S. election for House of Representatives (*7, 11*). Although combining forecasts within and between methods has a long history in the forecasting literature (*12–16*), so far there have been no attempts to combine information from all of these wisdom-of-crowds methods with information from traditional questions about own intentions at the individual participant level in a theoretically justified framework.

Here we provide the first method for integrating information from own intentions, election-winner, and social-circle questions into a single, theoretically justified forecast. This new Bayesian bootstrap method presents a Bayesian justification of how to best weight and combine information present in the answers to traditional and wisdom-of-crowds polling questions to forecast elections. It estimates the private and background evidence present in the answers to those questions into a forecast that weights their contributions accordingly. We then assess the accuracy of our forecasts from this new method, as well as the accuracy of own intentions and wisdom-of-crowds questions. We further investigate the relative contributions of different questions to the bootstrap forecast. Finally, we examine two reasons why some questions, in particular the social-circle expectations, contribute more to the Bayesian bootstrap forecast than others. It has been suggested that wisdom-of-crowds questions give more information about hard-to-reach or reluctant participants than traditional polling questions (*7*, *17*). But this conjecture has never been tested empirically. We show that social-circle expectations do indeed seem to capture less-well represented voters, such as younger and less educated, and hidden voters that are reluctant to reveal their voting preferences due to them being embarrassed, afraid of harassment, or intentionally obstructing their answers in polls. These results provide the first empirical support to the proposed reasons for the success of wisdom-of-crowds questions. All our analyses use data from large national samples (*N*>4000) surveyed before the 2018 U.S. Congressional election and the 2020 U.S. presidential election, as a part of USC Dornsife's Understanding America Study (UAS) (*18*).

The Bayesian Bootstrap

Previous research has shown that judgments of broader populations can be explained by models that assume that people base their judgments of broader populations on information from their social circles (*19*–*21*). It also suggested that people are quite good in estimating frequency of different characteristics in their social circles and that they are not just projecting their own characteristics (*19*, *22*). Based on these results, we will in the following assume that social-circle voting estimates are accurate. Social-circle knowledge, however, is not the only factor that can affect broader population judgments. That is particularly true for election forecasts, where people are constantly subject to information about the election from other sources. A complete informational accounting should therefore go beyond social-circle estimates to include the participants' own voting intentions and other public information or evidence that might be available.

The objective of the proposed bootstrap approach is to estimate and aggregate the totality of evidence, rather than aggregate any single type of polling question. To accomplish this, the forecast assumes that each participant's' election winner expectation, expressed in the form of vote shares, is itself an optimal Bayesian forecast in light of all evidence available to that participant. A participant's evidence may be conceptually divided into a private component that is reported directly, namely own voting intentions and social-circle expectations, and a background component, revealed indirectly via the election-winner expectations, which in our case is state-winner expectations. Background evidence partly derives from common sources, e.g., widely publicized polls, and partly from independent sources unique to each participant, e.g., personal social media and local gossip (see Figure S1).

Evidence is conceptualized as a sample of observations from an underlying Dirichlet distribution (a generalization of the beta distribution to more than two alternatives). The 'unit of account' is the participant's own voting intention, which contributes exactly one observation. A participant is then assumed to sample *N* individuals from their social circle, and combine this with a Dirichlet prior parameterized by a 'pseudocount' of *A*+*B* observations that represent background evidence. *A* denotes the common portion of background evidence, presumed the same for all participants, and *B* the independent portion. The term pseudocount is used in Bayesian statistics when the prior and posterior distributions are Dirichlet.

The results of this sampling determine the state-winner expectations at the level of an individual participant, treated in our model as in ideal Bayesian forecaster. For example, if a participant's sample of evidence is 8, 10, and 2 for Trump, Biden and Other candidates, this would imply state-winner expectations of 40%, 50% and 10% for the three options. The sample of 8 for Trump might come from 1 own vote for Trump, 3 votes from the social circle, and a pseudocount of 4 representing background (prior) evidence. Constructing the Bayesian bootstrap forecast involves one estimation step and one aggregation step.

<u>The estimation step</u>: To estimate parameters *N*, *A* and *B* we regress participants' state-winner expectations on their own intentions and social-circle expectations. The Supplementary Information provides detailed derivations, but the main idea can be outlined briefly.

Let x_i^r, y_i^r and s_i^r denote own voting intentions, state-winner expectations, and social-circle expectations for candidate *i* submitted by participant *r*. If the social-circle size and the amount of background evidence are the same across participants in a given state, then one can estimate these values by a constrained regression equation, $y_i^r = a_i + bx_i^r + cs_i^r + \varepsilon_i^r$, where ε_i^r is the residual (error) term. Because $\sum_i y_i^r = 1$, the parameters are constrained to $\sum_i a_i + b + c = 1$. The quantities, $a_i + \varepsilon_i^r$, bx_i^r and cs_i^r, are proportional to the amount of background evidence, own intentions evidence, and social-circle evidence in favor of candidate *i*, from the perspective of participant *r*. The ratio of regression parameters c/b is an estimate of the average social-circle size *N*. For example, if $c = b$, revealing that x_i^r and s_i^r have the same impact on state-winner expectations, that would suggest a social-circle size of only one individual. Similarly, the ratio a_i/b estimates the amount of background evidence favoring candidate i. If $b, c \approx 0$ and neither variable has impact on state-winner expectations, this would imply that participants are basing their election forecasts on background evidence alone, and in that case the best aggregate forecast is given simply by the intercepts a_i (which sum to 1 if $b = c = 0$ in the constrained regression). Note that a larger coefficient for social circles than own intentions can be viewed in other ways than a larger *N*, where the magnitude of the social-circle coefficient is larger than own intentions because the embodied evidence is "better" for some other reason. That might or might not be a problem for our method, depending on whether the social-circle sample is independent across individuals. Put another way, our method assumes that whatever evidence is presented via the social-circle question, that evidence is independent. Thus, sampling *N* individuals is one plausible scenario that would produce independent sampling of evidence, but it is not the only one.

Finally, the variance of residuals ε_i^r estimates the amount of independent relative to common background evidence. If participants with the same own and social circle answers also have roughly the same state-winner expectations, $\varepsilon_i^r \approx 0$, that implies that all background evidence is common.

The aggregation step: Using the parameter estimates *N*, *A*, *B* from the linear regression, the bootstrap forecast aggregates the evidence across all *n* participants. In the forecast, own intentions are weighted by sample size $+n$, social-circle expectations by $+nN$, and state-winner expectations by $A+nB$. The component *A* is not weighted by *n* to prevent double counting of shared evidence.

In the context of the U.S. elections, we calculate these forecasts for each U.S. state separately, and then combine them in the national-level forecasts by weighting them by the adult population in each state.

Results

Forecasting the 2018 U.S. House of Representatives election and the 2020 Presidential elections. Figure 1 shows the forecasting accuracy of own intentions, social-circle expectations, state-winner expectations, and the bootstrap forecast for these two elections on both the state and national levels. The forecasts for the 2020 election were preregistered (*23*). [2] A

[2] In this paper we focus on the new Bayesian bootstrap method, but the predictions for another method, the Bayesian Truth Serum (*42*), can be found in our preregistration for the 2020 election (*23*).

comparisons with other polls show that the results for our own intentions question are well within the range of most national polls that also use an own intentions question (Figure S4).

On the state level, in both 2018 and 2020 elections, the bootstrap forecast and wisdom-of-crowds questions are more accurate than own intentions questions in predicting vote shares for all parties (Figure 1A, top panel). The bootstrap forecast that uses regression parameters estimated in the 2018 election (i.e., no regression parameters were estimated from the 2020 data for this method) is also remarkably accurate. For the state-level margin between the two main parties/candidates, one of the wisdom-of-crowds questions — state-winner expectations — provides better forecasts than all other methods in both elections.

The results in Figure 1 also show that the bootstrap forecast typically outperforms all other methods in predicting national-level results for vote shares of all parties (Figure 1B, top panel), as well as for the Democrat-Republican margin (Figure 1B, bottom panel). In the 2018 election, the bootstrap forecast and the social-circle expectations have similar average errors across the three waves as the average of many polls from the website 538. In two out if three waves they have smaller errors than the average of polls (Table S1 in Supplementary information). In the 2020 election, the bootstrap forecast and social-circle expectations outperform the national-level forecasts of the average of many national polls across five waves before the election, and state-winner expectations outperform average polls for most of the waves (Table S2).

Contributions of different questions to the bootstrap forecast. The bootstrap forecast allows us to investigate the relative value of different questions for forecasting elections. For the 2018 election, averaging across states and polling waves, the bootstrap forecast assigns weights of approximately 11% to own intentions, 65% to social-circle expectations, and 24% to state-winner expectations. The results for the 2020 election are similar with weights of approximately 13% to own intentions, 59% to social-circle expectations, and 27% to state-winner expectations. These results suggest that the relative value of different questions is stable over time and contexts, even though the 2018 and 2020 elections were very different in terms of the political stakes, the overall societal situation and the form of the questions asked (i.e., in 2018 the questions asked about Democratic, Republican, or Other candidate, while in 2020 the Democratic and Republican presidential candidates were named). In particular, while covid-19 pandemics might have reduced people’s exposure to their social circles and hence decreased the usefulness of this information for election forecasts (*24*), the weight on social-circle expectations is only slightly smaller in 2020 compared to 2018.

Describing less-well represented voters using social-circle expectations. Why do social-circle expectations contribute so much to the bootstrap forecast? In large national polls with random recruitment of participants, such as ours, we expect little overlap in participants social-circles. This means that the correlation between our participants' social-circle estimates will be low. It can therefore benefit from the error-reducing properties of aggregating independent judgments, capitalizing on the main mechanisms behind traditional wisdom-of-crowds effects. There, the dependency between judgments is main factor that reduces the accuracy of aggregated judgments (*25–27*). In contrast, the state-winner expectations could be prone to correlated judgements, as all respondents can be influenced by the same background evidence about the election received from the media. In addition, social-circle expectations can provide information about the voting population beyond what is captured by own intentions questions, implicitly increasing representativeness (*7*, *17*). We explore this in two ways, by analyzing age and education questions, and by analyzing hidden voter questions.

First, we investigate how well social-circle questions can capture demographic characteristics, such as age and education. This is important, as it is well known that, for example, younger and lower educated people are often more difficult to reach by pollsters (*28*, *29*). We asked a sample of N=5,351 participants in the 2020 UAS panel to estimate the percentage of their social circles that have different age and education levels. Our results show that the percentages of social-circle estimates for age and education are closer to the population values (based on the 2019 American Community Survey) than the percentages of these people among our actual poll participants. In particular, the percentage of younger people and people with lower education estimated from social-circle estimates is closer to the population values than the percentage of these people among our actual poll participants even after weighting of responses (Figure S2 in the Supplementary information).

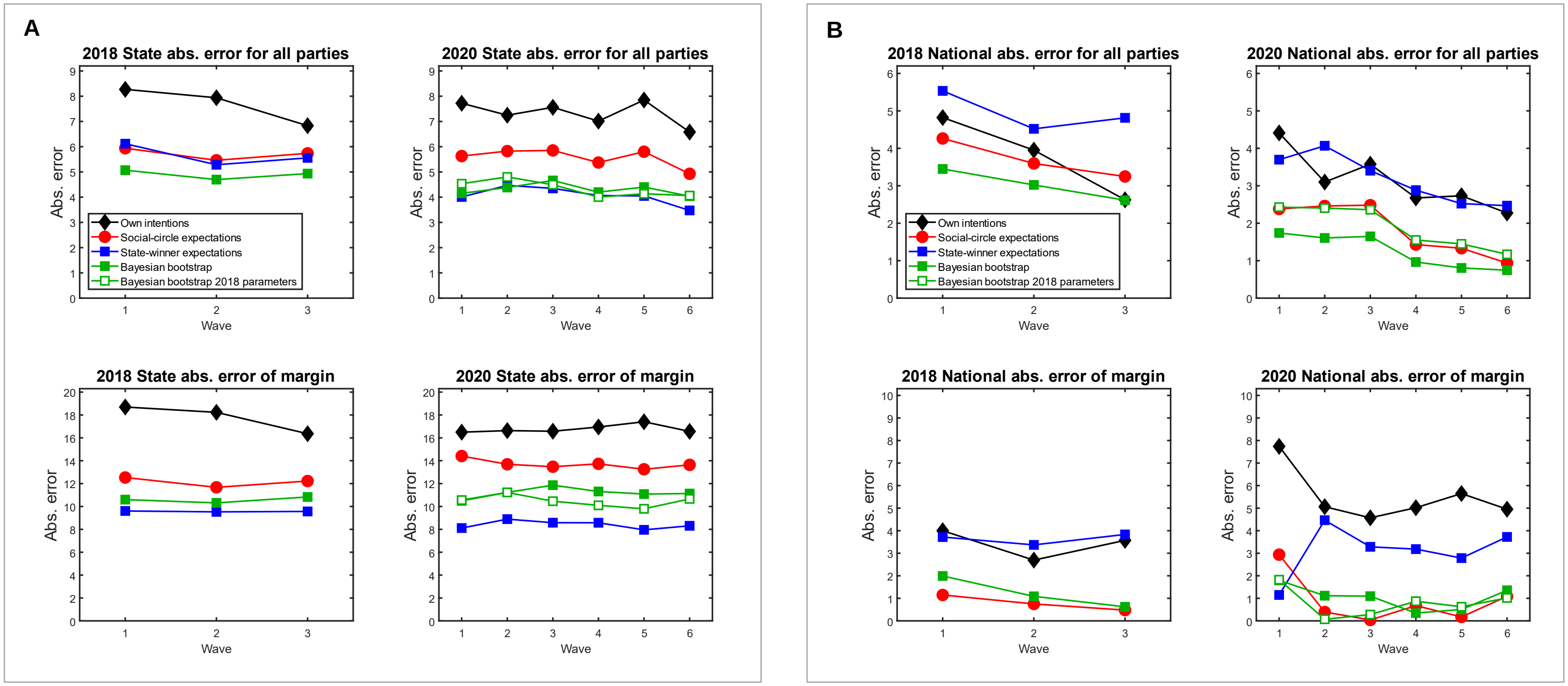

Figure 1. State and national level errors for different methods of forecasting the 2018 and 2020 U.S. Elections. Panel A: Average absolute error of state-level forecasts for the 2018 (left columns) and 2020 (right columns) U.S. election, for vote share of all parties (top panels) and absolute margin between the two main parties/candidates (bottom panels). Panel B: Absolute error of national forecasts for the 2018 (left columns) and 2020 (right columns) U.S. election, for vote share of all parties/candidates (Democrat/Republican/Other, top panels) and absolute margin between the two main parties/candidates (Democrat-Republican, bottom panels). For the 2020 election, there are two versions of the Bayesian bootstrap: one (filled green rectangles) that estimates parameters from the 2020 waves, and another that uses the parameters from the 2018 waves (unfilled green rectangles). Lower values indicate more accurate forecasts. Results are for different polling waves conducted before each election. All predictions are based on the survey responses multiplied with the probability to vote. For the national predictions, the state-level predictions were survey weighted and weighted with the population in each state. Waves refer to the window of dates where the participants could answer the polling questions, starting on August 22 in 2018 and August 11 in 2020. Information about sample sizes and polling dates can be found in Tables S1 and S2.

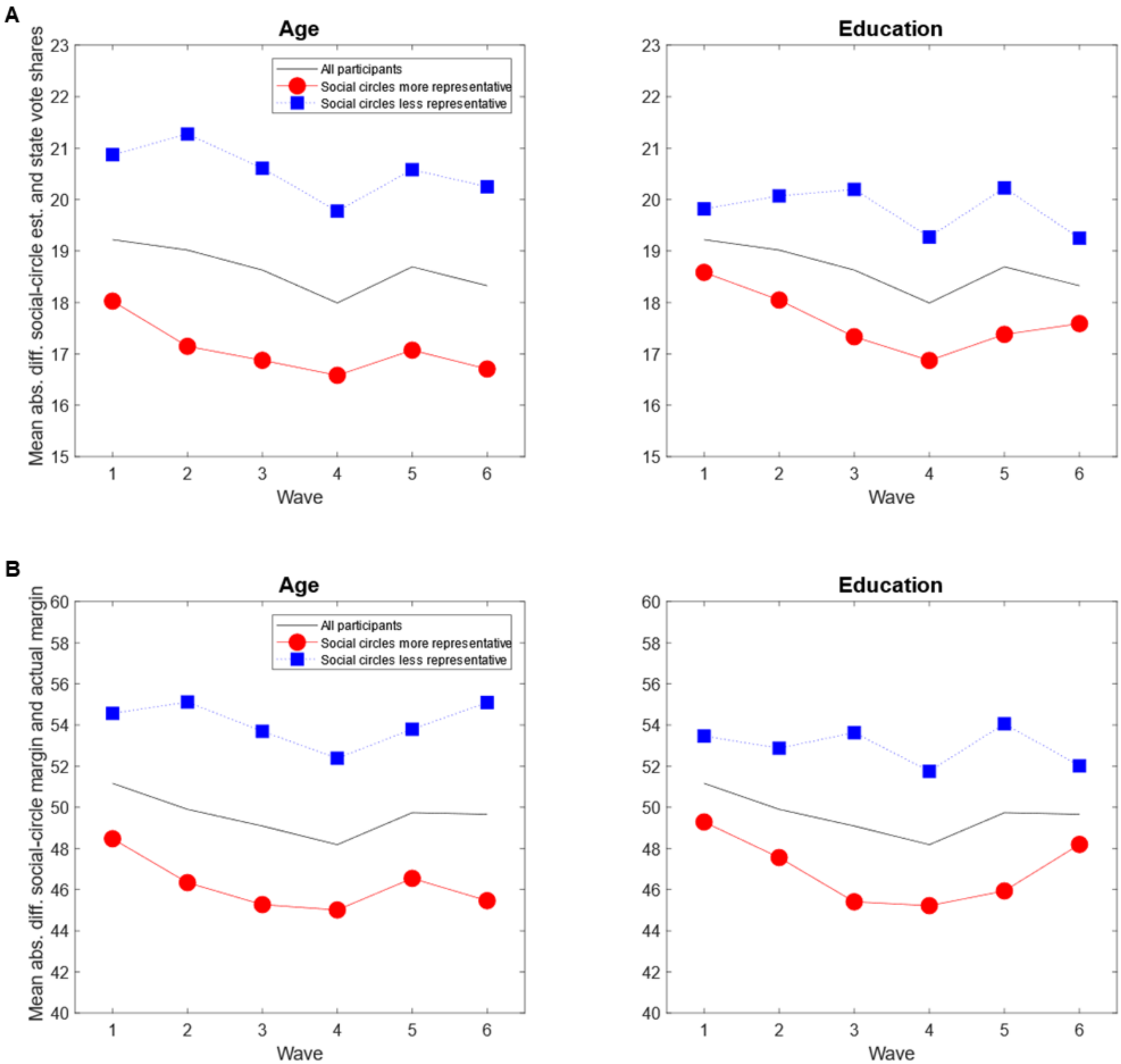

Figure 2. Representativeness of voting expectations in social circles for participants with social circles more representative by age and education (circles), and participants with social circles less representative by age and education (squares) (lower values indicate higher representativeness). Panel A: Mean absolute differences between individual social-circle estimates and state vote shares. Panel B: Mean absolute differences between individual social-circle estimates of the Biden-Trump margin and the actual margin. Participants were assigned to the two groups with median splits on the absolute differences between the social-circle estimates for age and education and state population values. As a comparison, the average differences for all participants are also shown (black lines).

If demographic representativeness of social circles is important for election forecasts, we expect that participants whose social circles are more similar to the state population in terms of age and education will also have voting expectations in their social-circles that are more representative of the voting intentions in the state population. Figure 2 shows the results for representativeness of voting expectations in social circles for groups of participants with social circles more representative by age and education, and participants with social circles less representative by age and education. The groups were determined by median splits on the absolute differences between the social-circle estimates for age and education and state population values. To measure the degree of representativeness in voting expectations, we calculated the absolute difference between individual social-circle estimates and the state vote shares in the 2020 election for all candidates (Panel A) and the average absolute difference between individual social-circle estimates of the Biden-Trump margin and the actual margin (Panel B). Across all six waves, participants with social circles more representative by age and education also have lower absolute differences between individual social-circle estimates and the state vote shares (Panel A) and the Biden-Trump margin (Panel B).

Second, the other group of less-well represented potential voters might be people who are reluctant to report their true voting intentions because of embarrassment, fear of harassment, or even willing obstruction of pollsters (*30–32*). To investigate the prevalence of these 'hidden voters', we asked the same sample of participants to estimate the percentage of their social circles who might be reluctant to reveal their opinions about Biden or Trump due to these reasons (see Methods for details of the questions asked). We use the answers to these questions to estimate the percentage of hidden voters in the voting population in each state. Note that for this estimate it is not enough to rely only on the percentage of hidden voters for a particular candidate. We need to take into account what percent of voters for that candidate is in participants' social circles and the likelihood that these potential voters will actually vote. In order to arrive at this estimate, we multiply participants' hidden voter estimates for a particular candidate in their social circle with both their estimates of the expected percentage of their social circle who will vote for that candidate and of the percentage that will actually vote. In a final step, we weight the responses in the same way as the forecasts to increase the national representativeness of the sample (see Weighting of forecasts in Methods). For these measures, the estimated percentages range from around five to thirteen percent, with higher estimates for the percentage of embarrassed, afraid of harassment, and willing obstructers among Trump compared to Biden voters (Figure S3, see also (*33*).

Participants whose social circles include only a few supporters of a particular candidate are more likely to report that their social contacts would be reluctant to reveal their opinion about that candidate because they are embarrassed or fear harassment (Table S3). We calculated the perceived support for each candidate in the participants' social circles by taking the signed differences between social-circle estimates of the percentages for Trump and Biden and the signed differences between Biden and Trump. We regressed each of the hidden voter questions for Trump (Biden) on the responses to the own intentions question and the perceived social-circle support for Trump (Biden). As shown in Table S3, controlling for own intentions, participants who perceive fewer supporters of Trump (or Biden) voters in their social circles, are also more likely to respond that more of their social contacts would be embarrassed and harassed if they would admit that they will vote for Trump (or Biden). In addition, there is evidence that own intentions also have an effect. If participants are more likely to vote for Trump (or Biden), those participants are also more likely to respond that more of their social contacts would be embarrassed or afraid of harassment if they would admit that they will vote for Trump (or Biden).

We can use participants' estimated percentages of hidden voters for both candidates in their social circles as a measure of how representative the voting expectations in their social circles are for the overall population of voters in their state. Given that the population of voters almost certainly includes some percentage of voters who would prefer to remain hidden to pollsters for various reasons (14, Figure S3), then participants with a larger estimated percentage of hidden voters in their social circles might have social circles that better reflect state voting patterns, compared with those participants who report knowing fewer or no hidden voters. To get a summary measure of the prevalence of hidden voters we summed the estimates for Biden and Trump for each of the hidden voter questions. We then used a median split on these estimates within each state to select participants that have a low or high percentage of hidden voters in their social circles. It is possible that the same participants have several of the reasons included in our hidden voter questions, so the resulting estimate of the total number of hidden voters is likely an overestimate. However, this should not hurt our analysis, which uses this sum only to do a median split and compare people with more or less hidden voters in their social circles. To measure the degree of representativeness of social-circle expectations, as before, we first calculated the absolute difference between individual social-circle estimates and the state vote shares in the 2020 election for all candidates. Then we calculated the average absolute difference between individual social-circle estimates of the Biden-Trump margin and the actual margin. Figure 3 shows the results for representativeness of social-circle expectations. For all three hidden voter questions, the absolute differences between individual social-circle estimates and the state vote shares (Panel A) and the Biden-Trump margin (Panel B) were lower for participants with a high percentage of hidden voters than for participants with low percentage of hidden voters.

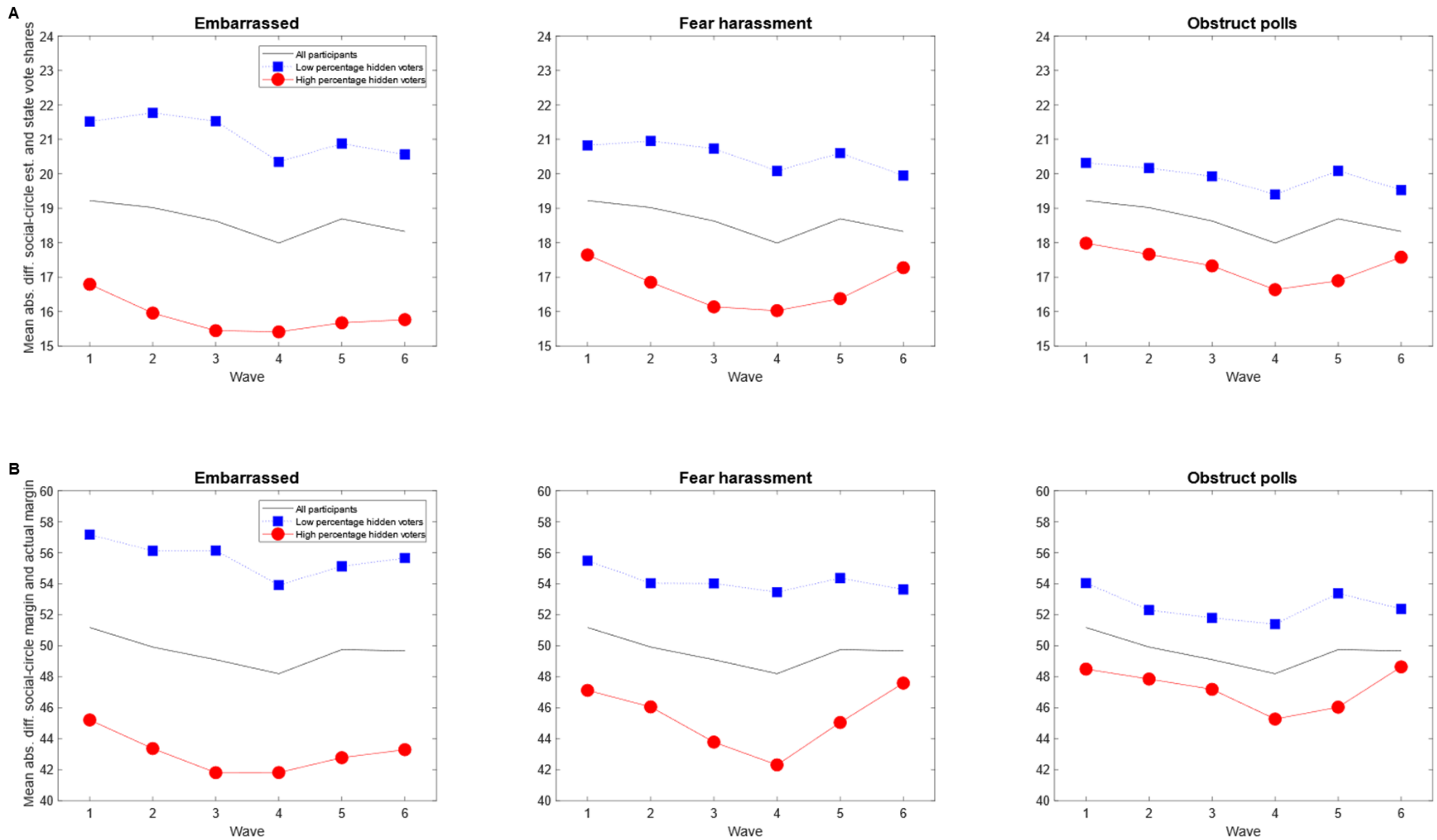

Figure 3. Representativeness of voting expectations in social circles for participants with high (circles), or low percentages (squares), of estimated hidden voters in their social circles (lower values indicate higher representativeness). Panel A: Panel A: Mean absolute differences between individual social-circle estimates and state vote shares. Panel B: Mean absolute differences between individual social-circle estimates of the Biden-Trump margin and the actual margin. Participants were assigned to the two groups with median splits on the sum of the estimated percentage of hidden voters for Biden and Trump. As a comparison, the average differences for all participants are also shown (black lines).

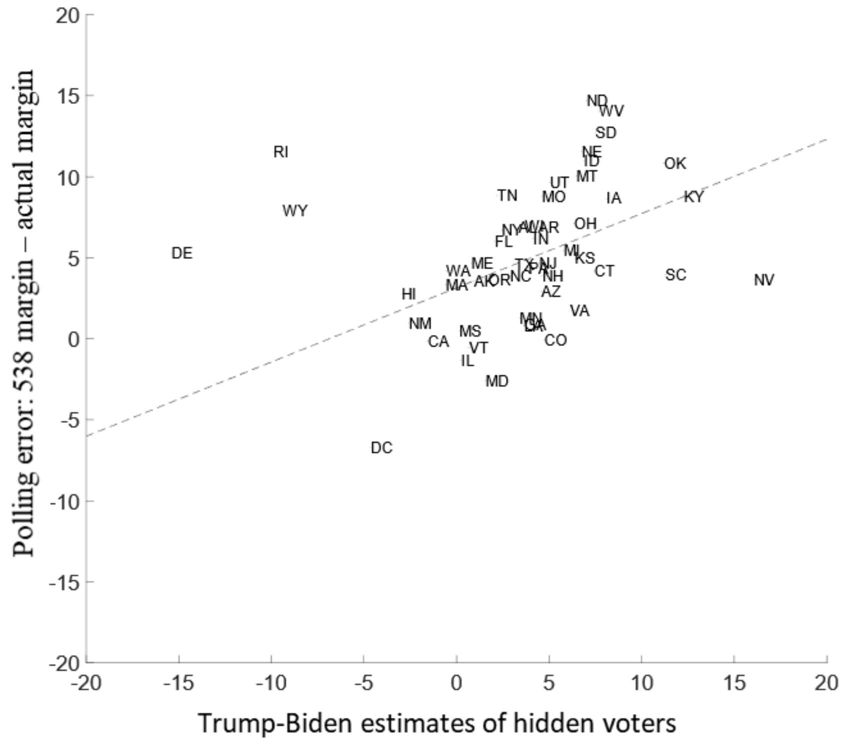

Figure 4. Relationship between average estimate of the difference between the percentages of hidden Trump and Biden voters based on estimates of our survey participants (*N*=5,331) and the error of predicted margin for Biden according to average state polls (as reported by 538.com). The positive correlation between the two measures suggests that hidden voters could have contributed to the prediction errors of state polls. On the y-axis, larger positive values mean a larger overestimation of Biden, and negative overestimation of Trump. The linear trend line was estimated with weighted least squares, with the square root of the number of participants in each state acting as weights. The data labels show the abbreviated state names.

These hidden voter results imply that hidden voters who were not reflected in the traditional polls might have contributed to the errors of traditional polling forecasts in the 2020 election. To investigate this further, we calculate the difference in the estimated percentage of hidden Trump and hidden Biden voters that are embarrassed, fear harassment, or willingly obstruct pollsters (see above for the calculation of the estimate of hidden voters). In Figure 4, we compare the average percentage of these differences for any reason with the error of average state polls (as reported by 538.com) predicting Biden's advantage over Trump in different states. Larger positive values on the y-axis suggest more overestimation of the Biden margin and larger positive values on the x-axis suggest more hidden Trump voters. This relationship is positive (*r* =

.47) suggesting that 5. To the extent that social-circle expectations provide evidence about voting intentions of these hidden voters, this could increase the value of these expectations in the bootstrap forecast.

Discussion

Taken together, our results suggest it is possible to develop questions and methods that could help the polling community be more accurate and regain the public's trust in predicting elections. The polling community can adapt and profit from novel technologies (*34*) such as the bootstrap forecast and wisdom-of-crowds questions.

The accuracy of the Bayesian bootstrap forecast demonstrates the benefits of theoretically justified weighting of different information sources. The bootstrap forecast outperforms forecasts based on own intentions questions and provides insight into the relative informational value of different polling questions. Across different societal contexts of the 2018 and 2020 U.S. elections, the relative contribution of the wisdom-of-crowds questions to the Bayesian bootstrap forecast, in particular the social-circle expectations, was much stronger than the contribution of own intentions. The large contribution of social-circle expectations to the Bayesian bootstrap forecast might be due to at least two factors that have previously been suggested in the literature, but never tested empirically (*7*, *17*). First, they seem to alleviate under-representation of some demographic groups. In particular, the percentage of younger people and people with lower education estimated from social-circle estimates is closer to the population values than the percentage of these people among our actual poll participants, even after applying survey weights. Furthermore, participants with social circles that are more similar to the state population in terms of age and education report voting expectations in their social circles that are more representative of the voting intentions in the state population. Second, social-circle expectations seem to reduce hidden voter biases. Our participants estimated that between five to thirteen percent of their social contacts were embarrassed, afraid of harassment, or willing to obstruct pollsters, with higher estimates among Trump compared to Biden voters. Participants with a larger estimated percentage of such hidden voters among their social contacts also report social-circle expectations that better reflect actual state voting patterns.

While the comprehensive analyses of the 2016 and 2020 presidential elections (*6*, *28*) found little difference between Trump's vote shares estimated by self-administered and interviewer-administered polls, suggesting low prevalence of shy or embarrassed voters, there is evidence that self-administered web surveys do not eliminate such bias entirely (*32*). Our analyses suggest that beyond embarrassment, a significant percentage of hidden voters in the 2020 U.S. election might have been people who feared harassment, or who aimed to deliberately skew polling results. The correlation we found between the percentages of hidden voters and the error of the predicted Biden-Trump margin according to an average of many state polls (Figure 4) points to the possibility that hidden voters could have contributed to the prediction errors of state polls. Further research on Bayesian bootstrap method should investigate different versions of wisdom of crowd questions, their application in different survey modes, using different ways of asking about candidates (by candidate name or by party), the effects of the number of social contacts participants have, and the effect of network structure (*35–37*)

In sum, it important to consider methods that are more robust to problems of sampling an increasingly distrustful electorate. Our result suggests that election predictions can be improved using novel approaches to combine the evidence from different types of polling questions, including traditional own-intentions questions as well as wisdom-of-crowds questions. The Bayesian bootstrap method can be used to obtain forecasts in surveys of other topics beyond elections, including those related to different public opinions, health behaviors, and economic trends.

Methods

Sample

Participants were members of the Understanding America Study (UAS) at the University of Southern California's (USC) Dornsife Center for Economic and Social Research (*18*). This longitudinal panel includes close to 9,000 U.S. residents, randomly selected from among all households in the United States using address-based sampling. Participants who agreed to participated, but did not log in to the UAS website, were reminded by a combination of e-mail, mail, and phone. Members of recruited households who did not have Internet access were provided with tablets and Internet service. The full sampling scheme can be found on the UAS website (*18*). All participants gave informed consent. The two studies were approved by the University of Southern California Institutional Review Board.

2018 U.S. House of Representatives election. Starting from August 22, 2018, all members of the panel were invited to answer our election poll questions. We asked the questions in three survey waves before the election, which was held on November 6, 2018. A wave is defined as being within which the participants had the opportunity to answer the polling questions. In a fourth wave after the election we asked each member which party they voted for. Table S1 shows sample sizes, starting and end dates for the different waves. Sample sizes in each wave were determined by the Center for Economic and Social Research's Understanding America Study as sufficient to provide estimates of election results within an acceptable margin of error (+/-3 p.p. at the 95% confidence level, the procedure is described on USC Dornsife Understanding America Study website https://uasdata.usc.edu/). Survey weights were constructed by a raking procedure that matched the sample to national population benchmarks based on the May 2016 Current Population Survey age by sex, race/ethnicity, sex by education, and household size by income (see the section on Weighting of forecasts below).

2020 U.S. Presidential election. Starting from August 11, 2020, 8,355 eligible voters who are active members of the panel were invited to answer our election poll questions. Each member who agreed to participate was randomized to respond on a pre-assigned day of the week, distributed so that the full sample participates over a 14-day period. Respondents have until their next assigned wave day (or 14 days after their assigned date) to complete the survey. Data for the full sample is nearly complete after the first 14 days, but not final until the end of the full 28-day wave. We analyze data based on the six first full 28-day waves. We only included answers from non-overlapping days. A post-election survey was conducted between November 4 and November 15 (N= 4749). Table S2 shows sample sizes and starting dates for the 6 survey waves before the election. Of those that completed all waves in 2018, and those that completed all waves in 2020, only 56% of those participating in 2020 also participated in the 2018 study.

Question texts

The question texts give the alternate wordings for the two elections in square brackets.

Likelihood to vote: "What is the percent chance that you will vote in the [2018 election for the U.S. House of Representatives? / 2020 U.S. presidential election?".]". *Own intentions:* "If you do vote in the [2018 election for the U.S. House of Representatives, what is the percent chance that you will vote for the Democratic candidate? For the Republican candidate? For another party's candidate? / 2020 U.S. presidential election, what is the percent chance that you will vote for: Joe Biden, Donald Trump, Other?]". *Social contacts definition that preceded social-circle questions*: "Now we would like you to think of your friends, family, colleagues, and other acquaintances who live in your state, are at least 18 years of age, and who you have communicated with at least briefly within the last month, either face-to-face, or otherwise. We will call these people your social contacts". *Social-circle likelihood to vote:* "What percentage of your social contacts [are likely to vote in your state in the 2018 election for the U.S. House of Representatives? / that live in your state are likely to vote in the 2020 U.S. presidential election?]". *Social-circle expectations:* "Out of all your social contacts who live in your state and are likely to vote in the [2018 election, what percentage do you think will vote for a Democratic candidate? For a Republican candidate? For another party's candidates? / 2020 U.S. presidential election, what percentage do you think will vote for: Joe Biden, Donald Trump, Other?]". *State-winner expectations:* "Of all people who live in your state and are likely to vote, what percentage do you think will vote for [a Republican candidate? For a Democratic candidate? For another party's candidate? / Joe Biden, Donald Trump, Other?]". Before the 2020 election, we also asked about: *social-circle size*: "How many people do you consider being your social contacts?". *Hidden voter questions*: "Think about your social contacts. If approached by pollsters, what percentage of your social contacts would: a) be embarrassed to reveal their opinions about Biden? _% b) be embarrassed to reveal their opinions about Trump? _%, c) be afraid that they will be facing negative consequences (e.g., harassment) if they reveal their opinions about Biden? _%, d) be afraid that they will be facing negative consequences (e.g., harassment) if they reveal their opinions about Trump? _%, e) try to skew poll results by saying they will vote for Trump although they intend to vote for Biden? _% f) try to skew poll results by saying they will vote for Biden although they intend to vote for Trump? _%". Before the 2020 election , we also asked two questions social-circle questions about *age and education*: "When asked, "How old are you?", what percentage of your social contacts would give each of the following answers:__% 18-29, __% 30-49, __% 50-69,__% 70+" and "When asked "What is your highest level of education?", what percentage of your social contacts would give each of the following answers?__% high school or less,__% some college or college degree,__% graduate degree".

Estimating hidden voters

We asked the hidden voter questions in wave 4 (N=5,331). To estimate the percentage of hidden voters for each of the three questions about embarrassment, fear of harassment, and obstruction of polls, we multiply participants' hidden voter estimates for a particular candidate in their social circle with both their estimates of the expected percentage of their social circle who will vote for that candidate and their estimate of the percentage that will actually vote. For Figure

S3, we also weighted the responses in the same way as the forecasts to increase the national representativeness of the sample (see Weighting of forecasts below).

Forecasts from poll data

Predictions for the 2020 election were preregistered (available at https://osf.io/zva6s and https://osf.io/x8jfk).

Forecasts based on own intentions. We multiplied each participant's measure of likelihood to vote by the measure of own intentions to provide a basis for vote-share forecasts. We then calculated the vote shares as a ratio of the average of these values and the average likelihood to vote, using a ratio estimator for the population mean (*7*, *39*). We calculated forecasts on the national and state levels. On the national level, we calculated means for each state, and then averaged the state forecasts weighted by state population size 18+ to obtain national forecasts (Table S1). On the state level, we calculated forecasts of state vote shares as means of forecasts for participants in a given state.

Forecasts based on social-circle expectations. We multiplied each participant's measure of social-circle likelihood to vote with the measure of social-circle expectations to provide a basis for vote-share forecasts. We used these individual values in the same ways as forecasts based on own intentions described above.

Forecasts based on state-winner expectations. These forecasts are the unaltered responses to the state-winner expectations question. The individual values are used in the same ways as forecasts based on own intentions described above.

Forecasts based on the Bayesian bootstrap. See the main text and the Bayesian bootstrap section in the Supplementary information.

Weighting of forecasts

To increase the national representativeness of our sample, participants' responses were weighted by poststratification weights that align the survey distributions to benchmarks for region, urbanicity, vote in 2016, sex, age, race, and education for the whole country. A description of the post-stratification weight procedure can be found in the document USC Dornsife National 2020 Election Tracking Poll Methodology available on the USC Dornsife Understanding America Study website (https://uasdata.usc.edu/). We apply the poststratification weights only to the national level forecasts, not to the state forecasts, as they are designed to align the sample with national level characteristics. In addition to the forecasts, we also weight the own and social-circle estimates for age and education in Figure S2 and the hidden voter estimates shown in Figure S3.

Election results

We used the national election results collected by The Cook Reports Election tracker for 2018 and 2020 (*40*, *41*). At the national level, we used the national popular vote as the benchmark. At the state level, we used the state's popular votes as the benchmark. We first computed the errors in predicting each states' popular vote and then averaged all values to get an overall measure of state-level accuracy.

Code availability

After registering as a data user on https://uasdata.usc.edu, the Matlab code used for different analyses are available from the corresponding author upon request.

Data availability

The USC poll data, based on the UAS surveys, can be downloaded from https://uasdata.usc.edu after registering on the UAS site as a data user.

References

1. 2020 election forecast, (available at https://projects.fivethirtyeight.com/2020-election-forecast/).
2. G. Dickie, Why polls were mostly wrong. *Sci. Am.* (2020), (available at https://www.scientificamerican.com/article/why-polls-were-mostly-wrong/).
3. J. W. Campbell, An embarrassing failure for election pollsters. *Conversat.* (2020), (available at https://theconversation.com/an-embarrassing-failure-for-election-pollsters-149499).
4. The polls for this election were off again. Is this the end of the industry? *NPR*, p. 2020.
5. G. Sherwood, If traditional polling is dead, what's next? *Hill* (2020), (available at https://thehill.com/blogs/pundits-blog/presidential-campaign/306456-if-traditional-polling-is-dead-whats-next).
6. J. D. Clinton, et al., "Task force on 2020 Pre-election polling: An evaluation of the 2020 general election polls" (2020), (available at https://www.aapor.org/Education-Resources/Reports/2020-Pre-Election-Polling-An-Evaluation-of-the-202.aspx).
7. M. Galesic, W. Bruine de Bruin, M. Dumas, A. Kapteyn, J. E. Darling, E. Meijer, Asking about social circles improves election predictions. *Nat. Hum. Behav.* **2**, 187–193 (2018).

8. A. Graefe, Accuracy of vote expectation surveys in forecasting elections. *Public Opin. Q.* **78**, 204–232 (2014).
9. A. Graefe, Forecasting proportional representation elections from non-representative expectation surveys. *Elect. Stud.* **42**, 222–228 (2016).
10. A. E. Murr, "Wisdom of crowds"? A decentralised election forecasting model that uses citizens' local expectations. *Elect. Stud.* **30**, 771–783 (2011).
11. W. Bruine De Bruin, Asking about social circles improves election predictions (BSPA 2020) (2020), (available at https://www.youtube.com/watch?v=v9WSmM8VeQ0).
12. V. A. Satopää, R. Pemantle, L. H. Ungar, Modeling probability forecasts via information diversity. *J. Am. Stat. Assoc.* **111**, 1623–1633 (2016).
13. PollyVote, (available at https://pollyvote.com/en/).
14. A. Graefe, Accuracy gains of adding vote expectation surveys to a combined forecast of US presidential election outcomes. *Res. Polit.* **2**, 205316801557041 (2015).
15. T. Clemen, Combining forecasts: A review and annotated bibliography. *Int. J. Forecast.* **5**, 559–583 (1989).
16. A. Graefe, J. S. Armstrong, R. J. Jones, A. G. Cuzán, Combining forecasts: An application to elections. *Int. J. Forecast.* **30**, 43–54 (2014).
17. D. M. Rothschild, J. Wolfers, "Forecasting elections: Voter intentions versus expectations," *SSRN* (2012), , doi:10.2139/ssrn.1884644.
18. Understanding America Study, (available at https://uasdata.usc.edu/index.php).
19. M. Galesic, H. Olsson, J. Rieskamp, A sampling model of social judgment. *Psychol. Rev.* **125**, 363–390 (2018).
20. C. Schulze, R. Hertwig, T. Pachur, Who you know is what you know: Modeling boundedly rational social sampling. *J. Exp. Psychol. Gen.* **150**, 221–241 (2021).
21. T. Pachur, R. Hertwig, J. Rieskamp, Intuitive judgments of social statistics: How exhaustive does sampling need to be? *J. Exp. Soc. Psychol.* **49**, 1059–1077 (2013).
22. M. Galesic, H. Olsson, J. Rieskamp, Social sampling explains apparent biases in judgments of social environments. *Psychol. Sci.* **23**, 1515–1523 (2012).
23. H. Olsson, W. B. De Bruin, M. Galesic, D. Prelec, A comparison of different methods to predict the 2020 United States presidential election: Addendum (2020), (available at https://osf.io/j54rz).

24. W. Bruine de Bruin, D. Bennett, Relationships between initial COVID-19 risk perceptions and protective health behaviors: A national survey. *Am. J. Prev. Med.* **59**, 157–167 (2020).
25. J. Surowiecki, *The wisdom of crowds* (Doubleday, New York, 2004).
26. C. P. Davis-Stober, D. V. Budescu, J. Dana, S. B. Broomell, When is a crowd wise? *Decision*. **1**, 79–101 (2014).
27. P. Atanasov, P. Rescober, E. Stone, S. A. Swift, E. Servan-schreiber, P. Tetlock, L. Ungar, B. Mellers, “Distilling the wisdom of crowds: Prediction markets versus prediction polls” (2015).
28. C. Kennedy, M. Blumenthal, S. Clement, J. D. Clinton, C. Durand, C. Franklin, K. McGeeney, L. Miringoff, K. Olson, D. Rivers, L. Saad, G. E. Witt, C. Wlezien, An evaluation of the 2016 election polls in the united states. *Public Opin. Q.* **82**, 1–33 (2018).
29. What the 2020 electorate looks like by party, race and ethnicity, age, education and religion, (available at https://www.pewresearch.org/fact-tank/2020/10/26/what-the-2020-electorate-looks-like-by-party-race-and-ethnicity-age-education-and-religion/).
30. Could “Shy” Trump Voters’ Discomfort With Disclosing Candidate Choice Have Skewed Telephone Polls? Evidence from the USC Election Poll, (available at https://healthpolicy.usc.edu/evidence-base/could-shy-trump-voters-discomfort-with-disclosing-candidate-choice-skew-telephone-polls-evidence-from-the-usc-election-poll/).
31. D. J. Hopkins, No more Wilder effect, never a Whitman effect: When and why polls mislead about black and female candidates. *J. Polit.* **71**, 769–781 (2009).
32. P. K. Enns, J. Lagodny, J. P. Schuldt, Understanding the 2016 US presidential polls: The importance of hidden Trump supporters. *Stat. Polit. Policy*. **8**, 41–63 (2017).
33. Study: Are election 2020 poll respondents honest about their vote?, (available at https://www.cloudresearch.com/resources/blog/election-2020-poll-respondent-honesty/).
34. R. M. Groves, Three eras of survey research. *Public Opin. Q.* **75**, 861–871 (2011).
35. S. L. Feld, Why your friends have more friends than you do. *Am. J. Sociol.* **96**, 1464–1477 (1991).
36. S. L. Feld, A. McGail, Egonets as systematically biased windows on society. *Netw. Sci.* **8**, 399–417 (2020).
37. N. A. Christakis, J. H. Fowler, Social network sensors for early detection of contagious outbreaks. *PLoS One*. **5**, e12948 (2010).

38. Understanding America Study: Weighting procedures, (available at https://uasdata.usc.edu/addons/documentation/UAS Weighting Procedures.pdf).
39. T. L. Gutsche, A. Kapteyn, E. Meijer, B. Weerman, The RAND continuous 2012 presidential election poll. *Public Opin. Q.* **78**, 233–254 (2014).
40. 2018 House Popular Vote Tracker, (available at https://cookpolitical.com/analysis/house/house-charts/2018-house-popular-vote-tracker).
41. 2020 National Popular Vote Tracker, (available at https://cookpolitical.com/2020-national-popular-vote-tracker).
42. D. Prelec, A Bayesian truth serum for subjective data. *Science (80-.).* **306**, 462–466 (2004).

Acknowledgements

We thank Jill Darling, Tania Gutsche, Marco Angrisani and the support from the whole team at the USC's Center for Economic and Social Research. H.O. and M.G. were partially supported by a grant from the National Science Foundation DRMS 1949432, and all authors by National Science Foundation MMS 2019982. W.B.d.B. was partially supported by Swedish Riksbanken Jubileumsfond's program on Science and Proven Experience. D.P. acknowledges the hospitality of All Souls College, Oxford during Michaelmas term 2020 when this paper was written. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies. This article relies on data from surveys administered by the Understanding America Study (UAS), which is maintained by the Center for Economic and Social Research (CESR) at the University of Southern California (USC), Dornsife. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of USC or UAS.

Author Contributions

H.O., M.G., W.B.d.B., and D.P. designed the research questions, data collection methods. and the survey questions. D.P. developed the Bayesian bootstrap method. H.O., D.P. and M.G. analyzed the data. All authors contributed to the writing of the paper.

Supplementary Information

Combining survey questions with a Bayesian bootstrap method yields accurate election forecasts

Henrik Olsson[1,a], Wändi Bruine de Bruin[b], Mirta Galesic[a], & Drazen Prelec[c]
[a]Santa Fe Institute, Santa Fe, NM, 87501; [b]Sol Price School of Public Policy, Dornsife Department of Psychology, Schaeffer Center for Health Policy and Economics, and Center for Economic and Social Research, University of Southern California, Los Angeles, CA 90089; [d]Sloan School of Management, Department of Economics, and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02142.

[1] Corresponding author: Henrik Olsson, Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501. Email: olsson@santafe.edu

The Bayesian bootstrap

This supplement sketches the theoretical argument for a Bayesian bootstrap approach for integrating stated own voting intentions with social-circle expectations and election-winner expectations. The supplement does not provide a complete definition, which will be developed elsewhere. The goal rather is to motivate the bootstrap forecast in the main text, and explain the heuristic steps and approximations involved in our estimation.

The core idea is that each respondent's election-winner expectations already reflect both private evidence reported in the survey (own intentions and social-circle expectations) as well as background evidence, e.g., what can be gleaned from news reports. The latter further divides into common (shared) evidence, available to everyone, and an independent individual evidence. Background evidence is expressed as a pseudocount sample of size $A+B$, where A is size of the common sample and B the independent individual sample. The social-circle estimate of each respondent is a sample of size N. We will assume that A,B,N do not vary across respondents.

The bootstrap election forecast attempts to estimate and aggregate the totality of this evidence. This requires, first, to establish the correct relative weighting of own, social circle, and background evidence (i.e., to estimate $N, A+B$), and, second, to avoid double counting common evidence (i.e., to estimate the division of $A+B$ between A and B). The common background evidence A should not be double-counted across respondents, so its contribution to the election forecast is diluted as the number of respondents increases. With a very large sample, the bootstrap forecast depends only on own intentions, social circle expectations, and estimated individual evidence, weighted in proportions $1:N:B$. With a small sample however, the forecast may be dominated by the common background evidence.

Below is the notation for reported information by respondent r, suppressing state IDs. n is the number of respondents (at the state level).

x_i^r — reported own intentions to vote for candidate $i = R, D, O$.

y_i^r — reported state-winner expectations for candidate $i = R, D, O$.

s_i^r — reported social-circle expectations for candidate $i = R, D, O$.

In addition to reported information, respondents receive background evidence about the election that is not reported in the polling questions:

α_i — % voting for candidate $i = R, D, O$ according to common background evidence

β_i^r — % voting for candidate $i = R, D, O$ according to individual background evidence of respondent r

$\bar{\beta}_i = \frac{1}{n} \sum_r \beta_i^r$ — average individual background evidence

We assume that background evidence (on the state level) is distributed according to a Dirichelet prior whose parameters $A\alpha_i + B\beta_i^r$ can be interpreted as the total amount of background evidence in favor of candidate i. Combining this with own intentions and social-circle expectations, x^r and s^r, the total amount of evidence in favor of candidate i, held by respondent r is therefore:

$$A\alpha_i + B\beta_i^r + x_i^r + Ns_i^r$$

Ideally, a respondent's state-winner expectations should match the expectations of the posterior distribution with respect to the above parameters, which is simply normalized total evidence:

$$y_i^r \approx \frac{A\alpha_i + B\beta_i^r + x_i^r + Ns_i^r}{A + B + 1 + N}$$

In the first step of the Bayesian bootstrap, state-winner expectations y_i^r are regressed against own intentions and social-circle expectations. This provides estimates of N, $A + B$ and the vote shares implied by background information. Intuitively, if own intentions variables have little impact on election expectations, that would indicate that the background evidence is precise and $A + B$ very large; similarly, the impact of social circle relative to own intentions in the regression provides an estimate of N. The second step then uses estimated parameters to generate the bootstrap forecast.

Assume now that the distributions of α_i and β_i^r are constant within each state, and A, B, N do not vary across individuals or states or options i. We estimate a constrained regression of y against x and s:

$$y_i^r = a_i + bx_i^r + cs_i^r + \varepsilon_i^r \quad i = D, R, O, \quad r = 1, \ldots, n$$

with the constraint $\sum_i a_i + b + c = 1$. We the interpret the coefficients as follows:

$$\begin{aligned} a_i &= \frac{A\alpha_i + B\bar{\beta}_i}{A+B+1+N} \quad i = D,R,O \\ b &= \frac{1}{A+B+1+N} \\ c &= \frac{N}{A+B+1+N} \end{aligned}$$

(The constraint is needed as $\sum_i \alpha_i = \sum_i \bar{\beta}_i = 1$ implies: $\sum_i a_i = (A+B)/(A+B+1+N) = 1-b-c$). From the estimated coefficients we back out the theoretical parameters using the coefficient for own voting intentions, b, as a 'unit of account:'

$$N = c/b$$

$$A\alpha_i + B\bar{\beta}_i = a_i/b$$

$$A+B = \sum_i (A\alpha + B\bar{\beta}_i) = \sum_i a_i/b$$

The first line estimates social-circle size N by comparing the regression coefficients for own intentions and social-circle expectations. The next two lines indicates that the intercepts a_i estimate background evidence $E_{\alpha+\beta}$ at the sample level.

The sample size B corresponding to the individual independent background evidence may be estimated from the residuals of the regression. The predicted value $\hat{y}_i^r$ for the respondent's state-winner expectations is:

$$\hat{y}_i^r = a_i + bx_i^r + cs_i^r \quad i = D,R,O, \quad r = 1,\ldots,n$$

or:

$$\hat{y}_i^r = \frac{A\alpha_i + B\bar{\beta}_i}{A+B+1+N} + \frac{1}{A+B+1+N}x_i^r + \frac{N}{A+B+1+N}s_i^r, \quad i = D,R,O, \quad r = 1,\ldots,n$$

The residual depends only the respondent's individual evidence, β_i^r:

$$\varepsilon_i^r = y_i^r - \hat{y}_i^r = \frac{B}{A+B+1+N}(\beta_i^r - \bar{\beta}_i) = \frac{B}{A+B}\frac{1}{(A+B+1+N)}((A+B)\beta_i^r - (A+B)\bar{\beta}_i)$$

or:

$$\sum_i Var(\varepsilon_i^r) = \frac{q}{A+B+1+N}\sum_i Var((A+B)\beta_i^r)$$

where $q = B/(A+B)$ is the quantity of interest, i.e., the fraction of background evidence that is individual and independent. Since α_i and β_i^r are presumed to have the same expectations, it is reasonable to estimate $Var((A+B)\beta_i^r))$ as the variance of the Dirichelet distribution with parameters $A\alpha_i + B\bar{\beta}_i = a_i/b$ already estimated from the linear regression:

$$\sum_i Var((A+B)\beta_i^r) = \sum_i \frac{(a_i/b)(A+B-(a_i/b))}{A+B+1}$$

Therefore:

$$\sum_i Var(\varepsilon_i^r) = \frac{q}{A+B+1+N} \sum_i \frac{(a_i/b)(A+B-(a_i/b))}{A+B+1}$$

yielding an estimate of the relative amount of independent background evidence:

$$q = \frac{(A+B+1)(A+B+1+N)}{\sum_i (a_i/b)(A+B-(a_i/b))} \sum_i Var(\varepsilon_i^r)$$

or:

$$q = \frac{(1-c)}{\sum_i a_i(\sum_{k\neq i} a_k)} \sum_i Var(\varepsilon_i^r)$$

The bootstrap forecast of fraction votes for option i weights background evidence by $(1-q+qn)$, own intentions by bn and social-circle expectations by cn:

$$\frac{(1-q+qn)\alpha_i + \sum_r x_i^r + N\sum_r s_i^r}{\sum_k (1-q+qn)\alpha_k + \sum_k \sum_r x_k^r + N\sum_k \sum_r s_k^r} = \frac{(1-q+qn)a_i + (b\sum_r x_i^r + c\sum_r s_i^r)}{\sum_k (1-q+qn)a_k + n(b+c)} \quad i = D, R, O$$

The weighting of background evidence by $(1-q+qn)$ ensures that common evidence is not overweighted. If $q = 0$, i.e., if all background evidence is common, then its influence in the bootstrap forecast rapidly declines with sample size. If $q = 1$, i.e., if background evidence is sampled independently by each respondent, then its influence scales just like reported own intentions and social-circle expectations.

Note that state-winner expectations are not directly weighted in the bootstrap forecast, but only indirectly through their impact on the parameters a_i, b, c and q estimated with the regression equation. However, in the large sample limit $(1-q+qn \approx qn)$ one may express the bootstrap forecast as a simple weighted average of the three inputs. Observe that the intercepts can be written in terms of the respondent inputs plus residuals, which in turn cancel out at the aggregate level:

$$a_i = y_i^r - (bx_i^r + cs_i^r + \varepsilon_i^r) = \frac{1}{n}\sum_r (y_i^r - (bx_i^r + cs_i^r)) \quad i = D, R, O, \quad r = 1, \dots, n$$

as $\sum_r \varepsilon_i^r = 0$. The numerator in the bootstrap forecast for i is therefore:

$$(1 - q + qn)a_i + (b\sum_r x_i^r + c\sum_r s_i^r) = (\frac{1-q}{n} + q)\sum_r y_i^r + (\frac{(1-q)}{n} + 1 - q)(b\sum_r x_i^r + c\sum_r s_i^r)$$

In the large sample limit $(n \to \infty)$ the bootstrap forecast weights state-winner expectations by $q/(q + (1-q)(b+c))$, own intentions by $b(1-q)/(q + (1-q)(b+c))$, social-circle expectations by $c(1-q)/(q + (1-q)(b+c))$. Using the exact weighting formula, one can show that the weight of state-winner expectations decreases with n iff $q < 1/2$.

Table S1

2018 U.S. House of Representatives election: National forecasts calculated as the weighted average of state forecasts based on different methods, compared to average 538.com national forecasts

	Election results (Nov 6) results	**Own intentions**	**Social-circle expectations**	**State-winner expectations**	**Bayesian bootstrap**	**538**
			Wave 1 (Aug 22 - Sep 11, *N*=4,511)			
Republicans	44.8	39.2	42.2	42.5	43.0	-
Democrats	53.4	51.8	49.7	47.4	49.6	-
Other	1.7	9.0	8.1	10.1	6.5	-
Democrats-Republicans	8.6	12.6	7.5	4.9	6.6	7.4
			Wave 2 (Sep 14 - Oct 4, *N*=4,259)			
Republicans	44.8	40.5	42.5	43.1	43.0	-
Democrats	53.4	51.8	50.4	48.4	50.5	-
Other	1.7	7.7	7.1	8.5	6.0	-
Democrats-Republicans	8.6	11.3	7.8	5.2	7.5	9.4
			Wave 3 (Oct 15 - Nov 5, *N*=5,038)			
Republicans	44.8	41.1	42.2	43.1	43.0	-
Democrats	53.4	53.2	51.2	47.9	51.0	-
Other	1.7	5.7	6.6	9.0	5.4	-
Democrats-Republicans	8.6	12.2	9.1	4.8	8.0	9.9

Note. National forecasts are average of state forecasts weighted by the size of a state population 18+. The 538 forecasts are the average forecasts made within the date range for our three waves.

Table S2

2020 U.S. presidential election: National forecasts calculated as the weighted average of state forecasts based on different methods, compared to average 538.com national forecasts

	Election results (as of Dec 17, 2020)	Own intentions	Social-circle expectations	State-winner expectations	Bayesian bootstrap	Bayesian bootstrap 2018 parameters	538
				Wave 1 (Aug 8-Aug 24 *N*=4,330)			
Trump	46.9	40.2	43.6	44.7	44.5	44.1	46.2
Biden	51.3	52.4	51.0	48.0	50.7	50.5	52.5
Other	1.8	7.3	5.4	7.4	4.0	5.3	1.3
Biden-Trump	4.5	12.2	7.4	3.3	6.2	6.4	6.4
				Wave 2 (Aug 25-Sep 7 *N*=4,583)			
Trump	46.9	42.2	44.8	46.0	45.9	45.1	46.2
Biden	51.3	51.7	49.7	46.0	49.3	49.6	52.6
Other	1.8	6.1	5.5	7.9	3.7	5.3	1.2
Biden-Trump	4.5	9.5	4.8	0.0	3.3	4.5	6.4
				Wave 3 (Sep 8–Sep 21 *N*=4,702)			
Trump	46.9	41.9	45.0	45.9	45.9	44.9	46.0
Biden	51.3	50.9	49.4	47.1	49.2	49.7	52.7
Other	1.8	7.2	5.5	6.9	3.6	5.1	1.3
Biden-Trump	4.5	9.0	4.4	1.2	3.4	4.8	6.6
				Wave 4 (Sep 22–Oct 5 *N*=4,277)			
Trump	46.9	42.9	45.5	46.3	46.2	45.3	45.9
Biden	51.3	52.4	50.6	47.6	50.3	50.7	52.8
Other	1.8	4.7	3.9	6.1	2.9	4.0	1.3
Biden-Trump	4.5	9.5	5.1	1.3	4.1	5.4	7
				Wave 5 (Oct 6–Oct 19 *N*=4,595)			
Trump	46.9	42.8	45.8	46.4	46.4	45.8	45.4
Biden	51.3	52.9	50.5	48.1	50.4	50.4	53.4
Other	1.8	4.3	3.7	5.5	2.7	3.8	1.2
Biden-Trump	4.5	10.1	4.6	1.7	3.9	4.5	8.1
				Wave 6 (Oct 20–Nov 2 *N*=4,835)			
Trump	46.9	43.4	46.6	46.4	46.8	46.4	45.4
Biden	51.3	52.4	50.0	47.9	50.0	50.0	53.4
Other	1.8	3.9	3.4	5.8	2.8	3.7	1.2
Biden-Trump	4.5	9.3	3.4	1.5	3.2	3.6	8

Note. National forecasts are average of state forecasts weighted by the size of a state population 18+. The 538 forecasts are the average forecasts made within the date range for our six waves.

Table S3

Results from regressing each of the hidden voter questions for Trump on the answers to the own intention question (Own intentions) for Trump and the signed differences (Social circle diff.) between the social-circle percentages of Biden and Trump (the Biden questions) and the Trump-Biden difference (Trump questions), and in the same way for the Trump questions. The regressions are across all participants in wave 6 that had answered the hidden voter questions and answered the own intention and social-circle questions (N=4,835)

	Biden questions				**Trump questions**			
	Est	*SE*	*t*	*p*	*Est*	*SE*	*t*	*p*
		Embarrassed				**Embarrassed**		
Intercept	21.92	0.69	31.94	0.00	24.84	0.80	31.17	0.00
Own intentions	0.02	0.01	1.92	0.05	0.06	0.01	4.05	0.00
Social circle diff.	-0.04	0.01	-4.14	0.00	-0.06	0.01	-5.70	0.00
		Fear harassment				**Fear harassment**		
Intercept	24.84	0.70	35.28	0.00	27.36	0.86	31.99	0.00
Own intentions	-0.04	0.01	-3.45	0.00	0.11	0.02	7.35	0.00
Social circle diff.	-0.06	0.01	-6.78	0.00	-0.08	0.01	-6.78	0.00
		Obstruct polls				**Obstruct polls**		
Intercept	12.26	0.53	23.32	0.00	14.12	0.65	21.64	0.00
Own intentions	-0.02	0.01	-2.14	0.03	0.05	0.01	4.51	0.00
Social circle diff.	-0.01	0.01	-0.98	0.33	-0.01	0.01	-1.39	0.16

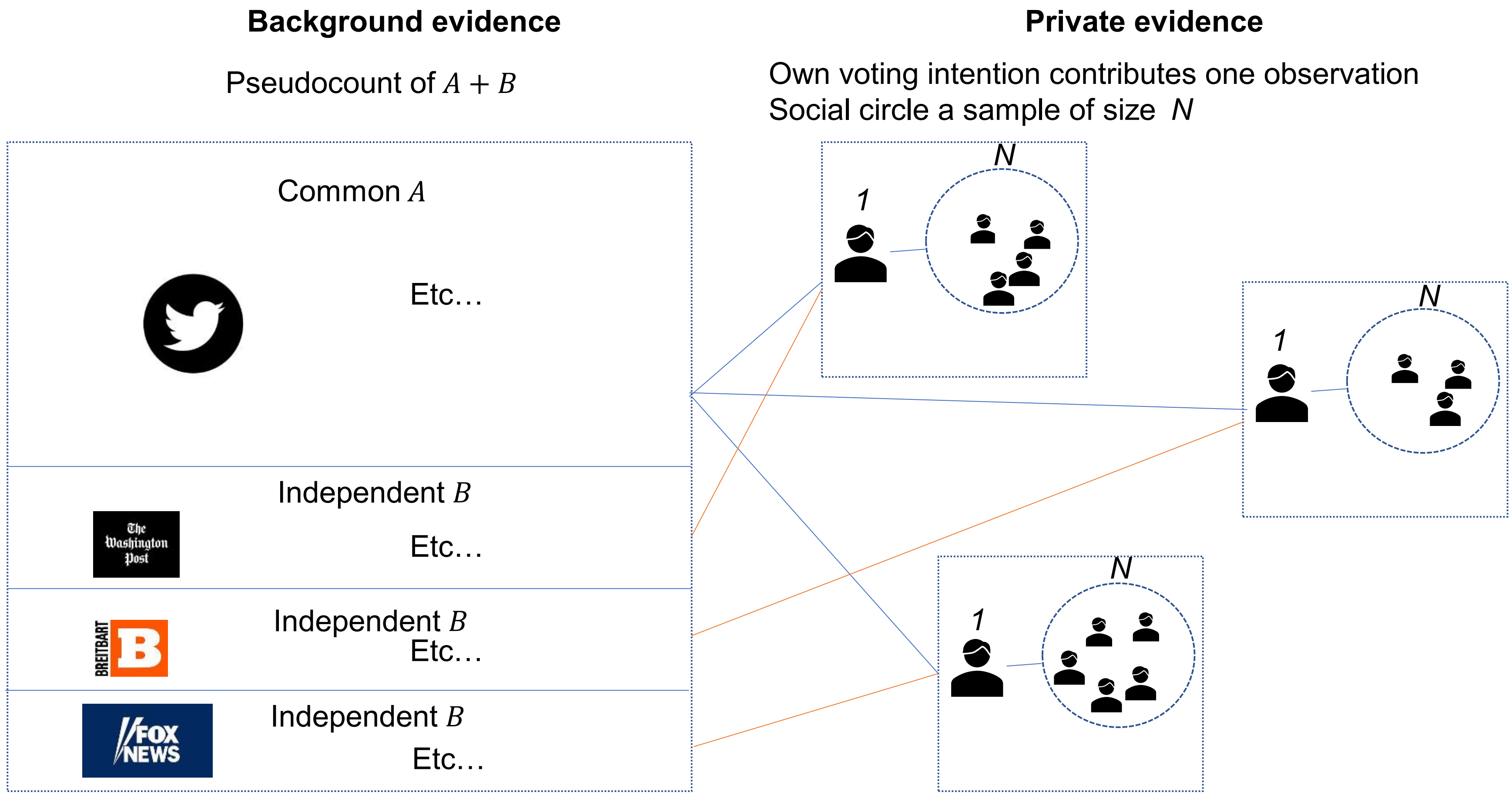

Figure S1. Main components of the Bayesian bootstrap: A participant's evidence divided into a private component that is reported directly and a background component, revealed indirectly via the election winner expectations. Background evidence partly derives from common sources and partly from independent sources unique to each participant. Evidence is conceptualized as a sample of observations. The 'unit of account' is the participant's own voting intention, which contributes exactly one observation. A participant is then assumed to sample N individuals from their social circle, and combine this with 'pseudocount' of A+B observations that represent common and independent background evidence.

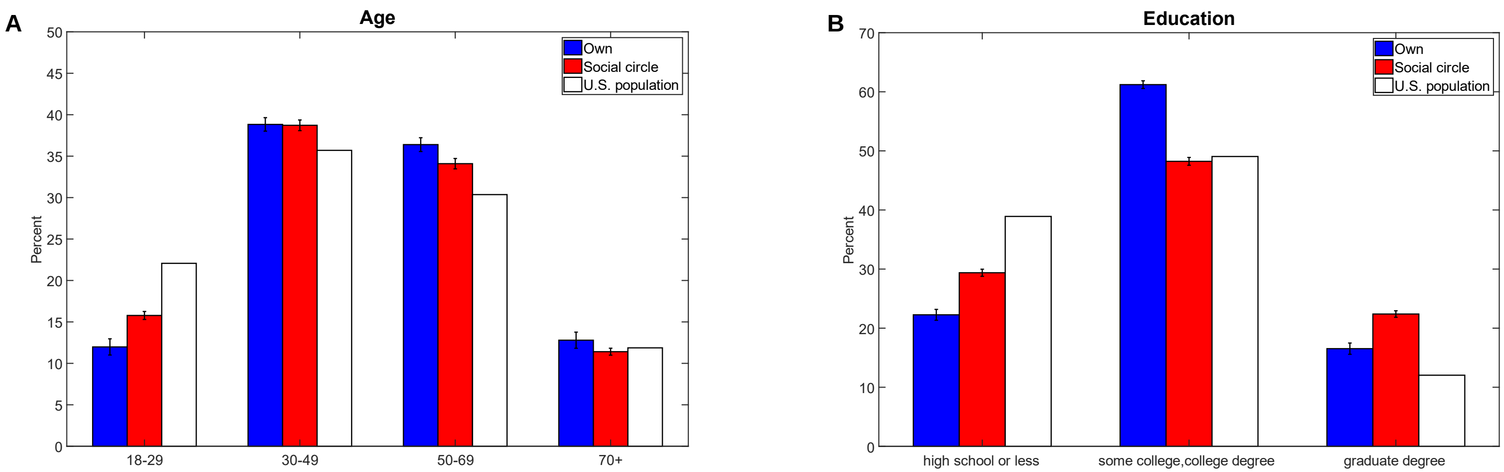

Figure S2. Panel A: Percentage of poll participants in 2020 (*N*=5,351) that reported that their age were in one of four age groups (blue bars), percentage of their reported social-circles in different age groups (red bars), and percentage of U.S. population based on Census data (white bars). Panel B: Percentage of poll participants (*N*=5,351) that reported to have different educational levels (blue bars), percentage of their reported social-circles in different educational levels (red bars), and percentage of U.S. population based on Census data from the 2019 American Community Survey (white bars). The answers to the age and education questions were weighted by poststratification weights (see Weighting of forecasts in the Methods section). Error bars are standard errors.

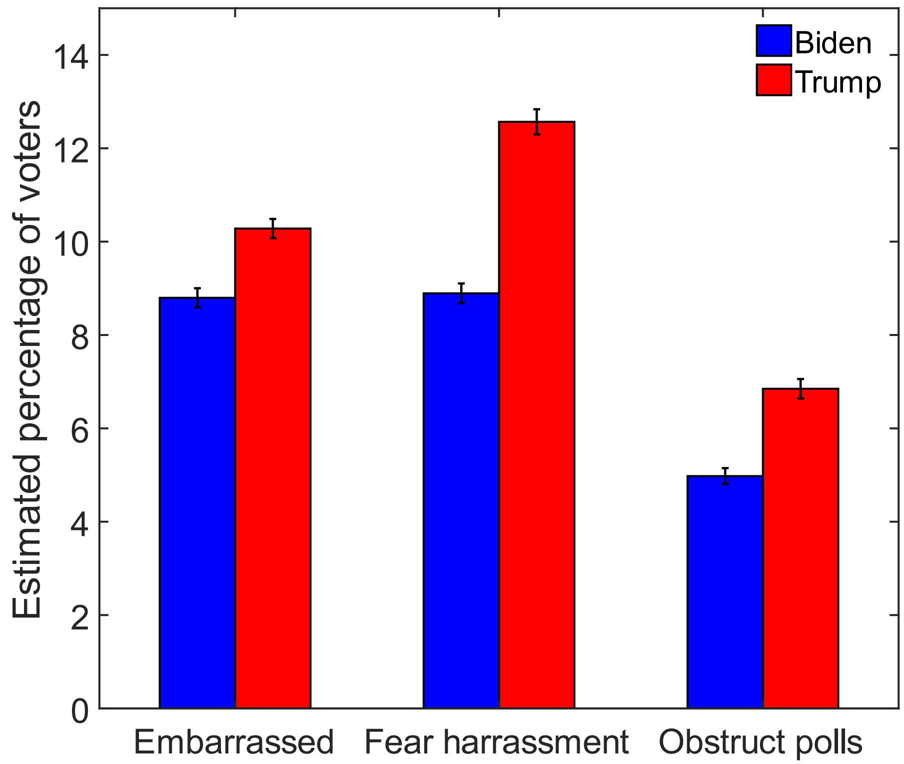

Figure S3. Estimated percentage of voters that would be embarrassed, fear harassment, or lie to obstruct pollsters when asked about their opinion about Biden and Trump, based on participants' estimates. Error bars are standard errors (*N*=5,331).

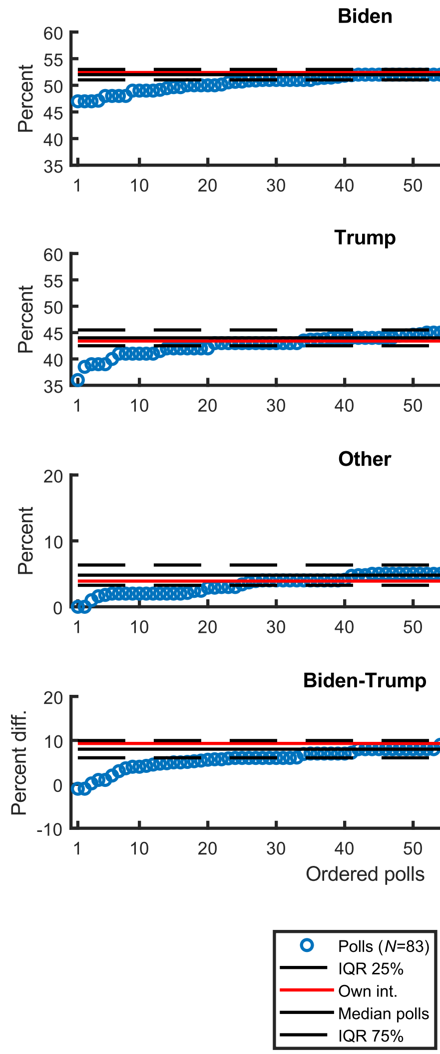

Figure S4. Percentages for national polls conducted between October 20 and November 2, 2020 as reported by 538. Full black lines show the median percentages of these polls and the dashed black lines show the interquartile range. Red lines are the results from the own intention question asked during the same dates (wave 6).