Projecting confidence: How the probabilistic horse race confuses and demobilizes the public

Sean J. Westwood[1], Solomon Messing[2], and Yphtach Lelkes *[3]

[1]Program in Quantitative Social Science, Dartmouth College
[2]Data Labs, Pew Research Center
[3]Annenberg School of Communication, University of Pennsylvania

March 6, 2018

Abstract

Horse race coverage in American elections has shifted focus from late-breaking poll numbers to sophisticated meta analytic forecasts that often emphasize candidates' probability of victory. We place this "probabilistic horeserace" in the context of Riker and Ordeshook (1968), and hypothesize that it will lower uncertainty about an election's outcome (perceived potential pivotality), which lowers turnout under the model. After demonstrating the prominence of probabilistic forecasts in election coverage, we use experiments to show that the public has difficulty reasoning about the probability of a candidates victory. Critically, when one candidate is ahead, win-probabilities convey substantially more confidence that she will win compared to vote share estimates. Even more importantly, we show that these impressions of probabilistic forecasts cause people not to vote in a behavioral game that simulates elections. In the context of the existing literature, the magnitude of these findings suggests that probabilistic horse race coverage can confuse and demobilize the public.

*We are grateful for helpful feedback from Kevin Arceneaux, Michael Delli Carpini, Claudia Deane, Michael Dimock, Dean Eckles, Ryan Enos, Jeffrey Friedman, Eitan Hersh, Dan Hopkins, Yusaku Horiuchi, Scott Keeter, Justin Grimmer, Justin H. Gross, Guy Grossman, Adam Hughes, Jocelyn Kiley, Courtney Kennedy, Onyi Lam, Herschel Nachlis, Andrew Mercer, Marc Meredith, Jonathan Mummolo, Sean Taylor, Rachel Weisel, and Stefan Wojcik

"I don't know how we'll ever calculate how many people thought it was in the bag, because the percentages kept being thrown at people—'Oh, she has an 88 percent chance to win!"' - Hillary Clinton quoted in Traister 2017

Horse race coverage focusing on campaign strategy and pre-election polling eclipses coverage of substantive issues in American elections (Iyengar, Norpoth, and Hahn, 2004), with an estimated four times as much coverage of the horse race as issue positions in 2016 (Patterson, 2016).[1] While political science has long considered the effects of horse race polls on behavior (e.g., Ansolabehere and Iyengar, 1994; Delli Carpini, 1984; Mutz, 1998), a new form of horse race coverage has emerged in recent elections: the probabilistic forecast.

The distinguishing feature of this *probabilistic horse race* is the presentation of election projections as probabilities of winning, $P(V_{share} > .5)$, instead of the expected vote share, $E(V_{share})$.[2] While probabilistic horse race coverage often employs sophisticated modeling, *any* electoral projection based on one or more polls can be presented *in either form.* This is true regardless of how the underlying data are aggregated, weighted, modeled to account for correlated errors and combined with other economic or non-survey data.[3]

Probabilistic forecasts—often consisting of complex meta-analyses based on hundreds of thousands of respondents—have grown in popularity in part because aggregating polls helps reduce bias and other forms of error compared to single one-off polls (Hillygus, 2011; Toff, 2017). This point was underscored in 2008 when probabilistic forecasts rose to prominence after at least one successfully predicted nearly every state's Senate race and presidential result (e.g., Silver, 2008). Since then, journalists have expressed "a growing interest in and reliance on polling aggregator websites fueled by demands for precise predictions" (Toff, 2017,

[1] The expanding coverage of the horse race parallels a dramatic proliferation of polls: leading up to the 1984 election twenty-seven polls were conducted between Labor Day and Election Day, by 2000 that number had grown to 245 (Traugott, 2005), and this trend shows no sign of slowing with 48,600 polls conducted in 2016 (Toff, 2017).

[2] Where V_{share} is the vote share for one of the top two candidates competing in a simple plurality voting system.

[3] Most major election forecasters in 2016 presented their projections both ways, but win-probabilities were generally more prominent than electoral college vote share estimates.

p. 1). Indeed, the number of articles returned by Google News that mention probabilistic forecasting during the height of the campaign (from August 1 until the day before an election) has grown by orders of magnitude since 2008 (See Table 1).[4]

Table 1: Number of articles mentioning terms related to probabilistic forecasts in Google News.

Year	# of articles
2008	907
2012	3,860
2016	15,500

In some ways, the widespread success and reliance on these forecasts represents a triumph of scientific communication. In addition to greater precision compared with one-off horserace polls, probabilistic forecasts can quantify how likely a given U.S. presidential candidate is to win using polling data and complex simulation, rather than leaving the task of making sense of state and national polls to speculative commentary about "paths to victory" (Silver, 2017). Furthermore, aggregating all polls reduces the ability of news outlets to focus on unusual polls that are more sensational or support a particular narrative (Searles, Ginn, and Nickens, 2016). But despite the rigor and popularity of these aggregated probabilistic forecasts, it is not clear that the public is well-equipped to *accurately* interpret them (e.g., Gelman and Azari, 2017; Sunstein, 2002; Keren, 1991; Gigerenzer et al., 2007). As we will show in this paper, the probabilistic horserace can make one candidate's victory appear inevitable, potentially demobilizing voters.

This paper proceeds as follows: First, we discuss how probabilistic forecasts fit into existing models of perceived pivotality and turnout. Next we document the prevalence of probabilistic forecasts in national political coverage. With a survey experiment, we then show that (1) presenting win-probabilities increases the public's certainty that the leading candidate

[4] Data collected on January 30, 2018 from https://goo.gl/5JhVHU, https://goo.gl/Cj7pTo, and https://goo.gl/qpP2wa. Query terms include `fivethirtyeight OR "princeton electoral consortium" OR "nate silver" OR "probability of winning" OR "election forecast"`.

will win, compared to expected vote share; (2) the public's judgments about a candidate's likelihood of victory diverge from accurate statistical interpretation—even when presented with an actual estimate of the probability of winning. Finally, we use a behavioral game to show that (1) probabilistic estimates have substantively meaningful effects on voting; (2) reported vote share estimates have no detectable effect; and (3) these effects are not contingent on the size of the voting group. The magnitude of the effects found here, the prevalence of probabilistic forecasts, and the small margins of recent presidential elections mean that these forecasts may have an impact on prominent elections.

Problems processing probabilistic forecasts

There are good reasons to expect probabilistic horse race coverage to cause more confusion than vote share projections, despite the equivalence of the underlying information. Small differences in the election metric most familiar to the public—vote share estimates—generally correspond to very large differences in the probability of a candidate's chance of victory, and a high degree of statistical sophistication is required to map between the two. Additionally, people tend to think in qualitative terms about the likelihood of events (Sunstein, 2002; Keren, 1991); if candidate A has an 85% chance of victory, they see victory the likely outcome (this may help explain why after the 2016 election, so many criticized forecasters for "getting it wrong," Lohr and Singer, 2017; Neyfakh, 2017)). But even more generally, one-off event probabilities—candidate A has an 85% chance of winning—are often misunderstood (Gigerenzer et al., 2007) compared to statements such as "if the election was repeated 1,000 times, candidate A would win 850 times; candidate B 150 times." Finally, people may conflate probabilistic forecasts with vote share projections, and incorrectly conclude that Candidate A is projected to win 85% percent of the vote rather than to having an 85% chance of winning the election.

Probabilistic forecasts can also convey unwarranted confidence because forecasters may not account for total survey error. An estimate of the probability of victory bakes in estimates of error, which recent work has found is often about twice as large as the estimates of sampling error provided in many polls (Shirani-Mehr et al., 2016). If the forecaster does not account for *total* survey error—including errors that may be correlated across surveys (Silver, 2014)—she will artificially inflate the estimated probability of victory or defeat.

Perceptions of pivotality

Scholars have long debated the consequences of horse race election coverage. Whether conventional horserace polls distract voters from issues (Boudreau and McCubbins, 2010; Hardy and Jamieson, 2005; Iyengar, Norpoth, and Hahn, 2004; Patterson, 2005) or provide useful information about candidates (Bartels, 1988; Mutz, 1998), they undoubtedly provide information to voters about candidate's relative pubic support and the closeness of a race.

Information about the closeness of a race can give voters a sense of whether their vote might matter, which ties into longstanding theories about why people vote. Work by Downs (1957) and Riker and Ordeshook (1968) on the calculus of voting points out that the strictly "rational voter" will not vote, because the actual odds of one person's vote being decisive in an election are near zero. Riker and Ordeshook (1968) attempted to resolve this paradox in two ways: first by assuming that citizens' sense of civic duty can motivate turnout and second by noting that citizens may *perceive* that their vote can influence the outcome of an election if it is close, despite those long odds. The latter is also consistent with the decision literature, which suggests that voters will tend to *over-estimate* the odds that they might cast the pivotal vote, because of the tendency to overweight the likelihood of salient but extremely rare events in decision-making (Tversky and Kahneman 1992; Barberis 2013; Fehr-Duda and Epper 2012)

The widely used formalization in Riker and Ordeshook (1968) follows: if P is the (perceived) probability of casting the decisive vote, B is the expected benefit of winning, D is the utility of voting or sense of "civic duty," and C is the cost of voting, then one should vote if $P \times B + D > C$. Matsusaka (1995) points out that this model does not account for the *information* available to the potential voter. Rather, the expected benefit B must be conditional on the voter's confidence in her expectations of the future consequences of policies that each candidate is likely to enact. This modification to earlier models better explains empirical patterns such as the association between higher turnout and phenomena related to better information about the B term, such as education, aggregate campaign spending, and elite level issue-polarization.

Similarly, a citizen's perception of the chances of casting a pivotal vote, P, depends on the information available to voters about the state of the race, around which there may be considerable uncertainty. If potential voters perceive meaningful uncertainty around P—or that the race is close enough that their vote *could* matter—then the payoff of voting, $P \times B$ in the model above, should be non-zero. This means turnout should be (negatively) affected by the extent to which potential voters perceive information about their odds of casting a pivotal vote as conclusive. We show in Study 1 that probabilistic forecasts increase potential voters' confidence about the who will win an election, compared with conventional horse race-style projections of vote share.

Past work provides good evidence that *uncertainty* in potential voters' perception of whether their vote could matter, P, affects turnout. Some of the best evidence comes from work that analyzes the effects of releasing exit polling results before voting ends, which clearly removes uncertainty. Work examining the effects of East Coast television networks' "early calls" for one candidate or another on West Coast turnout generally find small but substantively meaningful effects, despite the fact that the these calls occur late on election day (Delli Carpini, 1984; Sudman, 1986). Similar work exploiting voting reform as a natural experiment shows a full 12 percentage point decrease in turnout in the French overseas territories that voted after exit polls were released (Morton et al., 2015). These designs also isolate the effect of information about closeness from campaigns' tendencies to invest more in campaigns in competitive districts.

Other aggregate level studies find similar patterns consistent with a relationship between uncertainty and turnout. First of all, a large literature has demonstrated robust correlations between tighter elections and higher turnout (see Geys, 2006; Cancela and Geys, 2016, for reviews). Furthermore, Nicholson and Miller (1997) provide evidence from statistical models that prior election returns also explain turnout above and beyond campaign spending, particularly when good polling data is unavailable. With ANES data we show that from 1952-2016, people who said that one candidate will "win by quite a bit" in pre-election polling were less likely to vote, even after conditioning on prior turnout, year, party, and actual electoral college and popular vote margin (see Table A2 and Figure A3).

Field experiments provide additional evidence of a causal effect of how confidence in perceptions of electoral closeness can affect turnout. This literature finds substantive effects on turnout when polling results showing a closer race are delivered via telephone (among those who were reached, Biggers et al., 2017) but null results when relying on post cards to deliver closeness messages (for which it's not possible to verify the treatment was actually read, Gerber et al., 2017; Biggers et al., 2017).[5] Finally, one study conducted in the weeks leading up to the 2012 presidential election found higher rates of self-reported, post-election turnout when delivering ostensible polling results *less* consistent with the extant polling data showing a comfortable Obama lead (Vannette and Westwood, 2014).

Thus, perceptions related to whether one's vote matters, P, should vary based on two factors:

[5]Emphasizing the closeness of an election in the context of canvassing has a large effect on turnout compared to no contact, but is not necessarily stronger than other messages crafted to mobilize voters (Gerber and Green, 2000; Dale and Strauss, 2009; Enos and Fowler, 2014). However, these studies do not directly manipulate closeness.

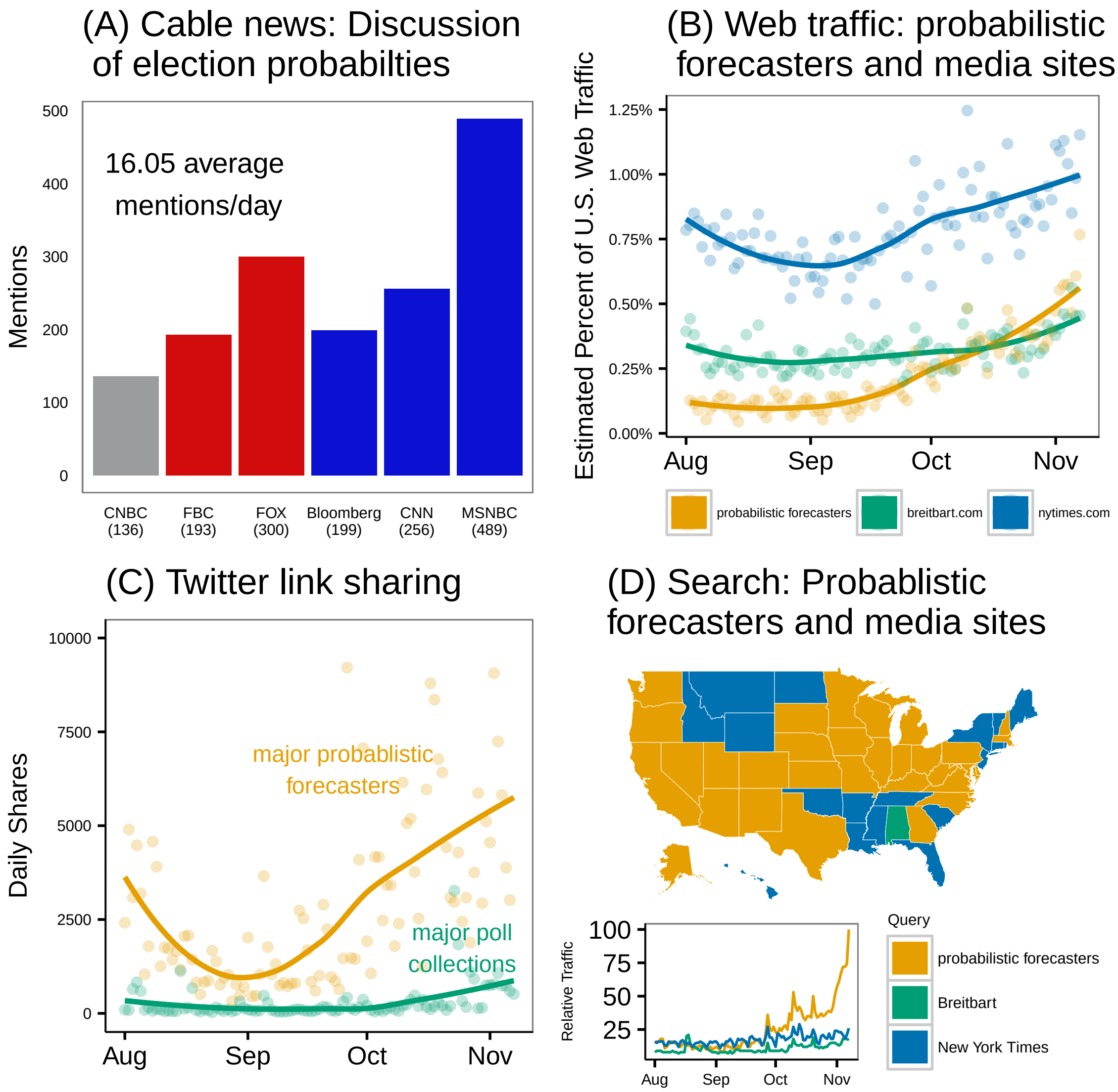
the information available to potential voters about their chances of casting the pivotal vote in an election, and their degree of certainty in that prediction. The probabilistic horserace may leave voters with considerably more confidence compared with traditional horserace coverage, particularly when news outlets tend to focus on unusual polls (Searles, Ginn, and Nickens, 2016) and offer speculative commentary about presidential candidates' potential "paths to victory," rather than a conclusive, quantitative prediction (Silver, 2017). We will show below that probabilistic forecasts lead to substantially more confident assessments of who will win an election compared with more conventional horse race-style vote share projections, even when the underlying polling data is identical. And furthermore, we will show that these assessments lower turnout in a behavioral economic game. First, however, we show that in 2016 probabilistic forecasts were widely available in a wide array of U.S. news outlets and other media.

The reach of probabilistic forecasts

Prior to showing how people respond to probabilistic forecasts, we first demonstrate that at least in the context of the 2016 presidential campaign, probabilistic forecasts were indeed widely available in U.S. news outlets and other media, and constitute an important part of the national conversation. In 2016, people visited websites offering forecasts at rates approaching major national media outlets, shared forecasts on social media at rates higher than major collections of polling data, and conducted millions of search queries to find these websites (Figure 1). Furthermore, they were covered extensively in print, with 16 thousand Google News search results in 2016,[6] and were a persistent source of discussion on cable news—particularly on channels with a more consistently liberal audience, as defined in (Mitchell et al., 2014; Bakshy, Messing, and Adamic, 2015).

[6] See Table 1, https://goo.gl/qpP2wa.

Figure 1: The reach of probabilistic forecasts in various media



Note: (A) Cable news mentioned election probabilities about 16 times each day, and did so more frequently on channels with more consistently liberal audiences. (B) Forecasting websites had more web traffic than Breitbart.com and about 50% of the traffic to NYTimes.com the day before the election (estimated percent of total U.S. web traffic each day). (C) Individuals sent tweets with links to major probabilistic forecasts a total of 281,661 times and to major collections of polls 28,416 times (tweets per day shown in plot). (D) Individuals searched Google for election forecasting sites more than they searched for two major news outlets in most states, including where the vote was close—PA, WI, and MI (states are colored such that they correspond to the largest source of searches). On the day before the election, search query traffic for Breitbart.com and NYTimes.com was 25% of election forecasting sites. All data from 8/1/2016 to 11/7/16; details on the data analyses used here are in the supporting materials.

Study 1: The perceptual consequences of probabilistic forecasts

Our first study shows that presenting the probability a candidate will win creates simultaneously more confident but less accurate beliefs about the ultimate victor than horse-race style vote share estimates. It relies on a dose-response experimental design and 4,151 respondents from wave 25 of Pew Research Center's American Trends Panel.

Participants saw a hypothetical U.S. Senate race, where "Candidate A supports the majority of the policies you support and is well qualified for the job" (implying co-partisanship) and "Candidate B does not share your views and is less qualified than Candidate A." They then read a hypothetical projection based ostensibly on recently fielded surveys analyzed by "a prominent group of statisticians." The actual projection was randomly assigned to present Candidate A's average projected vote share — $E(V_{share})$, probability of winning — $P(V_{share} > .5)$, or both (see Table 2).[7]

Table 2: Allocation of respondents to qualitative treatment cells in Study 1

		CI displayed	
Condition	Display order	No	Yes
Vote share	Only one display	813	818
P(win)	Only one display	875	0
Both	P(win) first	423	410
	Vote share first	413	399

For ease of interpretation, we pooled across two additional factors—the presence of margin of error or not, and order.[8] Among conditions displaying the vote share, a margin of error of +/-

[7] The fact that this is a Senate race might prompt questions regarding whether survey error might be substantially higher than in a typical presidential race, such that a high win probability would map to a larger difference in vote share than what we present here. Yet there is little evidence that respondents considered survey error in formulating their responses, based on the fact that presenting the margin of error alongside the vote share had no effect on any outcome.

[8] Results disaggregating these factors are consistent and are presented in Tables A7 and A8.

2% was displayed for half of participants in those conditions. Among conditions displaying both the vote share and the win-probability, half were randomly assigned to see the vote share appear first, half to see the win-probably first. Displaying the margin of error had no effect on judgments about the state of the race or certainty. Displaying the win-probability first resulted in slightly more extreme estimates of Candidate A's likelihood of victory and slightly more certainty about those judgments (See Table A8 in the Supplementary Information).

Estimates of Candidate A's vote share were randomly assigned to one of ten integer values between 45% and 55%. A plausible frequentist 95% CI of +/-2% was generated by simulating 20 surveys of 1000 people (see supplementary materials for details). Based on the same variance estimates, we also estimated the probability that Candidate A would get > 50% of the vote and win the hypothetical election, which ranged from 13% to 87% (see Table 3).[9]

Table 3: Quantitative dose-response treatments used in Study 1

$E(V_{share})$	95% CI	$P(V_{share} > .5)$
0.45	0.02	0.13
0.46	0.02	0.19
0.47	0.02	0.25
0.48	0.02	0.33
0.49	0.02	0.41
0.51	0.02	0.59
0.52	0.02	0.67
0.53	0.02	0.75
0.54	0.02	0.81
0.55	0.02	0.87

The numbers in Table 3 above rely on mapping the probability of victory to an estimate of the vote share accompanied by a 95% confidence interval, which in turn relies on the fact that both depend on the underlying distribution of a candidate's vote share. We estimated

[9]We use parameters typical of real world surveys; under certain unusual conditions that result in dramatically higher variance—for example, less than five responses per survey—win-probability numbers can be smaller than vote share estimates.

the expected vote share in our projection by the average of (hypothetical) survey sample means $\hat{\mu}_v = \frac{1}{N}\sum_i^N \bar{x}_i$ and the 95% confidence interval by $\hat{\mu}_v \pm T^{0.975}_{df=N} \times \frac{\hat{\sigma}_v}{\sqrt{N}}$, where i indexes each survey and N is the total number of surveys.[10] As is true of weighted, adjusted, and/or modeled estimate of $\hat{\mu}_v$ and $\hat{\sigma}_v$, the probability of victory, or $P(\mu_v > .5)$ can then be estimated by $1-\Phi\left(\frac{\hat{\mu}_v - .5}{\hat{\sigma}_v}\right)$. Alternatively non-parametric estimates based on simulation such as $\frac{1}{J}\sum_j^J I(\hat{\mu}_{v(j)} > .5)$ can be used to estimate the probability of victory, which is particularly useful when for example drawing J simulated electoral college outcomes. Similarly, the vote share and standard error thereof can be estimated by taking the average and standard deviation of $\hat{\mu}_{v(j)}$.

Respondents judged (1) the *share* of the vote they expected candidate A to receive, (2) how *likely* they thought candidate A was to win (both on a 0-100 point scale), and (3) how *certain* they were that candidate A would win or lose on a 5-point scale, which we transform so all values fall between 0 and 1.

Presenting aggregated survey estimates as a probability creates the impression that a candidate will win more decisively (Figure 2). Win-probabilities produced significantly more extreme judgments of vote share than when actual vote-share was displayed ($\beta = 1.49$, $T = 8.92$, $P < 2 \times 10^{-16}$, see also Table A6). For example, when presented with vote share estimates that Candidate A would win 55% of the vote, participants expected candidate A to get 56.5% of the vote (95% CI: 55.3%, 57.6%). When presented with the commensurate 87% probability that Candidate A would win, participants expected Candidate A to get 64.6% of the vote (95% CI 63.0%, 66.2%)—an 8 percentage point difference.

Effects on judgments about the likelihood of victory had similar effects, though participants gave highly inaccurate accounts of the likelihood of a candidate's victory, particularly for

[10] We do not need to employ complex modeling strategies that adjust for correlated errors or population covariates here, because we are drawing simulated data directly from a known and fully specified hypothetical population.

projections further away from a neck-and-neck race. For example, even when presented with a forecast putting Candidate A's chance of victory at 87%, the average participant said the likelihood of A's victory was 69.9% (95% CI 67.9%, 71.9%)–a 17 percentage point difference. When presented with commensurate vote share estimates that put Candidate A's share of the vote at 55%, respondents were more than 27 percentage points off—the average participant reported a 59.6% likelihood that Candidate A would win (95% CI 58.1%, 61.0%).

Respondents were also 7% more certain in their judgments about the state of the race when presented with "likely to win" compared to "vote-share" ($\beta = 0.07$, $T = 7.0$, $P = 3.7 \times 10^{-12}$, see also Table A6). In addition, around 93% of respondents reported the intent to hypothetically vote in all conditions. After conditioning on past voting history and party, we observed preliminary evidence that participants were slightly less likely to say they would hypothetically vote when seeing more extreme probabilistic forecasts (Table A9), though this finding should be considered exploratory.

Of course, it is significantly less effort to report the intent to vote in a survey than to expend the effort required to get to the polls on a Tuesday in the face of potential long lines and competing work/family obligations. Indeed, other work has found null results when examining the effect of closeness on self-reported intent to vote (Ansolabehere and Iyengar, 1994). Study 2 below attempts to better capture the economic trade offs entailed in voting, and shows that when voting presents a non-trivial cost, people do in fact vote at lower rates after viewing probabilistic forecasts that suggest that a win or loss is very likely.

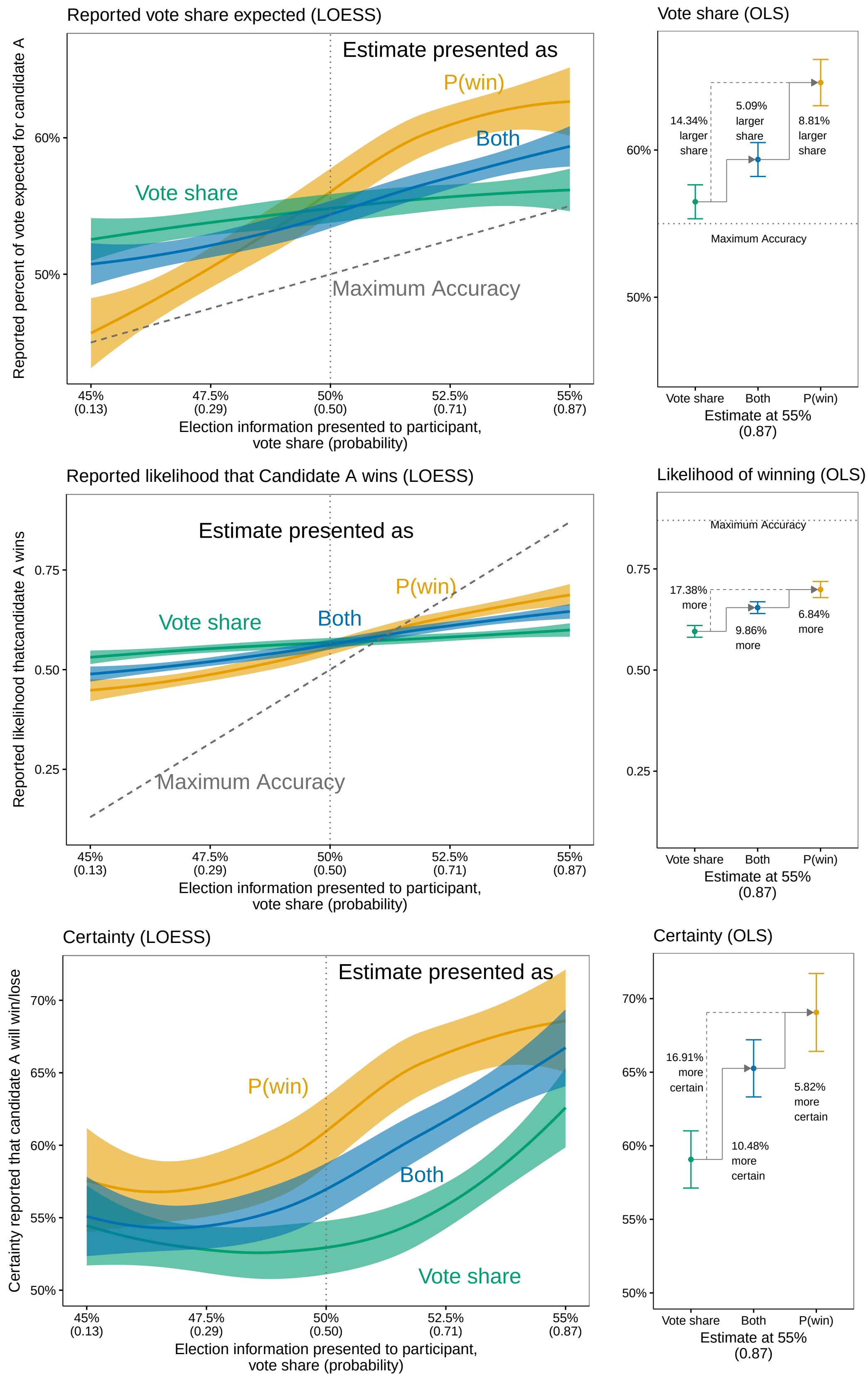


Figure 2: Effects of probabilistic forecasts on perceptions of an election

Probabilistic forecasts create the impression that the leading candidate will win more decisively, with more extreme judgments of anticipated vote share (top panel) and higher certainty in those judgments (bottom panel), even when accompanied by vote share projections. Participants are less accurate when attempting to judge the likelihood of winning (middle) than vote share (top). Plots on the right show differences when vote share is fixed at 55% (.87 probability). Lines fit using LOESS in plots on the left; results based on OLS regression in plots on the right, 95% confidence bands/intervals shown.

The strong effects documented above raise the question of how many people simply confused probabilistic estimates and vote share. This appears to be relatively rare. In the win-probability-only condition, 8.6% of respondents estimated vote share to be within 1% of the win-probability provided. In the condition that provided both win-probabilities and vote share projections, 2.1% of respondents estimated vote share to be within 1% of the provided win-probability. For comparison, in the vote share only condition wherein no win-probability number was provided, 0.6% of respondents estimated vote share to be within 1% of the equivalent win-probability number. Although some confusion between probabilities and vote share may have occurred, it did not drive these results (Table A7).

Robustness

One concern with Study 1 is that we used candidates for the U.S. Senate. It is possible that participants knew that state-level polling is nosier and less credible than the national polling used in presidential forecasts. We conducted a replication study that varied the candidate office (U.S. House/U.S. Senate/U.S. President) and found no differences between offices. We used a sample from Mechanical Turk (N=275) and a simplified design using two of the ten numerical values for probability/vote share (45% and 55%). We found no detectable differences in responses by the office sought by the hypothetical candidates (see Table A11).

Figure 2 also shows a bias in evaluating candidates that shared respondents views, consistent with a tendency toward "wishful thinking" when interpreting polling results Babad and Katz (1991); Dolan and Holbrook (2001). This wishful thinking effect attenuates significantly when presenting the win-probability for candidate B, who *doesn't* share the participant's views. In a replication of Study 1 (data from Qualtrics Panel, N=178) we varied the candidate reported to be ahead or behind. In addition to varying the candidate on which information would be provided, the numerical values were randomly varied: 41% chance of victory or 58% chance of victory. When the other candidate was reported to be ahead, respondents were less certain of victory ($\beta = -0.19$, $T = -2.661$, $P < 0.009$), reported a smaller expected vote share ($\beta = -13.20$, $T = -2.249$, $P = 0.03$), and a lower probability of victory ($\beta = -15.22$, $T = -2.30.0$, $P = 0.02$).

Study 2: How probabilistic forecasts affect behavior

In what follows, we show that when faced with the costs and benefits of voting in a behavioral game, more extreme probability estimates decrease voting. However, changes in projected vote share estimates have no detectable average effect on behavior. Our data come from 1,171 respondents (5,845 trials) drawn from a national online non-probability survey panel recruited by Qualtrics Panels.

Participants were instructed that they would ostensibly engage in a game with other participants who were completing the survey. Prior to the game, participants read instructions, reviewed examples and completed comprehension questions. Before each of five rounds, participants were randomly assigned to either Team A or Team B. At the start of the game (see Figure 3) they were given \$15. They were told that voting for their team (\$1 cost), increased the chance that their team would win. If their team won [lost], they would earn [lose] \$2. In our setup, \$1 is the cost to "vote." Before starting a round, we presented participants with a pre-vote poll, where we asked about vote intention. Participants were told this would be used along with the responses of other players and information from prior games to calculate the chance that their team would win (following the model for many forecasters). Participants were then shown the results ostensibly calculated from this poll, which included the two-team electoral vote share, randomly assigned to 40-60, and the probability that each team would win, randomly assigned to between 1-50 if a team's vote share was < 50, and 50-90 if the vote share was ≥ 50.

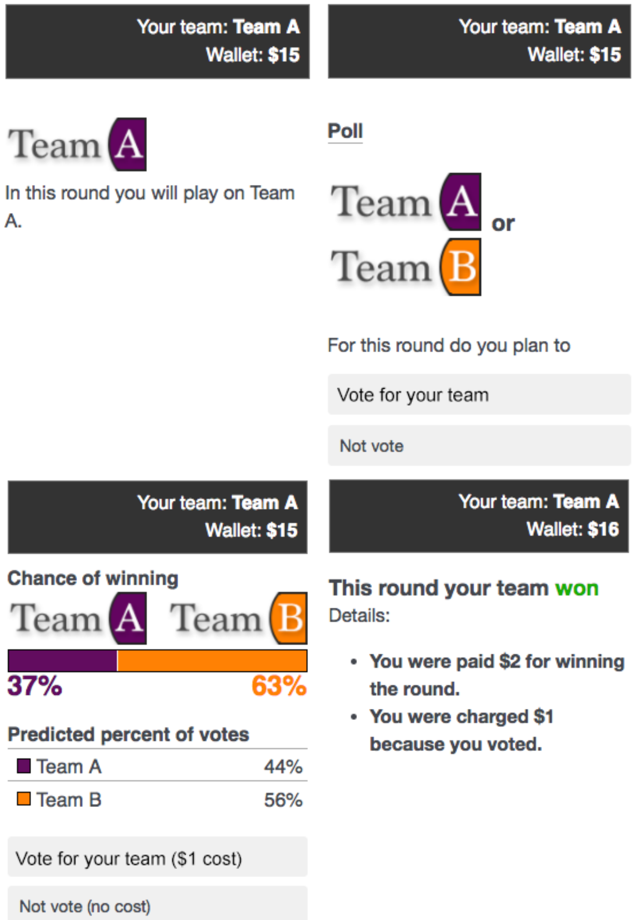


Figure 3: Study 2 stimuli and prompts

Note: Following team assignment (top left), participants took a poll (top right) about their intentions in the current round. After the system "processed" the poll, the projections were shown to participants and they were asked to decide if they wanted to actually vote for their team or abstain (bottom,left). Finally results were displayed (bottom right).

Game stages (stages 2-6 repeated in each of the five rounds):

1. Instructions, examples, and comprehension questions
2. Random assignment to Team A or Team B
3. Respondents polled on their voting intentions in the round
4. Presentation of vote share and probability of victory
5. Decision to vote ($1 cost) or not vote (no cost)

6. Feedback ($2 cost if the participant's team lost; $2 award if participant's the team won)

As the probability of winning diverged from 50-50, participants were less likely to vote ($\beta = -0.17$, $T = -4.1$, $P = 4.2 \times 10^{-5}$, see Figure).[11] However, we detected *no* effect of vote share extremity on voting ($\beta = -0.13$, $T = -0.7$, $P = 0.48$). Comparing the standardized effects of probability and vote share likewise reveals that probability has a much larger effect (see Table A18).

Given the cost of voting and payoffs, people will maximize their winnings if they only vote if $\$4 * P(decisive\_vote) > \$1$, or $P(decisive\_vote) > 0.25$, which corresponds to an extremely narrow band around even odds, and shrinks as the perceived number of players in the game increases (N). Figure clearly shows that people do not strictly maximize their winnings based on this calculus, but behave in a way more consistent with a qualitative assessment of whether their vote might matter.

[11] Data were analyzed using a multi-level model with random intercepts for each user because of repeated observations; p-values estimated based on Satterthwaite approximations to degrees of freedom. We found no evidence of directional differences, see Table A16.

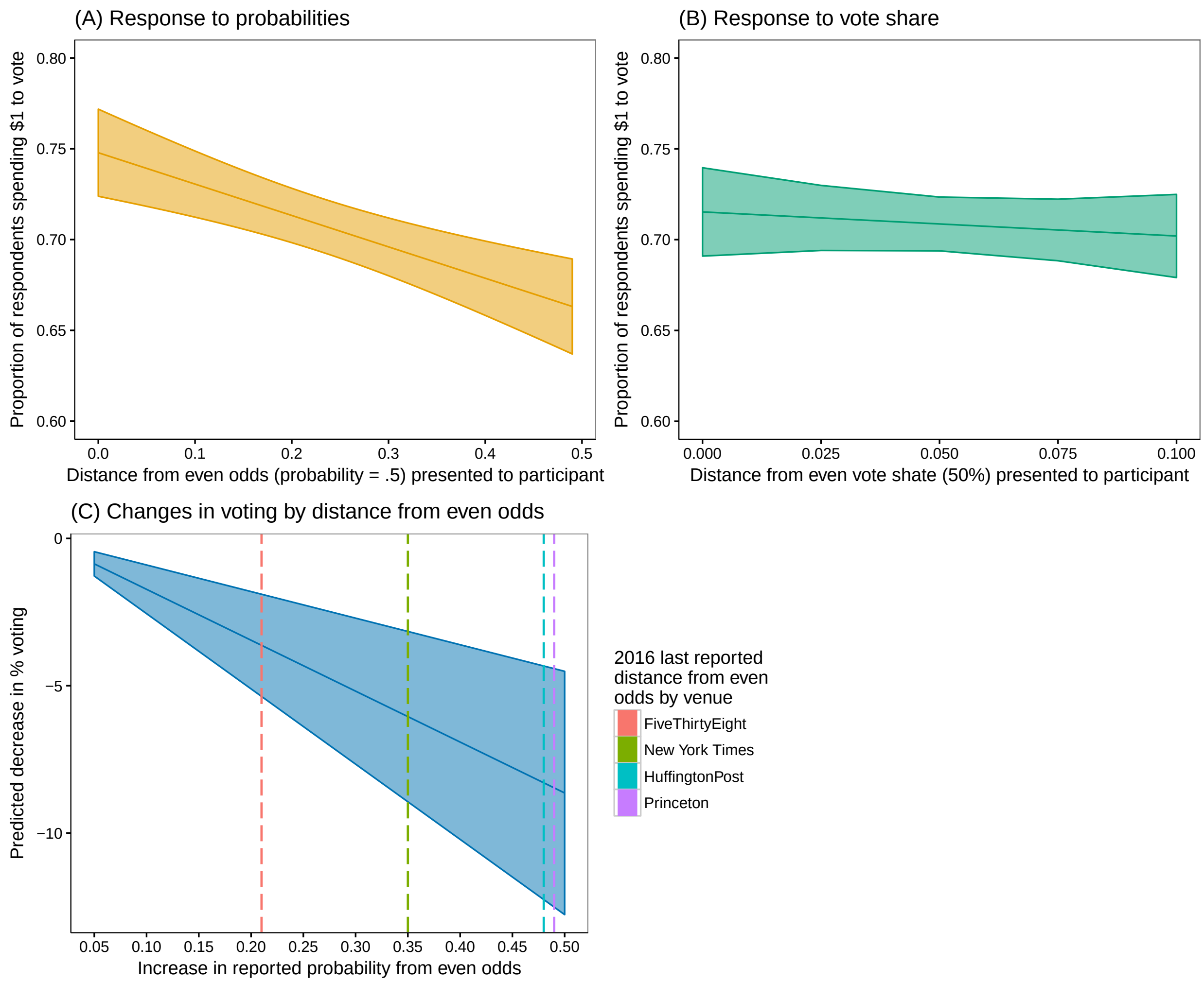


Figure 4: Effects of probabilistic forecasts on voting behavior

Note: When presented with both the probability of victory and vote share, participants are less likely to vote as the probability their team will win increases (A), but do not change behavior in response to differences in reported vote share (B). Changes in probabilities from even odds are compared with the final predicted Clinton advantage from various aggregators (C). Lines are marginal effects with 95% confidence bands.

For context, consider that 2016 forecasts reported win probabilities between 70% and 99%, giving Clinton an advantage ranging from 20% to 49%. Divergence of 20% from even odds in this game lowered voting by 3.4% (95% CI: [1.8%, 5.1%]) and divergence of 40% lowered voting by 6.9% (95% CI: [3.6%, 10.2%]). Although it is unclear how these results map to the real world, in some states the 2016 election was extremely close—Clinton ultimately lost by 0.7% in Pennsylvania, 0.2% in Michigan, 0.8% in Wisconsin, and 1.2% in Florida.

Robustness

These results are not conditional on a participants' understanding of probabilities. Following the game, participants completed the The Berlin Numeracy Test (Cokely et al., 2012), which presents respondents with a series of questions about probabilities in applied situations. We found no significant main effect of numeracy and no interaction between numeracy and forecasted probabilities displayed in the game (see table A19).

Furthermore, our results also do not depend on people confusing vote share with probabilities or incorrectly recalling either value. After each voting round, we asked respondents to report back to us the vote share and probability we supplied for the round. Of those respondents who gave a valid response, 161 respondents reported incorrect probabilities only once, while 36 made the mistake more than once. Removing respondents who made such a mistake does not substantively impact our results (Table A15).

Some people might have learned that the predictions were inaccurate (random) over time. However, with only five rounds, it would be challenging to detect that predictions were randomly assigned without conducting statistical analysis. Table A16 shows that there were no significant interactions between round and absolute distance between vote share and probabilistic forecast, suggesting that on average this did not affect our results.

Another question is whether respondents simply dismissed vote share as a noisy signal because unlike win-probability, it contains no information about precision. However, the polling data were presented as a census. And even if some respondents thought it was a sample or thought about other sources of error, such an effect would be inconsistent with the null effects of displaying measures of precision in Study 1. This presentation is also consistent with multiple probabilistic forecasting websites that present the probability of winning and vote share estimates without a margin of error.

Finally, the order of probabilities and vote share was fixed (first and second, respectively). It is possible that a primacy or recency effect could bias our attempts to compare the effects of vote share and probabilistic estimates. As part Study 3, we randomly varied the order of the information (at the participant-level). We detected no effect of order on behavior (see Table A20).

Study 3: The negligible consequences of group size in evaluating election probabilities

One way Study 2 differs from a real world election is that the number of voters is much smaller, which means that a person's vote is more likely to be pivotal. Players in Study 2 were not directly told how many other people were playing, which may have created ambiguity. In this study, we explicitly manipulated group size (N) as reported to participants in a repeated measures design, to gauge whether it attenuated the negative effect of probabilistic forecasts on turnout. We use 238 participants (1,190 trials) drawn from the Qualtrics Panel. We test this question with an interaction term between the probabilistic forecast and the group size.

In this study we told participants in the instructions: "Many other people are playing this game. Before each round you will be assigned to play with a random group of the total available players." For each round we randomly drew a value from a power of two table: 32, 64, 128, 256 and 512. To make the treatment less obvious we added random noise (drawn between [-3 and 3]) to these values for each round and for each respondent.

We do not find an effect of group N on behavior (Table A20) either in the interaction between group N and probabilities ($\beta = 0.00$, $T = 0.51$, $P = .61$) or in the interaction between group

N and vote share ($\beta = -0.00$, $T = -0.78$, $P = .44$). The inclusion of more people does not attenuate these effects—despite the lower likelihood of a pivotal vote all around, we still see similar effects to those in Study 2 above.

This suggests that people do not calculate pivotality in a manner consistent with a strict interpretation of rational choice theory. These data are more consistent with a model of the world (and much prior work Riker and Ordeshook, 1968; Keren, 1991; Sunstein, 2002) in which people have a qualitative notion of whether they might plausibly impact the vote—when the probability of a candidate's victory seems rather low or rather high, they appear less likely to vote.

Conclusion

We show that probabilistic horserace coverage confuses the public and, as odds diverge from 50-50, can have demobilizing effects compared to presenting vote share. We show using survey data that people provide highly inaccurate accounts of a candidate's likelihood of winning—even when provided with a projection of that quantity. Furthermore, presenting forecasted win-probabilities decreases the impression that an election is competitive compared to vote-share projections (Study 1). Higher win probabilities, but not vote share estimates, decrease voting in the face of the trade-offs embedded in our economic game (Study 2). This work is the first to compare how people respond to event forecasts against the quantitative projections of vote share on which they are based. These results suggest that an information environment with probabilistic forecasting can yield substantively different outcomes than one with only traditional horserace polls.

This work also shows that the aggregate effect of presenting win-probabilities on turnout may not always be symmetric across parties. The leading websites hosting probabilistic forecasts have tended to be shared more by liberal- than conservative social media users (Bakshy, Messing, and Adamic, 2015, see also Table A1) and in 2016, news outlets with more consistently liberal audiences were more likely to cover probabilistic projections. Furthermore, any effect on turnout is most likely to act on occasional voters, who tend to vote for liberal over conservative parties in the U.S. (Fowler et al., 2013). Additionally, one of our replication studies shows that focusing on the projected winner's probability of winning (e.g., Clinton's rather than Trump's in 2016), disproportionately makes supporters of the *predicted winner* overconfident in the likelihood of victory (Table A10). Finally, we use data from the American National Election Study (ANES) to show that in 2016, Democrats were more likely to say they expected one candidate to "win by quite a bit" than Republicans, and that this pattern was stronger in 2016 than in other recent elections (Figure A4).

The impression left by high-probability forecasts for a Clinton victory in 2016, which predicted the probability of a Clinton victory at between 70-99%, despite being based on electoral college vote share projections of 56-60% for Clinton (Lohr and Singer, 2017), may not have effectively conveyed the closeness the race in the electoral college to the public, potentially decreasing turnout.

The media visibility of probabilistic forecasts means their effects may not be limited to turnout—there may be additional downstream effects. Election coverage has secondary effects on donations and mobilization (Mutz, 1998), and political cynicism (Cappella and Jamieson, 1997). Furthermore, candidates' perceptions of the closeness of an election can affect campaigning and representation (Enos and Hersh, 2015; Mutz, 1997). These perceptions can also shape policy decisions—for example, prior to the 2016 election, the Obama administration's confidence in a Clinton victory was reportedly a factor in the muted response to Russian intervention in the election (Miller, Nakashima, and Entous, 2017).

These findings are not a criticism of the data underlying probabilistic forecasting, the statistical processes, or the ultimate accuracy of probabilistic forecasts. Instead, they speak to how people interpret these forecasts and behave based on those interpretations.

This work has other limitations worth noting. Study 1 relies on presenting a hypothetical election scenario, which provides a high degree of experimental control but can result in larger effects absent additional contextual information (Grimmer, Westwood, and Messing, 2014)—for example, media narratives about each candidate/campaign. The economic game presented in Study 2 tests the behavioral consequences of the economic trade-offs related to voting, but is not measuring actual voting in an actual election. Furthermore, participants may have had a heightened perception about the potential pivotality of their vote compared to a real world election due to the smaller pool of voters. However, in study 3, we explicitly manipulated group size as reported to participants, which did not attenuate the effects. Finally, studies were conducted using web-experiments embedded in surveys, thereby putting some limit on their external validity. While this design offers many benefits, such as reducing the high rate of non-compliance in field experiments and offering opportunities to investigate mechanisms, such as those in Study 3 and the Appendix, field experimental replications would lend higher ecological validity to these findings.

While forecasts are only one among many factors in play during an election, our work suggests that they can affect perceptions and ultimately outcomes, particularly in light of their media visibility in 2016. Given the far stronger effects for probabilistic forecasts compared with vote share projections, those concerned about the effects of polling—from depressing turnout (Mutz, 1998; Morton et al., 2015) to other political calculations (Enos and Hersh, 2015)—should be especially attentive to probabilistic forecasts.

References

Ansolabehere, Stephen D and Shanto Iyengar. 1994. "Of Horseshoes and Horse Races: Experimental Studies of the Impact of Poll Results on Electoral Behavior." *Political Communication* 11(4): 413–430.

Babad, Elisha and Yosi Katz. 1991. "Wishful thinkingagainst all odds." *Journal of Applied Social Psychology* 21(23): 1921–1938.

Bakshy, Eytan, Solomon Messing and Lada A Adamic. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science* 348(6239): 1130–1132.

Barberis, Nicholas. 2013. "The Psychology of Tail Events : A Note on Progress and Challenges." *American Economic Review* 103(3): 611–616.

Bartels, Larry M. 1988. *Presidential Primaries and the Dynamics of Public Choice*. Princeton, NJ: Princeton University Press.

Biggers, Daniel R, David J. Hendry, Alan S. Gerber and Gregory A. Huber. 2017. "Experimental Evidence about Whether (and Why) Electoral Closeness Affects Turnout.".

Boudreau, Cheryl and Mathew D McCubbins. 2010. "The Blind Leading the Blind: Who Gets Polling Information and Does it Improve Decisions?" *The Journal of Politics* 72(March): 513.

Cancela, Joao and Benny Geys. 2016. "Explaining voter turnout: A meta-analysis of national and subnational elections." *Electoral Studies* 42: 264–275.

Cappella, Joseph N and Kathleen Hall Jamieson. 1997. *Spiral of cynicism: The press and the public good*. Oxford University Press.

Cokely, Edward T, Mirta Galesic, Eric Schulz, Saima Ghazal and Rocio Garcia-Retamero.

2012. “Measuring risk literacy: The Berlin numeracy test.” *Judgment and Decision Making* 7(1): 25.

Dale, Allison and Aaron Strauss. 2009. “Don’t forget to vote: Text message reminders as a mobilization tool.” *American Journal of Political Science* 53(4): 787–804.

Delli Carpini, Michael X. 1984. “Scooping the voters? The consequences of the networks’ early call of the 1980 presidential race.” *The Journal of Politics* 46(3): 866–885.

Dolan, Kathleen A and Thomas M Holbrook. 2001. “Knowing versus caring: The role of affect and cognition in political perceptions.” *Political Psychology* 22(1): 27–44.

Downs, Anthony. 1957. “An Economic Theory of Political Action in a Democracy.” *The Journal of Political Economy* .

Enos, Ryan D and Anthony Fowler. 2014. “Pivotality and turnout: Evidence from a field experiment in the aftermath of a tied election.” *Political Science Research and Methods* 2(2): 309–319.

Enos, Ryan D and Eitan D Hersh. 2015. “Campaign Perceptions of Electoral Closeness: Uncertainty, Fear and Over-Confidence.” *British Journal of Political Science* pp. 1–19.

Fehr-Duda, Helga and Thomas Epper. 2012. “Probability and risk: Foundations and economic implications of probability-dependent risk preferences.” *Annu. Rev. Econ.* 4(1): 567–593.

Fowler, Anthony et al. 2013. “Electoral and Policy Consequences of Voter Turnout: Evidence from Compulsory Voting in Australia.” *Quarterly Journal of Political Science* 8(2): 159–182.

Gelman, Andrew and Julia Azari. 2017. “19 Things We Learned from the 2016 Election.” *Statistics and Public Policy* 4(1): 1–10.

Gerber, Alan, Mitchell Hoffman, John Morgan and Collin Raymond. 2017. One in a Million: Field Experiments on Perceived Closeness of the Election and Voter Turnout. Technical report National Bureau of Economic Research.

Gerber, Alan S and Donald P Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *The American Political Science Review* 94(September): 653–663.

Geys, Benny. 2006. "Explaining voter turnout: A review of aggregate-level research." *Electoral studies* 25(4): 637–663.

Gigerenzer, Gerd, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa M Schwartz and Steven Woloshin. 2007. "Helping doctors and patients make sense of health statistics." *Psychological Science in the Public Interest* 8(2): 53–96.

Grimmer, Justin, Sean J. Westwood and Solomon Messing. 2014. *The Impression of Influence: Legislator Communication, Representation, and Democratic Accountability*. Princeton University Press.

Hardy, Bruce W and Kathleen Hall Jamieson. 2005. "Can a Poll Affect Perception of Candidate Traits?" *Public Opinion Quarterly* 69(5): 725–743.

Hillygus, D. Sunshine. 2011. "The Evolution of Election Polling in the United States." *Public Opinion Quarterly* 75(December): 962–981.

Iyengar, Shanto, Helmut Norpoth and Kyu S Hahn. 2004. "Consumer Demand for Election News: the Horserace Sells." *The Journal of Politics* 66(1): 157–175.

Keren, Gideon. 1991. "Calibration and probability judgments: Conceptual and methodological issues." *Acta Psychologica* 77(3): 217–273.

Lohr, Steve and Natasha Singer. 2017. "How Data Failed Us in Calling an Election.".

Matsusaka, John G. 1995. "Explaining voter turnout patterns: An information theory." *Public choice* 84(1): 91–117.

Miller, Greg, Ellen Nakashima and Adam Entous. 2017. "Obama's secret struggle to punish Russia for Putin's election assault." *Washington Post* (June).

Mitchell, Amy, Jeffrey Gottfried, Jocelyn Kiley and Katerina Eva Matsa. 2014. Political Polarization & Media Habits. Technical report Pew Research Center's Journalism Project.

Morton, Rebecca B., Daniel Muller, Lionel Page and Benno Torgler. 2015. "Exit polls, turnout, and bandwagon voting: Evidence from a natural experiment." *European Economic Review* 77: 65 – 81.

Mutz, Diana C. 1997. "Mechanisms of Momentum: Does Thinking Make It So?" *The Journal of Politics* 59(1): 104–125.

Mutz, Diana C. 1998. *Impersonal Influence: How Perceptions of Mass Collectives Affect Political Attitudes*. Cambridge, UK: Cambridge University Press.

Neyfakh, Leon. 2017. "Nate Silver said Donald Trump had no shot. Where did he go wrong?".

Nicholson, Stephen P and Ross A Miller. 1997. "Prior beliefs and voter turnout in the 1986 and 1988 congressional elections." *Political Research Quarterly* 50(1): 199–213.

Patterson, Thomas. 2016. "News Coverage of the 2016 General Election: How the Press Failed the Voters.".

Patterson, Thomas E. 2005. "Of Polls, Mountains: U.S. Journalists and Their Use of Election Surveys." *Public Opinion Quarterly* 69(January): 716–724.

Riker, William H and Peter C Ordeshook. 1968. "A Theory of the Calculus of Voting." *American Political Science Review* 62(01): 25–42.

Searles, Kathleen, Martha Humphries Ginn and Jonathan Nickens. 2016. "For whom the Poll Airs Comparing Poll Results to Television Poll Coverage." *Public Opinion Quarterly* p. nfw031.

Shirani-Mehr, Houshmand, David Rothschild, Sharad Goel and Andrew Gelman. 2016. "Disentangling Bias and Variance in Election Polls.".

Silver, Nate. 2008. "Todays Polls and Final Election Projection: Obama 349, McCain 189.".

Silver, Nate. 2014. "The Polls Were Skewed Toward Democrats.".

Silver, Nate. 2017. "The Media Has A Probability Problem.".

Sudman, Seymour. 1986. "Do Exit Polls Influence Voting Behavior?" *Public Opinion Quarterly* 50(October): 331–339.

Sunstein, Cass R. 2002. "Probability neglect: Emotions, worst cases, and law." *The Yale Law Journal* 112(1): 61–107.

Toff, Benjamin. 2017. "The Nate Silver effecton political journalism: Gatecrashers, gatekeepers, and changing newsroom practices around coverage of public opinion polls." *Journalism* p. 1464884917731655.

Traister, Rebecca. 2017. "Hillary Clinton Is Furious. And Resigned. And Funny. And Worried." *New York Magazine* (May).

Traugott, Michael W. 2005. "The Accuracy of the National Preelection Polls in the 2004 Presidential Election." *Public Opinion Quarterly* 69(December): 642–654.

Tversky, Amos and Daniel Kahneman. 1992. "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and uncertainty* 5(4): 297–323.

Vannette, David L and Sean J Westwood. 2014. "Racing to the Polls: How Local Pre-election Polls Increase Turnout.".