Swaying the Public? Impacts of Election Forecast Visualizations on Emotion, Trust, and Intention in the 2022 U.S. Midterms

Fumeng Yang, Mandi Cai, Chloe Mortenson, Hoda Fakhari, Ayse D. Lokmanoglu, Jessica Hullman, Steven Franconeri, Nicholas Diakopoulos, Erik C. Nisbet, Matthew Kay

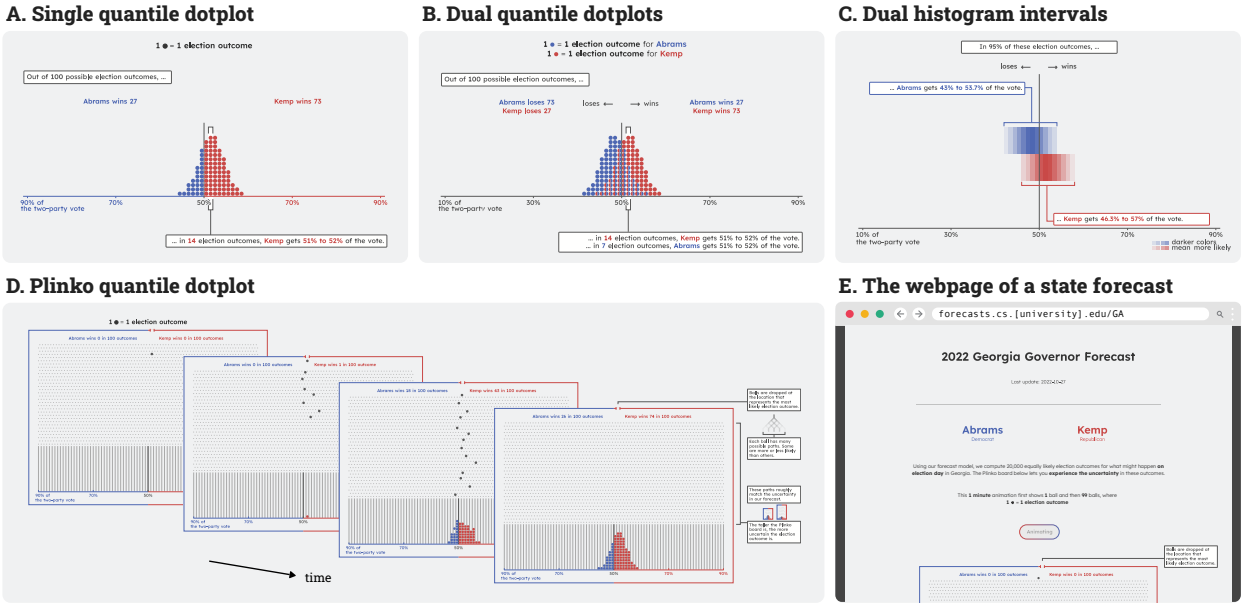


Fig. 1: **The four election forecast visualizations** in our longitudinal study: **A. Single quantile dotplot** (1-Dotplot), **B. Dual quantile dotplots** (2-Dotplot), **C. Dual histogram intervals** (2-Interval), and **D. Plinko quantile dotplot** (Plinko). In the 2022 U.S. midterm elections, we frequently updated this website to include the newest forecasts. Here all four visualizations show the forecast for Georgia on Oct. 27, 2022, which predicted a 73% probability of the Republican candidate Brian P. Kemp winning the governorship.

Abstract—We conducted a longitudinal study during the 2022 U.S. midterm elections, investigating the real-world impacts of uncertainty visualizations. Using our forecast model of the governor elections in 33 states, we created a website and deployed four uncertainty visualizations for the election forecasts: single quantile dotplot (1-Dotplot), dual quantile dotplots (2-Dotplot), dual histogram intervals (2-Interval), and Plinko quantile dotplot (Plinko), an animated design with a physical and probabilistic analogy. Our online experiment ran from Oct. 18, 2022, to Nov. 23, 2022, involving 1,327 participants from 15 states. We use Bayesian multilevel modeling and post-stratification to produce demographically-representative estimates of people's emotions, trust in forecasts, and political participation intention. We find that election forecast visualizations can heighten emotions, increase trust, and slightly affect people's intentions to participate in elections. 2-Interval shows the strongest effects across all measures; 1-Dotplot increases trust the most after elections. Both visualizations create emotional and trust gaps between different partisan identities, especially when a Republican candidate is predicted to win. Our qualitative analysis uncovers the complex political and social contexts of election forecast visualizations, showcasing that visualizations may provoke polarization. This intriguing interplay between visualization types, partisanship, and trust exemplifies the fundamental challenge of disentangling visualization from its context, underscoring a need for deeper investigation into the real-world impacts of visualizations. Our preprint and supplements are available at https://doi.org/osf.io/ajq8f.

Index Terms—Uncertainty visualization, Probabilistic forecasts, Elections, Emotions, Trust, Political participation, Longitudinal study

1 Introduction

Probabilistic election forecasts, exemplified by FiveThirtyEight's [1] and The Economist's [5] U.S. election models, provide dynamic estimates of uncertainty in electoral outcomes over time. These forecasts are typically presented using uncertainty visualizations, called election forecast visualizations. They are increasingly gaining public attention and media coverage, especially in high-profile elections, which are marked by negative campaigning, polarization, and misinformation. Residing in such an environment, election forecast visualizations may influence the general public's perception of elections and participation in the democratic process [86], making it crucial to understand their real-world impacts and consider the consequences of any design choices that could potentially alter election outcomes.

To this end, we capitalized on the opportunity of the 2022 U.S. November midterm elections[1] and conducted a longitudinal study investigating the effects of election forecast visualizations on emotions, trust in forecasts, and intention to participate in elections—all of which are potentially far-reaching impacts on the general public. We started with existing uncertainty visualizations (e.g., [31, 53, 76]) and a series of preliminary studies to inform the longitudinal study (Sec. 2). We then constructed our forecast model, built a forecasting website for the gubernatorial (governor) elections (Sec. 3), and deployed four election

The authors are with Northwestern University. E-mails: {fy, ayse.lokmanoglu, jhullman, franconeri, nad, erik.nisbet, mjskay}@northwestern.edu and {mandicai2028, chloemortenson2026, hoda}@u.northwestern.edu.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

[1] Brief background knowledge: the U.S. elections are dominated by the two-party system of Democrats and Republicans; midterm elections are held halfway through a president's four-year term and elect members of Congress and governors in various states; election day is the last day on which voters may cast a ballot. Supplements provide a glossary of election terms.

forecast visualizations (see Fig. 1). Our online experiment collected survey responses via Prolific, running from Oct. 18, 2022 to Nov. 23, 2022, involving 1,327 participants from 15 states in the U.S. (Sec. 4). As a result, this longitudinal study contributes:

- Demographically-balanced quantitative results of the impacts of the four forecast visualizations on people's emotions, trust in forecasts, and intentions for political participation (Sec. 5);
- Qualitative results elaborating why the general public does and does not trust an election forecast visualization and how they perceive forecasters' motivation (Sec. 6).

As a preview, we find that election forecast visualizations can intensify emotions, enhance trust in forecasts, and slightly increase intentions to participate in elections. The four visualizations show substantial differences: 2-Interval has the strongest impacts across nearly all measures, 1-Dotplot has the largest increase in trust after elections, and both exaggerate the differences between different partisan identities, depending on which of the two parties are predicted to win. Our experiment does not directly intervene in voting decisions. However, emotions can drive voter actions in elections [49, 81], and trust in forecasts can affect how people utilize the uncertainty information [41, 75], both hinting at the potential to change voter behavior and alter the election outcomes. Therefore, we refrain from making specific recommendations. Our findings suggest that while the uncertainty visualization literature has produced valuable recommendations for effective visual displays, those recommendations may not account for the complex social factors embedded in highly-charged political contexts with real-world implications. This necessitates further research, particularly ecologically valid studies, before we (as a field) can make confident design recommendations to the broader public.

2 Preliminaries

2.1 Related work

Studies on election forecast visualizations exist at the intersection of multiple fields, including uncertainty visualization, political communication, and journalism. Here we provide a summary, using icons to indicate the political landscape covered in each study or survey.

Uncertainty visualization Extant literature has developed various forms of uncertainty visualizations, such as summary plots (e.g., error bars [26, 53]), distributional plots (e.g., density plots [46, 53], fan charts [76]), discretized representations (e.g., quantile dotplot [31, 56], icon arrays [88]), and animations (e.g., [42, 48, 92]). These forms were assessed in tasks like improving probability perception (e.g., [56, 88]), trust in health care communication [75], transportation decision-making [31], and hurricane evacuation [65, 75]. Usually, frequency representations (e.g., a quantile dotplot) perform at least as well as probability representations (e.g., a density plot) [31, 53]. Per election forecast visualizations, data journalists have employed various forms (e.g., [1, 4]), but research in this area remains scarce [35].

Political communication Election forecasts and political polls are closely related. The former accentuates prediction uncertainty and is often based on polls, and the latter is extensively studied in political science. Election forecasts can confuse the general public and demobilize voters [86], while polls can shape both public opinion and policy [77], influence voter perception and behavior [21, 74], and affect voter turnout [15, 17], usually through perceptions of electoral competitiveness and the importance of a vote (pivotality) [23, 37, 39]. Also, there are other voter turnout theories like social pressure [38], civic duty [19], bandwagon [18] and underdog effects [17]. Partisan-motivated reasoning can be an important factor that affects the interpretation of polls, as people tend to reject polls that conflict with their pre-existing political beliefs [60, 67].

Journalism The media coverage of elections is sometimes referred to as a "horse race" due to its focus on who is ahead in the race [17]. The media may be incentivized to forecast elections for reputation and payoff [69] and to under- or over-report uncertainty for diverse reasons [47, 68]. Recently, election and COVID-19 forecasts informed predictive journalism [29, 87], focusing on conveying uncertain predictions, closely related to uncertainty visualization [35, 87].

Emotion is an important driver of political behavior [32, 51, 81,85]. Negative emotions like anger may spur actions requiring time and money (e.g., voting and donating), anxiety may drive less-costly actions like talking or information seeking, and positive emotions like enthusiasm reinforce existing behaviors [81]. Emotion is closely related to partisan identity and strength [40] and may have contributed to the polarization in the U.S. [51]. It is recognized as a design objective for communicative visualizations [63] and related to uncertainty tolerance, which can affect individuals' decision-making and well-being [41].

Trust Election forecasts can be a product of science, media, or politics. Thus, trust in election forecasts should be categorized as *institutional trust*, which is the perception that social institutions adequately perform their roles [20, 73], likely related to ideology and partisanship [71, 73]. Studies on trust in media and machine learning suggest that trust affects how people respond to the information being presented and their actions, such as consuming news [15] or following a prediction [90]. Here, an emerging topic is *trust calibration* [28, 90], which corrects undertrust and overtrust, especially after seeing an error. This is closely related to people losing trust in election forecasts due to prior failures in presidential forecasts.

2.2 Qualitative formative studies

To help delineate the problem and design space, we conduct three preliminary studies and report the key takeaways here. Readers can find a document detailing these studies in supplements.

Viewer survey Using a U.S. demographically-balanced sample (315 participants) provided by Prolific, we gathered responses from 146 participants (134 voted before) who visited election forecast websites with a graphic representation (e.g., FiveThirtyEight) in the 2016 and 2020 U.S. presidential elections. The survey data informs us that:

- Election forecast visualizations may impact affective responses. Participants recall that election forecasts invoked negative emotions (79, 59%) but could increase positive emotions (35, 26%).
- Election forecast visualizations may affect voter intentions. A small portion believes that election forecasts affect whether (79, <1%) and for whom (9, <1%) they vote, but over half of them (84, 63%) think election forecasts affect others' voting decisions.

Design space We examined the design space to explore visualization possibilities, prototyped over 40 visualizations (see supplements), selected ten representatives, and conducted a qualitative interview study.

Viewer interview In the qualitative interview study, we used a think-aloud protocol with 13 participants from the viewer survey above. The results of this study suggest that we should:

- Design salient but concise instructions or annotations [80]. Participants (12, 92%) mistakenly connect concrete visual representations (e.g., a dot) to real-world entities: a vote, a district, a poll, etc.
- Consider showing two distributions, which may better convey uncertainty. Participants (12, 92%) use visual cues like magnitude, distance, angle, shape, height, color, and numerosity, for reasoning [53]. Showing two distributions can be perceived as equal probabilities if the magnitude is the primary cue.
- Discard the most complex designs but keep some complexity. Participants (13, 100%) are averse to complexity and prefer simplicity, which also creates an illusion of certainty.

3 Forecast Website of the 2022 Governor Elections

We target governor elections in the 2022 U.S. midterm cycle. In congressional elections (i.e., the Senate or House), people usually consider it as one election and care about which party eventually has the majority of seats and controls it. Governor elections are a microcosm of a presidential election, but different states are usually independent of each other, allowing us to collect responses for various scenarios, such as different levels of uncertainty, *concordant* (when the actual winner matches the expected outcome) and *discordant* (when the expected winner is "wrong") cases. We choose to build our own website to have control over the study and visualization designs.

3.1 Forecast model

As the basis of this work, we adopt a Bayesian approach to forecasting the governor elections. Our approach is a modification of the Bayesian approach used for forecasting presidential elections [45, 64] but accounts for the specifics of 2022 and governor elections. It can be considered a Bayesian meta-analysis of polling results [79] that estimates the vote share between the Democratic and Republican parties in each state on each day. We use the polls collected and maintained by FiveThirtyEight [8] and assume no major third-party challenges [45].

We model each poll i as a Binomial sampling process, where the number of respondents indicating their support for the Democratic party is denoted by N_{DEM_i} with N_i being the total number of respondents supporting either party:

$$\begin{aligned} N_{\text{DEM}_i} &\sim \text{Binomial}(q_i, N_i) \\ \text{logit}(q_i) &\sim \theta_{\text{state}_i,\text{day}_i} \\ &\quad + \upsilon_{\text{pollster}_i} + \upsilon_{\text{method}_i} + \upsilon_{\text{population}_i} \\ &\quad + \zeta_{\text{state}_i} + \xi_{\text{poll}_i} \end{aligned}$$

The most important term is $\theta_{\text{state}_i,\text{day}_i}$, representing the underlying support for the Democratic party in state_i on day_i. The other terms represent different sources of bias: pollster effects $\upsilon_{\text{pollster}}$, polling method effects υ_{method}, polling population effects $\upsilon_{\text{population}}$, state-level error ζ, and measurement error ξ.

In presidential election forecasting, a model is usually assigned a Bayesian reverse random-walk prior [45, 64]. Given the dynamics of 2022 in the U.S. (e.g., Roe v. Wade overturned), we think a forward model is more suitable:

$$\theta_{\cdot,\text{day}_i} = \theta_{\cdot,\text{day}_i-1} + \epsilon_{\cdot,\text{day}_i}$$

where ϵ is day-to-day noise shared across states. The election outcomes are given by the predictions of the last day J (election day), and we transform them back to the linear space to get the vote share μ for the Democratic party $\mu_{\text{DEM}} = \text{logit}^{-1}(\theta_{\cdot,\text{J}})$, as well as for the Republican party $\mu_{\text{REP}} = 1 - \mu_{\text{DEM}}$. The priors of θ are previous election results, and other priors follow Heidemanns et al.'s 2020 U.S. presidential forecast [45]. Though our model cannot predict a state that does not have any polls, a number of polls are conducted in swing states, which have uncertain outcomes and are of most interest.

Because more polls are released as election day approaches, we update the forecast model throughout the experiment (see Sec. 3.4) and generate 20,000 posterior draws in each update. Each draw represents a possible election outcome. We release the R and Stan code as well as model outputs in a GitHub repository and add a link to the website; all can be found in supplements along with model alternatives.

3.2 Forecast performance

Pre-election For validity and ethical considerations, we must present reasonable forecasts and avoid creating misinformation. We compare our forecasts to FiveThirtyEight's on multiple days and provide an example comparison of the final forecasts on Nov. 8, 2022 (election day) in Fig. 2. Our forecasts agree with FiveThirtyEight's on the winners of all states except Arizona and Nevada, and the 80% predictive intervals are similar. The slight differences result in different probabilities, allowing us to be perceived as an independent forecasting website.

Post-election We also assess the forecast performance after election day with the intention to understand how it affects trust in forecasts. First, we consider the expected winners, and our final forecasts correctly predict 32 winners out of 33 states, except Nevada.[2] We then calculate the continuous ranked probability score (CRPS), a much-used scoring metric for probabilistic forecasts [91], defined as the distance between the CDF of forecast distribution and the step function of the outcome (0 = best, 1 = worst). We utilize the implementation of the R package scoringutils [52] and find that various approaches (e.g., different numbers of posterior draws) lead to similar scores. Thus, we use all 20,000 draws for scoring and obtain scores in the range of .005 to .071, with a median of .01 (e.g., Oregon) and a mean of .02 (e.g., Maryland), aligning with Fig. 2 and detailed in supplements.

[2]FiveThirtyEight correctly predicts winners in 35 of 36 states, except Arizona.

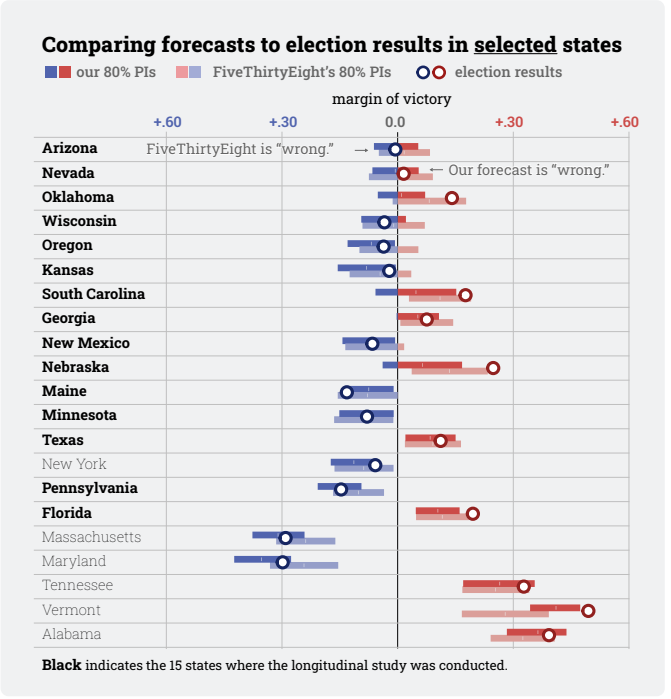


Fig. 2: **The performance of our forecasts**. We show 80% predictive intervals of vote share here, as FiveThirtyEight publicizes only means, 10% and 90% quantile points. Our forecasts are similar to FiveThirtyEight's in most states. According to The Associated Press's calls by Nov. 16, 2022 (see supplements), Nevada is the only state where our forecast does not match the election outcome (i.e., being "wrong").

3.3 Forecast visualizations

Guided by our preliminary studies (see Sec. 2.2), we decide on four uncertainty visualizations that are representatives of our prototypes without being overly intricate or potentially misleading. We strike a balance between showing one and two distributions, and select one continuous encoding and one animated design. These four visualizations cover three primary design dimensions: dimensionality, visual encoding, and animation. Each visualization conveys two quantities of our probabilistic forecasts: first, the predicted distribution(s) of two-party vote share, and second, the probability of a candidate (party) winning (or losing), which is a tail probability of the vote share distribution.

Single quantile dotplot (1-Dotplot, Fig. 1A) is a discrete outcome adaption of a probability density function [56]. It can reduce variance [56] and bias [53] in probabilistic estimates and improve everyday decision quality [31] or duration estimate [59]. FiveThirtyEight also uses a similar beeswarm plot to convey their forecasts for the 2020 U.S. general [1] and 2022 midterm [2] elections. To convey two-party vote share in one dotplot, the left half of x-axis shows the prediction of the Democratic party winning ($\mu_{\text{DEM}} > 50\%$), and the right half shows the prediction of the Republican party winning ($\mu_{\text{REP}} > 50\%$). Informed by our preliminary studies (see Sec. 2.2), we annotate three concepts:

The meaning of a dot; for example, **1 ● = 1 election outcome**, addressing the confusion that a dot may represent other real-world entities, like a district. Hovering over a dot also triggers a tooltip that explains the vote shares of that predicted outcome.

The probability of winning in frequencies [27]; for example, "**out of 100 possible election outcomes, Kemp wins 73**", which also leads to using 100 dots/posterior draws in a dotplot.

The most likely outcome and its interpretation, designed to illustrate the meaning of the height of a pile. For example, "**...in 14 election outcomes, Kemp gets 51% to 52% of the vote.**"

Dual quantile dotplots (2-Dotplot, Fig. 1B) adapt a quantile dotplot to show two distributions simultaneously. We align the baseline of the two dotplots, and use half dots ◐ when the distributions overlap, which addresses the space constraint and the occlusion issue [3]. The annota-

tions are the same as 1-Dotplot, except that they describe both parties as well as the probabilities of winning and losing. For example,

"1 ● = 1 election outcome for Abrams, 1 ● = 1 election outcome for Kemp."
"out of 100 possible election outcomes, Abrams wins 27, Kemp wins 73."
"...in 14 election outcomes, Kemp gets 51% to 52% of the vote."
"...in 7 election outcomes, Abrams gets 51% to 52% of the vote."

Dual histogram intervals (2-Interval, Fig. 1C) also display two distributions. They extend conventional interval representations [31, 53, 78], bin outcomes, and use illumination to encode probability density. They can be considered flattened histograms or discretized gradient plots [26, 46]. The Economist's 2021 German election forecasts also use a similar representation [4]. In our design, the annotations describe 95% prediction intervals and their interpretation. For example,

"...in 95% of these election outcomes, Abrams gets 43% to 53.7% of the vote, Kemp gets 46.3% to 57% of the vote."

Hovering over a bar also triggers a tooltip that explains the probability of the outcomes represented by that bar.

Plinko quantile dotplot (Plinko, Fig. 1D) is an animated visualization, which we designed to approximate the data-generating process with a physical analogy. Plinko is a popular pricing game featured on the long-running American game show "The Price is Right" [12]. The game is based on the Galton Board invented in 1889 [34], a device to demonstrate the central limit theorem.

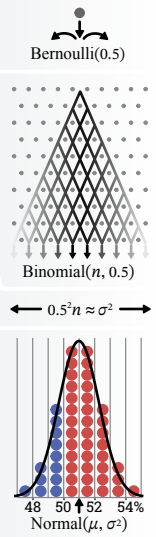
We employ the game as a physical and probabilistic analogy. The core concept is to approximate the Normal(μ, σ^2) distribution ( ) of predicted vote share using a Binomial distribution ( ) with a shifted mean. The Binomial distribution is resembled by a series of Bernoulli distributions, and each Bernoulli distribution is represented by a ball bounce on a peg. This design concept is further depicted as follows.

Each bounce has an equal chance of going left or right: a Bernoulli(q) distribution with $q = 0.5$.

A sequence of n bounces then simulates a Binomial(n, q) distribution, where $q = 0.5$ and n depends on the forecast distribution.

To derive n, we let the variance of the Binomial distribution match that of the Normal distribution (i.e., $q(1-q)n = \sigma^2$). This subsequently determines the height of the Plinko board. Intuitively, if a ball bounces through more rows of pegs (a larger n), its final location is more uncertain; thus, the **more uncertain** the forecast is, the **taller** the Plinko board is (e.g., Figs. 3D vs. 4B).

Because the mean of the Binomial distribution ($qn = 0.5n$) only matches the mean of the Normal distribution when $\mu = 50\%$, we shift it by $0.5n - \mu$. As a result, the mean and variance of the ball piles roughly match μ and σ^2, and the animation reflects a physical process generating roughly the same (quantile) forecast distribution.



We derive the ball trajectories from the possible combinations of the Bernoulli distributions, and include visual effects like acceleration, deceleration, and bouncing to reflect the physical analogy best. We fix the animation duration to be about 1 minute. In addition to the meaning of a ball, we annotate the key design concepts and animate them in dropping the first ball (see Fig. 1D and the videos in supplements):

"Balls are dropped at the location that represents the most likely election outcome. Each ball has many possible paths. Some are more or less likely than others. These paths roughly match the uncertainty in our forecast. The taller the Plinko board is, the more uncertain the election outcome is."

Lastly, we considered text representations. However, prior studies suggest that text representations may introduce larger variance [56], stronger bias [89], and worsen decision quality [31] compared to visual representations, and text presents much less information. In our preliminary studies, we also found that text elicited different mental models about election outcomes compared to *any* visual representations. Thus, we eliminate text representations for the ethical consideration that viewers should receive similar information, leaving further exploration to future work.

Post-election modifications (Fig. 3) To know how people adjust their trust in forecasts after elections, we update the visualizations to include election results and invite participants to return (see Sec. 4.3). We add the election results (actual vote shares) as one of the displayed outcomes. In 1-Dotplot and 2-Dotplot (Figs. 3AB), we use ○ ○ to indicate the actual vote shares and use ◌ ◌ when they are in a bin alone. In Plinko, we also change the coloring of the election outcome ball in the animation for comparability (see Fig. 3D). In 2-Interval, we annotate the election results onto the intervals (Fig. 3C). Additionally, we fade out the previous annotations to reduce visual clutter and add an explanation for the election result. For example,

"○ = the actual result (added after election day)."
"Abrams got 46.2% of the vote and lost, Kemp got 53.8% of the vote and won."

3.4 Forecast website

We design our website with the intention of appealing to the general public and resembling a professionally-produced forecast website. To achieve this, we also obtain design feedback from an election forecast visualization designer and ensure the website is colorblind-inclusive.

When visiting the website, a visitor lands on a page featuring a tile grid U.S. map at the top (Fig. 4A). We choose this design for its simplicity and frequent use by U.S. media outlets for navigating state-level information (e.g., [13]). We include minimal information in this map to reduce the influence on our study of visualizations while being realistic. We color-code each tile by the predicted margins (alternatively by election winners after knowing the results). Below the map, we include the following sections: "When will this website be updated?", "How do we forecast the elections?", "What data is recorded?", "About us", and "Sharing this site".

A visitor can click on a tile to enter the state forecast page (Fig. 4B). The top of the page depicts the two candidates' names. One of the four forecast visualizations is displayed as the topline (Figs. 1E and 4B), followed by a button to unfold survey questions (Fig. 4C). The page also displays "Polls used in the model", "How other days look in the model" (line graphs of vote share and the probabilities of winning over time, Fig. 4B), and the options to "Explore other states".

The website was updated roughly every other day beginning on Sept. 25, 2022 to include the newest polls and forecasts, and strictly every weekday from Oct. 17, 2022 to Nov. 8, 2022 (election day). A week later, it was updated with post-election modifications, including a headline to indicate the sources of election results. For example, "According to The Associated Press (AP), **99%** of the votes are reported. **Abrams** got **46.2%** of the two-party vote, **Kemp** got **53.8%** of the two-party vote, and **Kemp won** Georgia's governorship." All the updates and pages are responsive to different screen sizes, optimized for desktops/laptops, and tested in Chrome, Firefox, and Safari. The visualization is randomly assigned on the first visit, and stored in the browser's local storage. The website is hosted at https://forecasts.cs.northwestern.com.[3]

4 Longitudinal Experiment

Using the website, we conduct a three-wave online longitudinal experiment during the 2022 U.S. midterm elections.

4.1 Measures

Informed by related work (Sec. 2.1) and our preliminary studies (Sec. 2.2), we focus on three measures: emotions, trust in forecasts, and intention for political participation. We opt for self-reported measures, as they are commonly employed in political science (e.g., [73, 84]), easy to complete in a short amount of time, and minimally intrusive to the user experience of a forecasting website.

Emotion We select 10 emotion items from the Positive and Negative Affect Schedule (PANAS-X) [84]: angry, attentive, ashamed, confident, happy, nervous, relaxed, sad, surprised, and tired. They represent 10 emotion subscales: *hostility*, *attentiveness*, *guilt*, *self-assurance*, *joviality*, *fear*, *serenity*, *sadness*, *surprise*, and *fatigue*, respectively.

[3] As of the publication date, the website has been moved to https://forecasts.cs.northwestern.com/2022-governors-elections.

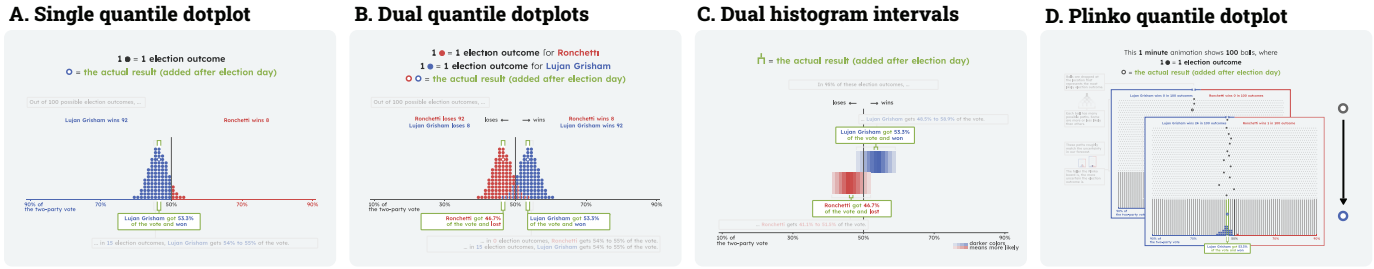


Fig. 3: **Post-election modifications to the visualizations**. (A) 1-Dotplot, (B) 2-Dotplot, and (D) Plinko use ○ ○ (but ◌ ◌ when the election result is in a bin alone); Plinko starts with ○ and changes to ○ ○ when the ball falls into a bin. (C) 2-Interval uses ⑃. The state here is New Mexico.

These 10 emotion items are most relevant to the election context, as seen in the literature and our preliminary studies (see Sec. 2.2). Each item is rated on a 7-point Likert scale ranging from not at all (1) to extremely (7). In practice, they are presented in randomized order and grouped into two groups of five (see Fig. 4C). They are later combined into positive, negative, and surprised emotions in our analysis.

Trust in forecasts We adopt the human-computer trust measure [66] to account for political [16, 73] and media [62] trust, and use two scales: *cognitive* and *affective* trust [66, 70]. Cognitive trust is the belief about the trustee's ability, measured by five items in randomized order: accuracy, fairness, reliability, trustworthiness, and understandability. Affective trust is the emotional bond, measured by faith and personal attachment. Each item is rated on a 7-point Likert scale from 1 to 7 (e.g., "This forecast is inaccurate 1 2 3 4 5 6 7 accurate") and later combined into the two trust scales. We also collect a free-text explanation for trust.

Participation intention Political participation is associated with multiple activities [82]. We evaluate two costly activities: voting and campaign contributions (e.g., "...does this forecast make you more or less likely to vote...?" and "...does this forecast make you more or less likely to contribute money or time...?"). We also elicit people's perceptions of party peers' intentions. These result in four questions, each rated on a 7-point Likert scale from less likely (1) to more likely (7). After election day, we adjust the wording to refer to participating in future elections. We collect a short free-text explanation for the ratings.

Demographics We record age, sex, education, race, and residential state in order to use post-stratification [25, 57] and approximate a representative sample over these characteristics (see Sec. 5.1 below). We elicit partisan leaning using a multiple-choice question (e.g., "Generally speaking, when it comes to political parties in the United States, how would you best describe yourself? A strong Democrat...") [44].

Other questions We ask the following open-ended questions to gain deep understandings: "What is your first impression of this forecast?", "What do you think the forecasters' intention is?", and "Is there anything you find confusing?". We also collect ideology, media trust, trust in the electoral process, trust in democracy, etc. The question wording and secondary analysis for these are provided in supplements.

4.2 Participant recruitment

Among all the states with a governor election that our forecast model could predict (e.g., those with available polling data), we choose 15 states with uncertain elections (see Fig. 2). Most of them are known as swing (or battleground) states and without major third-party challengers: Arizona, Florida, Georgia, Kansas, Maine, Minnesota, Nebraska, Nevada, New Mexico, Oklahoma, Oregon, Pennsylvania, South Carolina, Texas, and Wisconsin. This selection intends to assess how the general public experiences uncertainty, especially when forecasts do or do not match the actual outcomes (i.e., being "wrong"). As it is impossible to conduct a full pilot (i.e., until elections end) before deciding on the sample size nor to control it precisely (i.e., dropout), we decide to start by requesting 100 participants from each state.

We recruit participants from Prolific.co, and use Prolific APIs to screen them based on their profiles reported to Prolific, restricting residential states to the 15 states specified above. In states with large populations (e.g., Texas), we request 30 Democratic-affiliated, 30 Republican-affiliated, and 40 Independent participants to match the U.S. partisanship split [10]. In states with small populations (Maine, Nebraska, Nevada, and New Mexico), we remove the partisan constraint due to the limited number of available participants on Prolific. We follow Prolific's guideline on conducting longitudinal studies [6] and request participants who have at least 15 approvals.

Those profiles are not always reliable (e.g., participants may have moved to another state). We later filter out participants who do not live in the states of interest using the demographic information collected and constrain to desktop/laptop users.

4.3 Experimental design and procedure

Our goal is to measure the impacts of election forecast visualizations over time. We use a three-wave panel design and invite participants to return, as summarized in Fig. 5. Both pre- and post-election stages are of interest: the former gauges the effects when the forecasts are predicting the future in an ecologically validated environment, while the latter is essential for trust calibration once the election results are known [90]. To reduce priming and carryover effects and shorten experiment time, we collect demographics in the first wave, allowing us to screen participants and obtain "baselines" for emotions and trust.

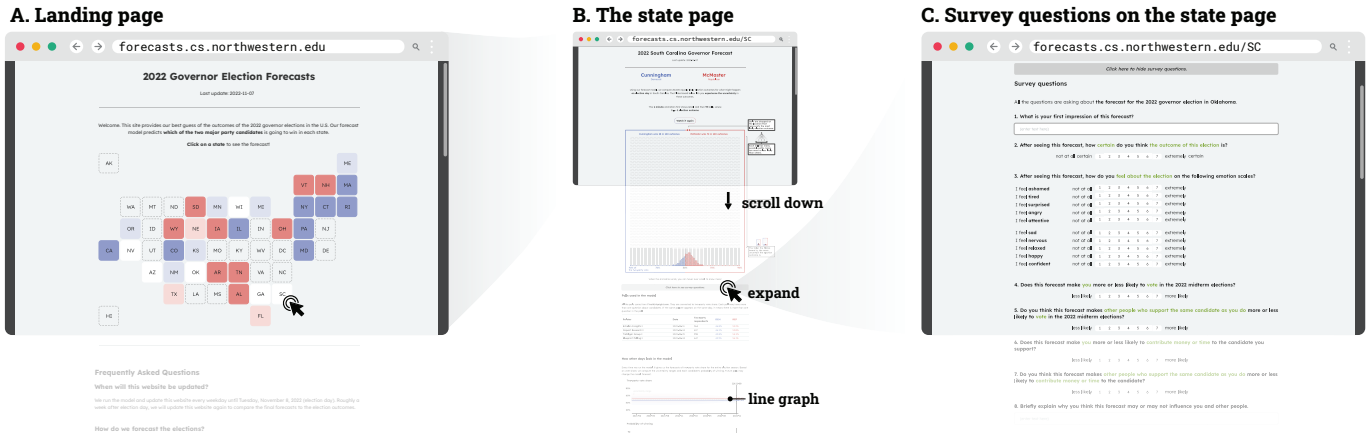


Fig. 4: **Our forecasting website** has (A) one main page with a U.S. grid map on the top. A visitor can click on a state to enter (B) the state forecast page. The visitor views the state forecast and (C) answers the survey questions.

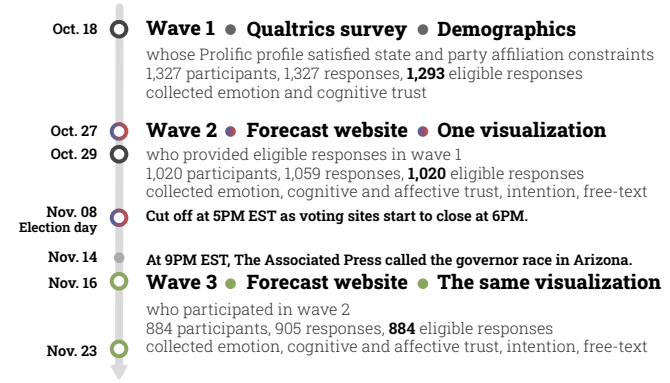


Fig. 5: **The design of the three-wave longitudinal study**, running from Oct. 18, 2022 to Nov. 23, 2022.

Wave 1 was a Qualtrics survey which collected residential state, partisanship, 10 emotion items, cognitive trust in election forecasts, demographics, and more. Affective trust and political participation intention concerned a specific forecast, and thus were measured in later waves. We started on Oct. 18, 2022, removed the partisanship constraint for the states with less than 100 responses on Oct. 27, 2022, and cut off on Oct. 29, 2022, when most states had no new responses. In total, we collected 1,327 responses from 1,327 participants; **1,293 responses**/participants came from the 15 states specified above and were eligible for the next wave.

Wave 2 invited back all eligible participants from wave 1, directed them to our forecasting website, and assigned them to one of the forecast visualizations. We stored the assignments in the browsers' local storage and on the server to impose the same visualization during later visits and waves. We explicitly instructed participants that, to receive payment, they must click on their residential state in the U.S. map (see Fig. 4A), view the governor forecast, and submit the survey on that page. We also informed participants that they were welcome to explore the website and visit it later. We started on Oct. 27, 2022 and cut off at 5PM EST, Nov. 8, 2022 (election day), because voting sites began closing at 6PM and elections would be called soon. In total, we collected 1,059 responses from 1,020 participants; **1,020 responses**/participants were eligible for the next wave.

Wave 3 invited back all eligible participants from wave 2, directed them to our forecasting website, and showed them the post-election updates using the same visualization in wave 2. Due to delayed vote counting in Arizona, The Associated Press called the winner at 9PM EST, Nov. 14, 2022. Constrained by this, we started on Nov. 16, 2022 and cut off on Nov. 23, 2022, as the U.S. Thanksgiving holiday was approaching. In total, we collected 905 responses from 884 participants; **884 responses**/participants were eligible.

Summary This study was approved by our IRB office. Each wave had a separate consent and paid each participant $1.50, $2, and $2, respectively. The median and 95% coverage of response completion time were 4.83 [2.23, 16.50], 4.84 [1.92, 14.61], and 3.67 [1.42, 11.62] minutes, respectively; Plinko usually took 1 minute longer. In wave 1, 1,121 participants (87%) reported that they registered to vote, 49 were undecided, and 123 responded "No" or others. In wave 3, 680 participants (77%) reported that they voted in the 2022 midterm elections, 184 responded "No", and 20 responded "Prefer not to say" or others. State breakdowns are available in supplements.

Pilot Before each wave, we recruited participants from Iowa, Michigan, Massachusetts, Pennsylvania, Texas, and Vermont to test the website, fine-tune the instructions, and estimate the attrition rate. We released small batches, invited participants to return in a few days, and observed that 60% to 100% of the participants would come back. Using consistent recruitment titles, pre-informing participants about the following waves, and reminding them via Prolific's message system largely improved the rate. In total, we had 115, 73, and 65 pilot participants for the three waves. The pilot data were used in pre-registration to decide the quantitative analysis models.

5 Quantitative Results

5.1 Pre-registered analyses

Modeling To estimate the interaction effects and to post-stratify data, we analyze the pilot data and decide on Bayesian multivariate multilevel ordinal regression. We choose a multivariate structure because each measure consists of multiple correlated items, and we choose ordinal regression for analyzing Likert format data, following practices in visualization [78] and psychology [22]. We estimate a separate model for emotions (10 items), cognitive trust (5 items), affective trust (2 items), and political participation intention (4 items), sharing the same pre-registered formula and priors, as we do not have separate hypotheses for each. Briefly, our models estimate the following effects: (1) interactions among visualizations, waves, candidates' probabilities of winning, partisanship, predicted winner, forecast correctness, and forecast performance (CRPS); (2) days from election day to incorporate the changes over time; (3) state, age, sex, education, and race each as random intercepts to allow post-stratification; and (4) participants as correlated random intercepts. The data, code, and model alternatives are available in supplements.

Post-stratification is a statistical technique to improve the accuracy and representativeness of survey data. It weights the estimates for each respondent by the estimated prevalence of different demographic and geographic groups [61] and is often used with multilevel modeling [36, 57]. We use post-stratification to generate **a demographically-balanced representative sample for the 15 states** (though this is an approximation). Following Kastellec et al. [54,55], we use the 5% Public Use Microdata Sample (PUMAS) from the U.S. 2020 Census (via tidycensus [83]) to compute the cross-tab percentages for all combinations of age (18 or older, voting age), sex, education, and race in each state, and calculate the weighted mean of each posterior draw in accordance with the random effects of these demographics. As such, our results are **the average effects on the 15 states**. The exception is that in wave 3, we separate the 14 states where the forecasts correctly predicted the election winners, from Nevada, the one and only discordant state, which likely drives the effects of a "wrong" forecast.

Composite measures Following guidelines on analyzing Likert format data [22, 43], we combine posterior draws of different items to get composite measures and check Cronbach's α for internal consistency [24]. Analyzing individual items can be useful, but we combine them for result interpretability and legibility. We average confident, happy, relaxed, and attentive to generate positive emotions (α: .80 [.78, .81]). We average angry, sad, nervous, ashamed, and tired to get negative emotions (α: .87 [.86, .88]), leaving surprised a separate scale. Similarly, we average accuracy, fairness, reliability, trustworthiness, and understandability to generate a cognitive trust scale (α: .91 [.90, .92]), and average faith and personal attachment to get an affective trust scale (α: .85 [.84, .87]). For participation intention, we pre-registered three combinations, but our later qualitative analysis suggests that voting and contributing to campaigns are two different activities, perceived differently. Thus, we report each item, acknowledging the reliability issue.

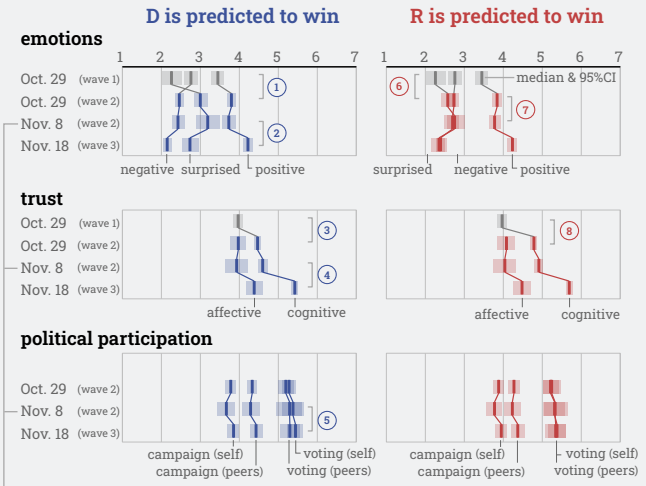
5.2 Results

In retrospect, the estimated effects are small for several model terms. Thus, we report the estimates of an average forecast (average uncertainty and median performance). We report Oct. 29 (10 days before election day), Nov. 8 (election day), and Nov. 18 (10 days from election day) as the representatives of three waves. We weight the overall estimates to roughly match the partisan leaning[4] in these 15 states and to balance Democrats and Republicans (**40% Democrats**, **40% Republicans**, 20% Independents) [11]. We first present the overall effects of election forecast visualizations over time and then break them down by visualization types and partisanship, showing medians and 95% credible intervals (CIs; Bayesian analogy to confidence intervals). No overlap with zero for the CIs after subtraction indicates substantial effects. Additional details are available in supplements, such as the results of Independents.

[4]Partisan leaning is different from party affiliation. Prolific provides the latter. However, partisan leaning seems to show stronger effects in our pilot data.

The effects of election forecast visualizations over time

Overview: average partisan leaning and visualizations (from Oct. 29, 2022 to Nov. 18, 2022)



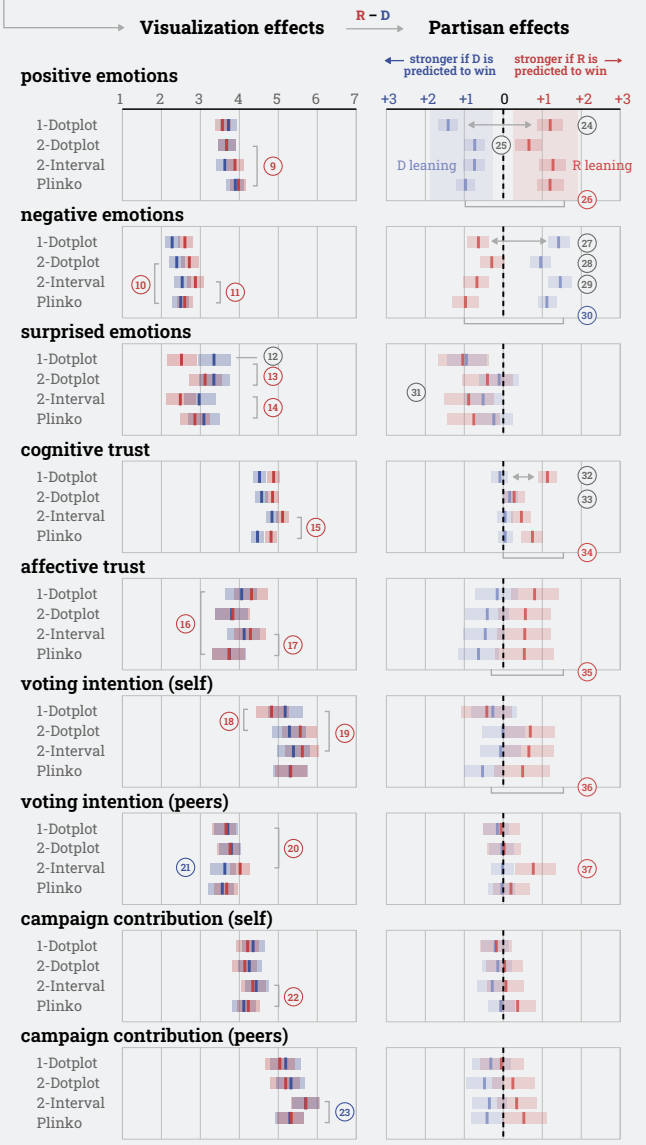
The differences are reported as median [95% CI].

In this overview, we observe the following: ① when a Democratic candidate is predicted to win (i.e., the expected winner), seeing forecast visualizations can *decrease* people's negative emotions by -0.30 [-0.45, -0.15] Likert points, *increase* positive emotions by 0.35 [0.19, 0.51], make people feel *more* surprised by 0.74 [0.44, 1.02], and ③ improve cognitive trust by 0.50 [0.38, 0.62]. ② After elections, a correct forecast can *decrease* negative emotions by -0.28 [-0.39, -0.16], *increase* positive emotions by 0.49 [0.37, 0.62], make people feel *less* surprised by -0.46 [-0.70, -0.22], and ④ *increase* both cognitive (0.82 [0.74, 0.92]) and affective (0.44 [0.26, 0.64]) trust. Also, after elections, people think that ⑤ they and their peers are slightly more likely to contribute to campaigns (e.g., 0.18 [0.026, 0.37]). The effects are similar when a Republican candidate is predicted to win, but we see ⑥ almost no change in negative emotions, smaller increases in surprised emotions (0.32 [-0.01, 0.63]), but ⑧ larger increases in cognitive trust (0.82 [0.68, 0.96] cf. ③).

In sum, election forecast visualizations can change people's emotions and increase their trust in forecasts; after elections, people also change their emotions, trust, and slightly their intentions to participate in future elections. These changes are much larger than those caused by time alone (e.g., ⑦ between Oct. 29 and Nov. 8).

The effects of visualizations and partisan leaning

Details: breakdowns by visualizations or people's partisan leaning (Nov. 8, 2022)



We dive into Nov. 8, 2022 (0 in our model) for a deeper investigation.

Visualization effects

We observe the following: **2-Interval** generally shows *the strongest effects*, especially when a Republican candidate is predicted to win. It *increases* negative emotions (e.g., ⑪ 0.28 [0.039, 0.50] Likert points more than Plinko), cognitive and affective trust (e.g., ⑮ 0.29 [0.12, 0.48] and ⑰ 0.55 [0.062, 1.07] more than Plinko), voting intention (e.g., ⑲ 0.78 [0.31, 1.30] more than 1-Dotplot), perceived peers' voting intention (e.g., ⑳ 0.66 [0.14, 1.07] more than 1-Dotplot), intention for campaign contribution (e.g., ㉒ 0.35 [0.059, 0.68] more than Plinko), and perceived peers' intention for campaign contribution (e.g., ㉓ 0.32 [0.00, 0.66] more than Plinko), but **2-Interval** results in the *least* surprised emotions (e.g., ⑭ -0.36 [-0.81, 0.08] less than Plinko). **1-Dotplot** makes people feel ⑫ the *most* surprised when a Democratic candidate is predicted to win, but the *least* surprised when a Republican candidate is predicted to win; it also gains people's affective trust (e.g., ⑯ 0.58 [0.010, 1.09] more than Plinko), but is perceived as the least likely to increase voting intention (e.g., ⑱ -0.73 [-1.21, -0.23] less than 2-Dotplot). 1-Dotplot also increases cognitive trust the most in wave 3 (see supplements). **Plinko** makes people feel the *most* positive (e.g., ⑨ 0.31 [0.051, 0.54] more than 2-Dotplot), the *least* negative when a Republican candidate is predicted to win (e.g., ⑩ -0.28 [-0.50, -0.039] less than 2-Dotplot). In sum, the differences in visualizations are small but substantial, and they interact with the effects of which party is predicted to win.

Partisan effects

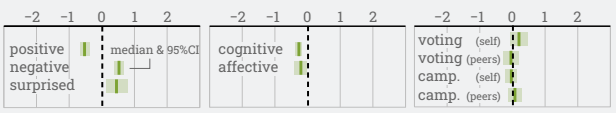
Republicans ㉖ feel more positively and ㉞ ㉟ trust the forecasts more when their candidate is predicted to win, compared to how Democrats feel when a Democratic candidate is predicted to win. That said, Republicans' cognitive trust is affected by which party is predicted to win. Democrats ㉚ feel more negatively about the opposite party predicted to win than Republicans do. ㉛ Both Democrats and Republicans feel more surprised when a Democratic candidate is predicted to win than when a Republican candidate is predicted to win. In sum, both Democrats and Republicans are biased when the forecast predicts a Republican candidate is going to win, but in different ways.

Interactions between visualization and partisanship

1-Dotplot creates *large differences* between the two parties in ㉔ positive emotions, ㉗ negative emotions, and ㉜ cognitive trust. **2-Interval** shows similar effects and makes Republicans think ㊱ their peers are more likely to vote for their party winning than what it does to Democrats. **2-Dotplot** appears to mitigate these partisan differences in ㉕ positive and ㉘ negative emotions, and ㉝ cognitive trust.

The effects of a "wrong" forecast in Nevada

Average partisan leaning and visualizations (Nov. 18, 2022)



Because Nevada is the only discordant state, we report it separately, post-stratifying posteriors to match the state demographics and comparing it to a hypothetically "correct" forecast for Nevada. The "wrong" forecast *decreases* positive emotions by -0.54 [-0.70, -0.39] Likert points, *increases* negative emotions by 0.51 [0.38, 0.65], and makes people feel *more* surprised by 0.43 [0.11, 0.77]. It slightly decreases both cognitive (-0.28 [-0.38, -0.18]) and affective (-0.22 [-0.41, -0.03]) trust, making people think they are more likely to vote (0.19 [-0.09, 0.46]) in the future. We do not find any conclusive partisanship or visualization effects here.

Discussion

The differences in Democrats' and Republicans' feelings about the two parties are likely caused by affective polarization [49, 81] and distinct voting cultures [33]. People identifying as Republicans view co-partisans positively, and people identifying as Democrats view opposing partisans negatively [50]. Also, Republicans' trust seems more influenced by visualization types, particularly 1-Dotplot. The observations that people are more surprised when a Democratic candidate is predicted to win *might* be related to their expectations for the 2022 midterm elections. Similarly, if the differences in which of the two parties are predicted to win can be further confirmed in future studies, we *might* need to consider asking people to register their favored party before showing them election visualizations. Factors like this are rooted deeply in U.S. society, inadvertently contributing to polarization and reinforcing partisan-motivated reasoning. This challenges the objective interpretation of election forecast visualizations, as further elaborated in our qualitative results below.

6 Qualitative Results

6.1 Analysis

We analyze the free-text responses to gain deeper insights into how people perceive election forecast visualizations. We do not distinguish whether participants respond to their states and analyze all 8,010 responses (≈115,000 words), consisting of 1,059 responses for each open-ended question in wave 2 and 905 for each in wave 3 (see Sec. 4.3; all are required). One coder started with open coding for each question, grouped the codes into axes, merged similar axes [58], and verified the code assignments, which were not mutually exclusive. As model-based post-stratification is not feasible here, the results may be skewed towards Democrats and the states with larger populations. We report relevant codes below and provide our codebook in supplements.

6.2 Results

Trust and distrust in election forecast visualizations

In wave 2, we have responses expressing trust (440, 42%), distrust (289, 27%), hesitation to judge (115, 11%), and mixed feelings (181, 17%). In wave 3, we have responses describing trust (609, 67%), distrust (135, 15%), hesitation to judge (37, 4%), and mixed feelings (95, 11%).

Pre-election trust (wave 2) The main reason for trusting a forecast visualization is its alignment with pre-existing knowledge or beliefs about election outcomes (263, 25%). Participants may not articulate a specific reason but reveal a high propensity to trust (141, 13%) [72] (e.g., *"I assume its accurate"*) or find visualizations understandable (132, 12%). Other reasons to trust forecasts include using and aggregating polls (88, 8%), trustworthy sources (40, 4%), employing scientific methods (60, 6%), and considering uncertainty (24, 2%).

Post-election trust (wave 3) All of the above reasons reappear but diminish. The primary reason becomes the forecast matching election results (343, 38%) and aligning with their expectations or pre-existing knowledge (67, 7%). Participants may focus on the correct winner (62, 7%), the election result falling within the predicted range (67, 7%), or both (6, <1%). Some people assess the most or second most likely outcome (15, 2%), the probability of winning (8, <1%), and whether the forecast contains the election outcome (7, <1%).

Pre-election distrust (wave 2) The main causes of distrust include distrust in polls or poll sources (83, 8%), low trust propensity (73, 7%) (e.g., "*I don't put too much faith in forecasts*"), forecast complexity or ambiguity (72, 7%), and disagreement with prior knowledge (71, 7%). Participants may not trust the forecast source or overlook poll sources (53, 5%) and feel a lack of method explanation or transparency (51, 5%).[5] They may deem forecasting impossible (43, 4%), recall prior failures (27, 3%) or be concerned with uncertainty (27, 3%).

Post-election distrust (wave 3) All the above reasons for distrust reoccur, with visual complexity or ambiguity becoming the main (69, 8%). Participants distrust forecasts due to wide outcome ranges (19, 2%), deviation from election results (20, 2%), or the need for more elections to build trust (15, 2%). Few consider forecasts incorrect (9, 2%) but it may be easy to predict a red or blue state (14, 2%).

Discussion Trust in forecasts appears to be heavily influenced by prior beliefs. When a forecast aligns with people's pre-existing knowledge and accurately predicts an outcome, it can improve their trust and alleviate much of the distrust, with specific design choices being a small portion, as observed for trust in machine learning [90]. One opportunity here might be to show multiple forecasts [75], but this must be done carefully, considering the specific context and visual complexity. Another opportunity may lie in designing post-election visualizations to explain how the outcome relates to the forecast. This communication of model calibration in comparison with forecasts is not typical in uncertainty visualization (also see [7, 9]), but it may foster institutional trust over time by appropriately acknowledging uncertainties.

Forecast visualizations & participation intentions

In wave 2, many participants think they would vote regardless or decide based on other factors (302, 29%), and some contribute time and funds irrespectively (33, 3%). Some do not care about election forecasts (72, 7%) or feel the forecasts merely confirm prior knowledge (33, 3%).

Participants think the visualizations remind people to vote (123, 12%), with uncertainty causing insecurity (112, 11%). They think pivotal votes in close elections can motivate voting (191, 18%) [30], while decisiveness can suppress it (102, 10%) [19]. Seeing their candidates winning may suppress (100, 9%) or encourage (74, 7%) voting, sometimes due to bandwagon effects (18, 2%) [18]. Seeing their candidate losing may suppress voting (61, 6%) or cause an underdog effect to encourage voting (92, 9%) [17]. Considering the consequences of the opposite party winning may encourage voting (32, 3%). Emotions also play a role (109, 10%), with both positive and negative emotions potentially encouraging (18, 2%; 22, 2%) or suppressing voting (17, 2%; 20, 2%).

In wave 3, all the above codes recur, and a similar portion (279, 31%) insist forecast visualizations do not affect voting intentions. However, positive (118, 13%) and negative (213, 24%) emotions become the main drivers. Positive emotions may motivate future voting due to winning reinforcement (57, 6%) or showing a payoff for voters' efforts and underscoring the importance of voting (59, 7%). Losing the election with a narrower margin (79, 9%) can inspire hope for future wins and motivates voting (26, 3%) [32]. Conversely, negative emotions like loss, regret, and anger motivate voting (99, 11%) [81], while sadness, tiredness, hopelessness, and disappointment discourage it (34, 4%) [85]. Also, winning a large margin (54, 6%) and accurate forecasts (6, <1%) may suppress people's intentions to vote in the future [19, 86].

[5] Our partial mouse movement logs suggest that >60% of participants might not have scrolled down to "How do we forecast the elections?" and "About us" on the landing page (Fig. 4A), and >10% of participants might not have scrolled down to see polls on state pages (Fig. 4B). Even if they did, they could distrust out of unfamiliarity as we had no record of forecasting U.S. elections.

Discussion These responses suggest a complex relationship between viewing forecast visualizations and voting intentions, which extends beyond what is currently observed in political science literature. Some perceptions appear to be a result of understanding the uncertainty, but more are closely related to the political context and media environment. This raises questions about the purpose of election forecast visualizations. Are they solely intended to convey an accurate prediction with well-calibrated uncertainty, or do they have other communicative goals [63] that could lead to under- or over-reporting uncertainty [35, 68] such as avoiding being "wrong" or raising voter awareness? If these motivations exist, how do viewers perceive them? We delve into these questions below.

Perceptions of forecasters' motivation

In wave 2, most participants consider the forecast website an information source (1012, 95%), with some considering it a decision aid for informed voting (17, 2%). They perceive the forecast website as unbiased, accurate, or realistic (184, 17%), providing election information (150, 14%), or compiling information (68, 6%).

Participants are aware this is a research study (129, 12%), thinking the website is collecting data for a [non-]profit institution (58, 5%) or conducting an observational study (60, 6%) without intervening in their decisions. However, a small fraction thinks the forecast visualization is fake or feels manipulated (7, <1%). Some think the website can have broader impacts (234, 22%) or express concerns about the impacts (30, 3%). Positively, participants think the website targets voters (53, 5%) and can influence voting (without propaganda) (97, 9%), encourage voting (94, 9%), educate people (59, 6%), and raise voters' awareness (26, 2%).

A substantial fraction thinks the motivation behind the website is immoral (103, 10%). It may have a political agenda, made to influence voting or persuade people to vote for a particular party (46, 4%), get profit/traffic/newsworthiness (22, 2%), suppress voting (14, 1%), show support for a party (6, <1%), be biased (13, 1%) or manipulative (7, 1%).

Discussion Recall that we attempted to resemble a professionally-produced forecast website and "accurately" visualize the uncertainty. However, it becomes apparent that many people interpret our intentions in ways that go beyond these goals, potentially reflecting back on their reaction to the forecast visualizations presented. While these misperceptions are understandable in a political context, they also highlight the importance of examining how forecast visualizations are interpreted by participants to seek solutions for misperceptions. To understand these, we analyze the qualitative differences in visualizations and their potential impacts on viewers' perceptions below.

Qualitative differences in visualizations

Participants' impression 2-Dotplot and 2-Interval receive fewer positive impressions than 1-Dotplot and Plinko (e.g., 23/259 cf. 64/278), but the former two have more participants describing takeaways from the forecasts (e.g., 97/259 cf. 66/264) and referring to prior knowledge (e.g., 56/259 cf. 37/264). 2-Dotplot seems the most confusing, with the fewest responses about clarity (173/259 cf. 220/278 to 214/278 in wave 2; 181/226 cf. 190/228 to 204/224 in wave 3).

Trust 1-Dotplot is less associated with trust propensity (e.g., 25/264 cf. 41/258) but more with prior knowledge or belief (e.g., 84/264 cf. 54/278 in wave 2; 21/228 cf. 6/226 in wave 3). 1-Dotplot and 2-Dotplot are less associated with forecast accuracy (e.g., 70/228 cf. 84/224). Plinko renders a sense of being unscientific, random, or irrelevant to elections (32/278 in wave 2; 13/224 in wave 3), which appears to undermine trust.

Participation intention 2-Dotplot is perceived as a reminder to vote more than others (40/259 cf. 26/278 to 25/258). 2-Interval has fewer underdog (15/258 cf. 26/278 to 27/264) but more pivotality responses (76/258 cf. 30/264 to 52/259). 1-Dotplot and Plinko have more responses of voting regardless (54/258 to 62/278 cf. 41/259 to 42/264).

Forecasters' motivation Plinko receives more responses about entertainment (e.g., 18/278 cf. 0) and uncertainty (e.g., 26/278 cf. 4/264 to 11/258) but fewer about forecast accuracy (24/278 cf. 43/259 to 60/264). 2-Interval has the most responses about informing the public (166/258 cf. 144/278 to 151/264). 1-Dotplot (33/264) and 2-Dotplot (31/259) have more responses about immoral purposes (cf. 14/278 to 19/258).

Discussion The four visualization designs generate qualitative differences in how people relate the information to prior beliefs. Factors such as the clarity of a design, the presence of two distributions, and even confusion can all be interpreted as part of propaganda [63]. While we do not systematically assess design variables in this study, it is suggested that showing one or two distributions might be interpreted differently, with two distributions potentially being perceived as indicating a closer election and therefore encouraging people to vote. Interestingly, confusion caused by 2-Dotplot seems to lead to smaller differences among different party identities in our quantitative results. In contrast, a comprehensible visualization with a concrete representation, such as 1-Dotplot, appears to worsen polarization by causing people to connect the information more strongly to their prior beliefs.

7 General Discussion and Conclusion

We contribute empirical knowledge on election forecast visualizations, spanning the fields of uncertainty visualization, political communication, journalism, affect, and trust. We recognize that readers might feel overwhelmed at this point, as are we. While the extant literature in uncertainty visualization inspired this study and guided us throughout the visualization design and longitudinal experiment, it appears insufficient to explain everything we observed. The real-world environment is much more complex than a typical controlled experiment, generating results that we normally could not observe, and a single study is insufficient to answer all questions about election forecast visualizations. One highlight is that the interactions with partisanship seem to violate our assumptions about how people interpret visualizations. We do not know if these findings are generalizable to other U.S. elections. Given such complexity, we deliberately chose not to offer explicit recommendations, with the aim to provoke thought and discussion.

There are other imperfections in this study, some of which are related to ethics. For example, we had hoped that more of our "close" forecasts would go one way or the other, resulting in several states with "wrong" forecasts to compare against those with "correct" forecasts, but this happened in only one state. In the future, should we instead try to adjust forecasts to make it more likely get different outcomes to happen in experiments (e.g., for forecasts close to 50%, we could nudge it to the other side of 50% for half of the participants)? Would it be ethical? We designed visualizations and annotations to the best of our ability within the real-world timeline, and calibrated the model to the best of our ability, although we may not have collected the most useful data if we had known the future. Also, the post-stratified estimates may not perfectly represent the U.S. general public, as selection, attrition, and response biases likely occur. In fact, our sample appears to be far more politically active than the general turn-out in the 2022 U.S. midterm elections (77% vs. 46%) [14]. Our study is tailored to the U.S. elections, but different countries have different electorate systems and political environments, where our findings may not generalize. In the U.S., governor elections typically receive less attention than congressional races, particularly presidential elections, which may lead to different results. We hope that our findings and insights can lay the groundwork for similar studies during the 2024 U.S. general election.

Highlighted findings

- Election forecast visualizations can change emotions and enhance people's trust in forecasts, slightly affecting perceptions and intentions for participating in elections;
- The differences in forecast visualizations are small but substantial, and they interact with the effects of the predicted winning party;
- 2-Interval generally has the strongest effects, especially when a Republican candidate is predicted to win, and 1-Dotplot can create large differences between individuals of different party identities;
- When a forecast aligns with a person's prior belief and accurately predicts an outcome, it can largely improve trust and alleviate much of the distrust;
- There is a misalignment between forecasters' (designers') motivations and viewers' perceptions of those motivations, with forecasts and design choices potentially being linked to prior beliefs and interpreted through a political lens.

ACKNOWLEDGMENTS

This research is supported by NSF IIS-2107490, NSF IIS-1901485, NSF IIS-2126598, and NSF 2127309 to the Computing Research Association for the CIFellows Project. The authors thank the following people for their feedback on this research: Yuan (Charles) Cui, Lily W. Ge, Ziyang Guo, Lane Harrison, Maryam Hedayati, Hyeok Kim, Taewook Kim, David H. Laidlaw, Priyanka Nanayakkara, Abhraneel Sarma, Evan Peck, Dongping Zhang, Kaiyu Zheng, and especially thank Anna Wiederkehr. The authors also thank anonymous reviewers for their valuable comments. The country flags are from Twitter, and the browser svg is credited to Vecteezy.

SUPPLEMENTAL MATERIALS

The authors provide the following materials at https://doi.org/osf.io/ajq8f: (1) two videos of the website, (2) three documents for survey questions, preliminary studies, and terms, (3) the forecast model code and alternatives, (4) pre-registration, (5) code, data, and models for quantitative analysis, (6) secondary analysis, and (7) codebook for qualitative analysis.

REFERENCES

[1] 2020 Election Forecast. https://projects.fivethirtyeight.com/2020-election-forecast. 1, 2, 3

[2] 2022 Election Forecast. https://projects.fivethirtyeight.com/2022-election-forecast. 3

[3] Attacking discrimination with smarter machine learning. https://research.google.com/bigpicture/attacking-discrimination-in-ml/. 3

[4] The Economist's German election model forecasts three possible governing coalitions. https://www.economistgroup.com/group-news/the-economist/the-economists-german-election-model-forecasts-three-possible-governing. 2, 4

[5] Forecasting the U.S. elections. https://projects.economist.com/us-2020-forecast/president. 1

[6] How do I set up a longitudinal / multi-part study? https://researcher-help.prolific.co/hc/en-gb/articles/360009222733-How-do-I-set-up-a-longitudinal-multi-part-study-. 5

[7] How good are FiveThirtyEight forecasts? https://projects.fivethirtyeight.com/checking-our-work. 8

[8] Latest polls. https://projects.fivethirtyeight.com/polls. 3

[9] Metaculus track record. https://www.metaculus.com/questions/track-record/. 8

[10] Party affiliation. https://news.gallup.com/poll/15370/party-affiliation.aspx. 5

[11] Party affiliation by state. https://www.pewresearch.org/religion/religious-landscape-study/compare/party-affiliation/by/state. 6

[12] The Price is Right. https://priceisright.com/games. 4

[13] What you need to know about the measles outbreak. https://www.washingtonpost.com/graphics/health/how-fast-does-measles-spread. 4

[14] Where voter turnout exceeded 2018 highs. https://www.washingtonpost.com/politics/interactive/2022/voter-turnout-2022-by-state. 9

[15] M. Agranov, J. K. Goeree, J. Romero, and L. Yariv. What Makes Voters Turn Out: The Effects of Polls and Beliefs. *Journal of the European Economic Association*, 16(3):825–856, 2017. doi: 10.1093/jeea/jvx023 2

[16] C. J. Anderson and A. J. LoTempio. Winning, losing and political trust in america. *British Journal of Political Science*, 32(2):335351, 2002. doi: 10.1017/S0007123402000133 5

[17] S. Ansolabehere and S. Iyengar. Of horseshoes and horse races: Experimental studies of the impact of poll results on electoral behavior. *Political Communication*, 11(4):413–430, 1994. doi: 10.1080/10584609.1994.9963048 2, 8

[18] M. Barnfield. Think twice before jumping on the bandwagon: Clarifying concepts in research on the bandwagon effect. *Political Studies Review*, 18(4):553–574, 2020. doi: 10.1177/1478929919870691 2, 8

[19] A. Blais. *To Vote or Not to Vote?: The Merits and Limits of Rational Choice Theory*. University of Pittsburgh Pre, 2000. 2, 8

[20] B. T. Blankenship and A. J. Stewart. Threat, trust, and trump: identity and voting in the 2016 presidential election. *Politics, Groups, and Identities*, 7(3):724–736, 2019. doi: 10.1080/21565503.2019.1633932 2

[21] C. Boudreau and M. D. McCubbins. The blind leading the blind: Who gets polling information and does it improve decisions? *The Journal of Politics*, 72(2):513–527, 2010. doi: 10.1017/S0022381609990946 2

[22] J. D. Brown. Likert items and scales of measurement. *Statistics*, 15(1):10–14, 2011. 6

[23] J. Cancela and B. Geys. Explaining voter turnout: A meta-analysis of national and subnational elections. *Electoral Studies*, 42:264–275, 2016. doi: 10.1016/j.electstud.2016.03.005 2

[24] A. Christmann and S. Van Aelst. Robust estimation of cronbach's alpha. *Journal of Multivariate Analysis*, 97(7):1660–1674, 2006. doi: 10.1016/j.jmva.2005.05.012 6

[25] M. Christofoletti, T. R. B. Benedetti, F. G. Mendes, and H. M. Carvalho. Using multilevel regression and poststratification to estimate physical activity levels from health surveys. *International Journal of Environmental Research and Public Health*, 18(14), 2021. doi: 10.3390/ijerph18147477 5

[26] M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE TVCG*, 20(12):2142–2151, 2014. doi: 10.1109/TVCG.2014.2346298 2, 4

[27] L. Cosmides and J. Tooby. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58(1):1–73, 1996. doi: 10.1016/0010-0277(95)00664-8 3

[28] E. J. de Visser, M. M. M. Peeters, M. F. Jung, S. Kohn, T. H. Shaw, R. Pak, and M. A. Neerincx. Towards a theory of longitudinal trust calibration in human–robot teams. *International Journal of Social Robotics*, 12(2):459–478, 2020. doi: 10.1007/s12369-019-00596-x 2

[29] N. Diakopoulos. Predictive journalism: On the role of computational prospection in news media. *Tow Center for Digital Journalism*, 2022. 2

[30] J. Duffy and M. Tavits. Beliefs and voting decisions: A test of the pivotal voter model. *American Journal of Political Science*, 52(3):603–618, 2008. doi: 10.1111/j.1540-5907.2008.00332.x 8

[31] M. Fernandes, L. Walls, S. Munson, J. Hullman, and M. Kay. Uncertainty displays using quantile dotplots or cdfs improve transit decision-making. In *ACM CHI*, 2018. doi: 10.1145/3173574.3173718 1, 2, 3, 4

[32] C. Finn and J. Glaser. Voter affect and the 2008 U.S. presidential election: Hope and race mattered. *Analyses of Social Issues and Public Policy*, 10(1):262–275, 2010. doi: 10.1111/j.1530-2415.2010.01206.x 2, 8

[33] J. Freeman. The political culture of the democratic and republican parties. *Political Science Quarterly*, 101(3):327–356, 1986. 8

[34] F. Galton. *Natural Inheritance*, vol. 42. Macmillan, 1889. 4

[35] A. Gelman, J. Hullman, C. Wlezien, and G. Elliott Morris. Information, incentives, and goals in election forecasts. *Judgment and Decision Making*, 15(5):863880, 2020. doi: 10.1017/S1930297500007981 2, 9

[36] A. Gelman, J. Lax, J. Phillips, J. Gabry, and R. Trangucci. Using multilevel regression and poststratification to estimate dynamic public opinion. *unpublished manuscript, Columbia University*, 2, 2016. 6

[37] A. Gerber, M. Hoffman, J. Morgan, and C. Raymond. One in a million: Field experiments on perceived closeness of the election and voter turnout. *American Economic Journal: Applied Economics*, 12(3):287–325, 2020. doi: 10.1257/app.20180574 2

[38] A. S. Gerber, D. P. Green, and C. W. Larimer. Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, 102(1):3348, 2008. doi: 10.1017/S000305540808009X 2

[39] B. Geys. 'Rational' theories of voter turnout: A review. *Political Studies Review*, 4(1):16–35, 2006. doi: 10.1111/j.1478-9299.2006.00034.x 2

[40] E. W. Groenendyk and A. J. Banks. Emotional rescue: How affect helps partisans overcome collective action problems. *Political Psychology*, 35(3):359–378, 2014. doi: 10.1111/pops.12045 2

[41] Y. Gu, S. Gu, Y. Lei, H. Li, et al. From uncertainty to anxiety: How uncertainty fuels anxiety in a process mediated by intolerance of uncertainty. *Neural Plasticity*, 2020. doi: 10.1155/2020/8866386 2

[42] P. K. Han, W. M. Klein, B. Killam, T. Lehman, H. Massett, and A. N. Freedman. Representing randomness in the communication of individualized cancer risk estimates: Effects on cancer risk perceptions, worry, and subjective uncertainty about risk. *Patient Education and Counseling*, 86(1):106–113, 2012. doi: 10.1016/j.pec.2011.01.033 2

[43] S. E. Harpe. How to analyze likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, 7(6):836–850, 2015. doi: 10.1016/j.cptl.2015.08.001 6

[44] P. S. Hart. One or many? The influence of episodic and thematic climate change frames on policy preferences and individual behavior change. *Science Communication*, 33(1):28–51, 2011. doi: 10.1177/1075547010366400 5

[45] M. Heidemanns, A. Gelman, and G. E. Morris. An updated dynamic bayesian forecasting model for the U.S. presidential election. *Harvard Data Science Review*, 2(4), 2020. doi: 10.1162/99608f92.fc62f1e1 3

[46] J. Helske, S. Helske, M. Cooper, A. Ynnerman, and L. Besançon. Can visualization alleviate dichotomous thinking? Effects of visual representations on the cliff effect. *IEEE TVCG*, 27(8):3397–3409, 2021. doi: 10.1109/TVCG.2021.3073466 2, 4

[47] J. Hullman. Why authors don't visualize uncertainty. *IEEE TVCG*, 26(1):130–139, 2020. doi: 10.1109/TVCG.2019.2934287 2

[48] J. Hullman, P. Resnick, and E. Adar. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PLOS ONE*, 10(11):1–25, 11 2015. doi: 10.1371/journal.pone.0142444 2

[49] S. Iyengar, G. Sood, and Y. Lelkes. Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, 76(3):405–431, 09 2012. doi: 10.1093/poq/nfs038 2, 8

[50] S. Iyengar and S. J. Westwood. Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59(3):690–707, 2015. doi: 10.2307/24583091 8

[51] A. Iyer, T. Schmader, and B. Lickel. Why individuals protest the perceived transgressions of their country: The role of anger, shame, and guilt. *Personality and Social Psychology Bulletin*, 33(4):572–587, 2007. doi: 10.1177/0146167206297402 2

[52] A. Jordan, F. Krüger, and S. Lerch. Evaluating probabilistic forecasts with scoringrules. *Journal of Statistical Software*, 90(12):1–37, 2019. doi: 10.18637/jss.v090.i12 3

[53] A. Kale, M. Kay, and J. Hullman. Visual reasoning strategies for effect size judgments and decisions. *IEEE TVCG*, 27(2):272–282, 2021. doi: 10.1109/TVCG.2020.3030335 1, 2, 3, 4

[54] J. P. Kastellec, J. R. Lax, M. Malecki, and J. H. Phillips. Polarizing the electoral connection: Partisan representation in supreme court confirmation politics. *The Journal of Politics*, 77(3):787–804, 2015. doi: 10.1086/681261 6

[55] J. P. Kastellec, J. R. Lax, and J. Phillips. Estimating state public opinion with multi-level regression and poststratification using R. *unpublished manuscript, Princeton University*, 2010. 6

[56] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson. When (ish) is my bus? User-centered visualizations of uncertainty in everyday, mobile predictive systems. In *ACM CHI*, 2016. doi: 10.1145/2858036.2858558 2, 3, 4

[57] L. Kennedy and A. Gelman. Know your population and know your model: Using model-based regression and poststratification to generalize findings beyond the observed sample. *Psychological Methods*, 26(5):547, 2021. doi: 10.1037/met0000362 5, 6

[58] S. H. Khandkar. Open coding. *University of Calgary*, 23:2009, 2009. 8

[59] M. Koval and Y. Jansen. Do you see what you mean? Using predictive visualizations to reduce optimism in duration estimates. In *ACM CHI*, 2022. doi: 10.1145/3491102.3502010 3

[60] O. Kuru, J. Pasek, and M. W. Traugott. Motivated reasoning in the perceived credibility of public opinion polls. *Public Opinion Quarterly*, 81(2):422–446, 05 2017. doi: 10.1093/poq/nfx018 2

[61] J. R. Lax and J. H. Phillips. Gay rights in the states: Public opinion and policy responsiveness. *American Political Science Review*, 103(3):367386, 2009. doi: 10.1017/S0003055409990050 6

[62] T.-T. Lee. Why they don't trust the media: An examination of factors predicting trust. *American Behavioral Scientist*, 54(1):8–21, 2010. doi: 10.1177/0002764210376308 5

[63] E. Lee-Robbins and E. Adar. Affective learning objectives for communicative visualizations. *IEEE TVCG*, 29(1):1–11, 2023. doi: 10.1109/TVCG.2022.3209500 2, 9

[64] D. A. Linzer. Dynamic bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association*, 108(501):124–134, 2013. doi: 10.1080/01621459.2012.737735 3

[65] L. Liu, A. P. Boone, I. T. Ruginski, L. Padilla, M. Hegarty, S. H. Creem-Regehr, W. B. Thompson, C. Yuksel, and D. H. House. Uncertainty visualization by representative sampling from prediction ensembles. *IEEE TVCG*, 23(9):2165–2178, 2017. doi: 10.1109/TVCG.2016.2607204 2

[66] M. Madsen and S. Gregor. Measuring human-computer trust. In *Australasian Conference on Information Systems*, vol. 53, pp. 6–8, 2000. 5

[67] G. J. Madson and D. S. Hillygus. All the best polls agree with me: Bias in evaluations of political polling. *Political Behavior*, 42(4):1055–1072, 2020. doi: 10.1007/s11109-019-09532-1 2

[68] C. F. Manski. The lure of incredible certitude. *Economics & Philosophy*, 36(2):216245, 2020. doi: 10.1017/S0266267119000105 2, 9

[69] I. Marinovic, M. Ottaviani, and P. Sorensen. Forecasters' objectives and strategies. In *Handbook of Economic Forecasting*, vol. 2, pp. 690–720. Elsevier, 2013. doi: 10.1016/B978-0-444-62731-5.00012-9 2

[70] D. J. McAllister. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1):24–59, 1995. doi: 10.5465/256727 5

[71] A. M. McCright, K. Dentzman, M. Charters, and T. Dietz. The influence of political ideology on trust in science. *Environmental Research Letters*, 8(4):044029, 2013. 2

[72] S. M. Merritt, H. Heimbaugh, J. LaChapell, and D. Lee. I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors*, 55(3):520–534, 2013. doi: 10.1177/0018720812465081 8

[73] E. C. Nisbet, K. E. Cooper, and R. K. Garrett. The partisan brain: How dissonant science messages lead conservatives and liberals to (dis)trust science. *The Annals of the American Academy of Political and Social Science*, 658:36–66, 2015. 2, 4, 5

[74] P. Oleskog Tryggvason. *Under the influence? Understanding medias coverage of opinion polls and their effects on citizens and politicians*. PhD thesis, University of Gothenburg, 2021. 2

[75] L. Padilla, S. Dryhurst, H. Hosseinpour, and A. Kruczkiewicz. Multiple hazard uncertainty visualization challenges and paths forward. *Frontiers in Psychology*, 2021. doi: 10.3389/fpsyg.2021.579207 2, 8

[76] L. Padilla, R. Fygenson, S. C. Castro, and E. Bertini. Multiple forecast visualizations (MFVs): Trade-offs in trust and performance in multiple covid-19 forecast visualizations. *IEEE TVCG*, 29(1):12–22, 2023. doi: 10.1109/TVCG.2022.3209457 1, 2

[77] T. E. Patterson. Of polls, mountains: U.S. journalists and their use of election surveys. *Public Opinion Quarterly*, 69(5):716–724, 01 2005. doi: 10.1093/poq/nfi065 2

[78] A. Sarma, S. Guo, J. Hoffswell, R. Rossi, F. Du, E. Koh, and M. Kay. Evaluating the use of uncertainty visualisations for imputations of data missing at random in scatterplots. *IEEE TVCG*, 29(1):602–612, 2023. doi: 10.1109/TVCG.2022.3209348 4, 6

[79] H. Shirani-Mehr, D. Rothschild, S. Goel, and A. Gelman. Disentangling bias and variance in election polls. *Journal of the American Statistical Association*, 113:607–614, 2018. doi: 10.1080/01621459.2018.1448823 3

[80] C. Stokes, V. Setlur, B. Cogley, A. Satyanarayan, and M. A. Hearst. Striking a balance: Reader takeaways and preferences when integrating text and charts. *IEEE TVCG*, 29(1):1233–1243, 2023. doi: 10.1109/TVCG.2022.3209383 2

[81] N. A. Valentino, T. Brader, E. W. Groenendyk, K. Gregorowicz, and V. L. Hutchings. Election night's alright for fighting: The role of emotions in political participation. *The Journal of Politics*, 73(1):156–170, 2011. doi: 10.1017/S0022381610000939 2, 8

[82] J. W. Van Deth. Studying political participation: Towards a theory of everything. In *Joint Sessions of workshops of the European Consortium for Political Research*, pp. 6–11. Grenoble, 2001. 5

[83] K. Walker and M. Herman. *tidycensus: Load U.S. Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*, 2023. R package version 1.3.2. 6

[84] D. Watson and L. A. Clark. The PANAS-X: Manual for the positive and negative affect schedule-expanded form. 1994. 4

[85] C. Weber. Emotions, campaigns, and political participation. *Political Research Quarterly*, 66(2):414–428, 2013. doi: 10.2307/23563153 2, 8

[86] S. J. Westwood, S. Messing, and Y. Lelkes. Projecting confidence: How the probabilistic horse race confuses and demobilizes the public. *The Journal of Politics*, 82(4):1530–1544, 2020. doi: 10.1086/708682 1, 2, 8

[87] B. Witzenbergera and N. Diakopoulos. Election predictions in the news: How users perceive and respond to visual election forecasts. *Information, Communication & Society*, pp. 1–22, 2023. doi: 10.1080/1369118X.2023.2230267 2

[88] C. Xiong, A. Sarvghad, D. G. Goldstein, J. M. Hofman, and C. Demiralp. Investigating perceptual biases in icon arrays. In *ACM CHI*, 2022. doi: 10.1145/3491102.3501874 2

[89] F. Yang, M. Hedayati, and M. Kay. Subjective probability correction for uncertainty representations. In *ACM CHI*, 2023. doi: 10.1145/3544548.3580998 4

[90] F. Yang, Z. Huang, J. Scholtz, and D. L. Arendt. How do visual explanations foster end users' appropriate trust in machine learning? In *ACM IUI*, 2020. doi: 10.1145/3377325.3377480 2, 5, 8

[91] M. Zamo and P. Naveau. Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts. *Mathematical Geosciences*, 50(2):209–234, 2018. doi: 10.1007/s11004-017-9709-7 3

[92] D. Zhang, E. Adar, and J. Hullman. Visualizing uncertainty in probabilistic graphs with network hypothetical outcome plots (NetHOPs). *IEEE TVCG*, 28(1):443–453, 2022. doi: 10.1109/TVCG.2021.3114679 2