The Backstory to "Swaying the Public": A Design Chronicle of Election Forecast Visualizations

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

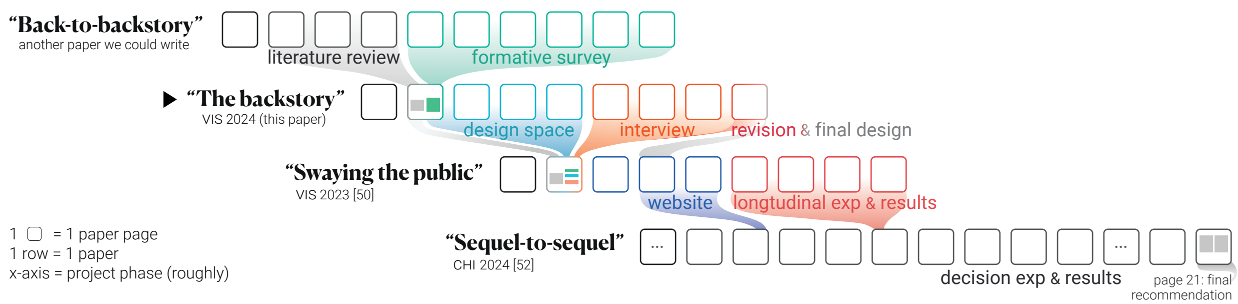


Fig. 1: **The storyline and the scope of each paper**; "probability correction" [51] and "election sausage" [17] are also highly relevant.

Abstract—A year ago, we submitted an IEEE VIS paper entitled "Swaying the Public? Impacts of Election Forecast Visualizations on Emotion, Trust, and Intention in the 2022 U.S. Midterms" [50], which was later bestowed with the honor of a best paper award. Yet, studying such a complex phenomenon required us to explore many more design paths than we could count, and certainly more than we could document in a single paper. This paper, then, is the unwritten prequel—the backstory. It chronicles our journey from a simple idea—to study visualizations for election forecasts—through obstacles such as developing meaningfully different, easy-to-understand forecast visualizations, crafting professional-looking forecasts, and grappling with how to study perceptions of the forecasts before, during, and after the 2022 U.S. midterm elections. This journey yielded a rich set of original knowledge. We formalized a design space for two-party election forecasts, navigating through dimensions like data transformations, visual channels, and types of animated narratives. Through qualitative evaluation of ten representative prototypes with 13 participants, we then identified six core insights into the interpretation of uncertainty visualizations in a U.S. election context. These insights informed our revisions to remove ambiguity in our visual encodings and to prepare a professional-looking forecasting website. As part of this story, we also distilled challenges faced and design lessons learned to inform both designers and practitioners. Ultimately, we hope our methodical approach could inspire others in the community to tackle the hard problems inherent to designing and evaluating visualizations for the general public.

Index Terms—Uncertainty visualization, probabilistic forecasts, design space, animation

1 INTRODUCTION

This story began in November 2016. Despite mainstream media news outlets forecasting a victory for Hillary Clinton, the election night revealed Donald Trump's unexpected ascent to the presidency.[1] That night, juggling a homework deadline, I[2] found myself intermittently refreshing the election map and woke up to a transformed world the following day. Fast forward to the autumn of 2020: amid a global pandemic and a relentless election *week* that painted a picture of uncertainty for countless individuals, I perched anxiously in front of my desktop, watching the last author's stream of tweets.

These personal experiences were unforgettable. It was not until April 2021 that Matthew Kay (Matt) and I finally talked about our shared experiences, and agreed on a vision to investigate uncertainty visualizations for U.S. election forecasts. In January 2022, I started a post-doc at Northwestern University. Our early conversations were often punctuated with "I don't know." The wall of questions grew taller with each meeting: What even is (or isn't) an election forecast? What visualization types are meaningful to study in a political context? How do laypeople understand forecasts? How can we study what people actually think about election forecast visualizations? What if we ran a study during the upcoming midterm elections? How would we even do that, logistically?

A year later, we had completed a longitudinal study of forecast visualization perceptions during the 2022 U.S. midterm elections, along with ten collaborators from the domains of journalism, political communication, and perceptual psychology. We had developed and deployed four meaningfully different but well-composed visualizations, engaged thousands of potential voters, and measured perceptions of the forecasts before, during, and after the elections [50]. Looking back, we made substantial progress on many of the questions we once thought insurmountable. That said, our original paper [50] is just the end result of what we learned. To get there, we first needed to get a handle on the design space and build a professional-quality forecasting website—just so we could study those visualizations in an ecologically valid way. Doing that required an entire additional paper's worth of work, and generated an additional paper's worth of knowledge.

This, then, is that additional paper: the backstory. In keeping with the nature of a backstory, we choose to write it in a slightly informal tone to preserve honesty. Our story unfolds with a formative survey, in which we collected people's experiences with election forecasts through Prolific (Sec. 2). To provide a foundation for our exploration, we formally defined a design space tailored for U.S. election forecasts, cataloging dimensions like data transformations, visual channels and layouts, and animated narratives (Sec. 3). Following this, we qualitatively evaluated ten representative prototypes in the design space through interviews with 13 participants from the formative survey (Sec. 4). This study produced six core insights into how people interpret uncertainty visualizations and reason about probability in a U.S. election context—such as confounding win probability with vote share, and erroneously forming connections between concrete visual representations (like dots) and real-world entities (like votes). Informed by these insights and further discussions with former forecast website designers, we revised our initial designs to ensure comprehensibility for a lay audience (Sec. 5). As we concluded our jour-

- *All authors are with Northwestern University. Emails: {fy, ayse.lokmanoglu, nad, erik.nisbet, mjskay}@northwestern.edu and {mandicai2028, chloemortenson2026, hoda}@u.northwestern.edu.*
- *For any question, please contact the first author at fy@northwestern.edu*

[1]The U.S. elections are dominated by the left-leaning Democratic and right-leaning Republican parties; election day is the last day on which voters may vote, usually the first Tuesday in November.

[2]The personal pronoun "I" is used to refer to the first author wherever suitable.

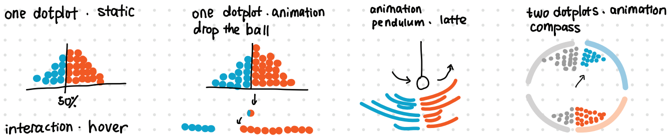


Fig. 2: **Sketches from our early stage**. Colors were modified to match the other figures. See supplementary materials for the original versions.

ney, we distilled the primary challenge we had encountered: ensuring that viewers in the wild would interpret our visualizations as intended (Sec. 6). In our efforts to address this challenge, we also acquired critical design knowledge: incorporating extensive annotations can remove ambiguity for a reader when interpreting a visualization.

Our journey yielded a rich body of novel knowledge, comprising **a design space**, **interview insights**, **our research process**, and **the design lessons** we learned. These elements, in their entirety, constitute the "**contributions**" of this paper. While our design space and interview insights have the potential to ignite future explorations in the realm of uncertainty visualizations, we also hope that readers will find value in knowing our process and lessons learned from designing visualizations for a real-world event and massive audiences. Additionally, we provide our survey questions, interview protocol, videos, sketches, and prototype code at `https://www.doi.org/10.17605/osf.io/ygq2v`.

2 BACKGROUND

The collective anxiety when immersed in deep uncertainty on election night (or even, in 2020, election *week*) left a strong impression on us. However, these anecdotal experiences may not directly translate to actionable research. To explore research directions, we first appealed to the literature in uncertainty communication and political science, and reviewed the existing practices of media news outlets at that time.

2.1 Literature Review

Uncertainty Communication. The literature on uncertainty communication is rich, including contributions from our co-authors [22, 32, 35]. A number of representations have been proposed, including summary plots (e.g., error bars [20, 32]), distributional plots (e.g., density plots [28, 32], fan charts [39]), discretized representations (e.g., quantile dotplots [22, 35], icon arrays [49]), and animations [27, 29, 53]. In particular, Gelman et al. [23] and Padilla et al. [40] discussed possible design spaces for uncertainty visualizations. These guidelines do not adequately address the wide range of design possibilities for uncertain visualizations. Thus, we chose to define a design space first (Sec. 3).

Political Science and Economics. The most relevant topic is how political polls affect voters' perceptions and actions. The literature suggests that polls can shape public opinion [41], influence voter perception [15, 38], and affect voter turnout [12, 13]. The key is the perception of electoral closeness and pivotality (the importance of a vote) [18, 24, 25], which can cause bandwagon [14, 15, 21, 46] and underdog effects [13]. A few works specifically studied election forecasts [23, 33, 48], yielding concerns about how they might confuse and influence voting behaviors. All were helpful in our survey design, but unclear about the role of visual representations in these processes.

Design Practices up to 2022. We also reviewed the designs used by media news outlets when showing likely winners in elections, including The New York Times' animated needle [10], FiveThirtyEight's bee swarm plots [1, 2] and histograms [1, 2], and The Economist's textual summary [6] and gradient intervals [5] (see Fig. 3). The New York Times also experimented with a dice-spinning animation to let viewers experience uncertainty [9]. Most of these outlets have a map to show

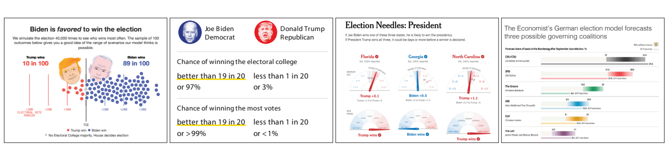


Fig. 3: **Common examples of election forecasts**: FiveThirtyEight (2020) [1], The Economist (2020) [6], New York Times (2020) [10], and The Economist (2021) [5].

and let users interact with state-level data [1, 2, 6, 8]. For live election forecasts published on election night, common practices include text summaries and gradient bar graphs [19]; bar graphs and maps [31, 42, 45] are often used to show the current state of vote tallies.

Election Forecast vs. Winner Projection. One distinction we noticed is the difference between pre-election forecasting (prediction of winner before or on election day) and post-election winner projection (live election forecasts after voting sites have closed). These two are both predictions of winners and are connected through the true uncertainties in election outcomes. Pre-election forecasting may be more influential because of its time span (e.g., several months) and potential consequences (e.g., changing trust and voting behavior), while post-election winner projections can reach a wider audience and are reported by all major news outlets (e.g., CNN, The New York Times). We **decided to study pre-election forecasting** based on our interests at the time. Later, Mandi (the second author) investigated post-election winner projection [17], which may interest a curious reader.

2.2 Formative Survey

We did not find sufficient answers from the literature, particularly regarding how people comprehend election forecasts, let alone their visual representations. Thus, we decided to conduct a formative survey to collect empirical data on people's experiences with election forecasts. This survey was crafted through several brainstorming sessions among the authors. To ensure the quality and relevance of the data, it had two branches: a short branch to filter out less-experienced participants and a long branch to collect in-depth insights from those with substantial experience in engaging with election forecasts.

Survey Design. The first short branch checked whether participants visited any major election forecasting website (e.g., CNN, FiveThirtyEight, The Economist, RealClearPolitics). The second long branch followed the short branch, and had questions asking their experience with election forecasting websites (e.g., FiveThirtyEight), election forecasts' effects on their decisions, their perception of how election forecasts affect others, along with free-text explanations. The exact questions can be found in supplementary materials.

Participants. We requested a U.S. demographically-balanced sample via Prolific.com and obtained 315 participants; 156 participants entered the long branch, and 145 (63 female, 77 male, 5 others) accomplished it; 134 participants reported that they usually vote in U.S. presidential elections.

Outcome. Because the purpose of this survey was to generate possible research themes, we did not perform any inferential statistical analysis. Instead, we conducted thematic analyses of participants' free-text responses (145 participants $\times$ 9 questions). We had briefly reported the key results from the thematic analyses in Sec. 2.2 of our previously-published paper [50]: they showed the importance of emotions, and revealed a large gap between self-perception of political behaviors (e.g., voting) and the perception of other people. The other results also extended or reinforced the observations reported in the political science literature. We will look for another opportunity (perhaps a backstory to a backstory) to report those findings more fully, or readers can refer to supplementary materials. **This survey was critical**, testifying to the potential value of our research direction and highlighting possible measures that could be worth investigating. This gave us a solid foundation from which we could begin designing visualizations for U.S. election forecasts.

3 DESIGN SPACE

We planned to follow a systematic approach to designing election forecast visualizations, as we would be facing a real-world event and massive audiences. However, we did not find any sufficiently well-defined design space or any direct guidance, as most prior wisdom was tailored to other contexts. Facing this situation, we decided to map out the design space for election forecasts ourselves, with the dual aim of illuminating this space and paving the way for future exploration of uncertainty visualization designs. Drawing upon existing uncertainty visualization literature and current (as of 2022) practices, as well as our

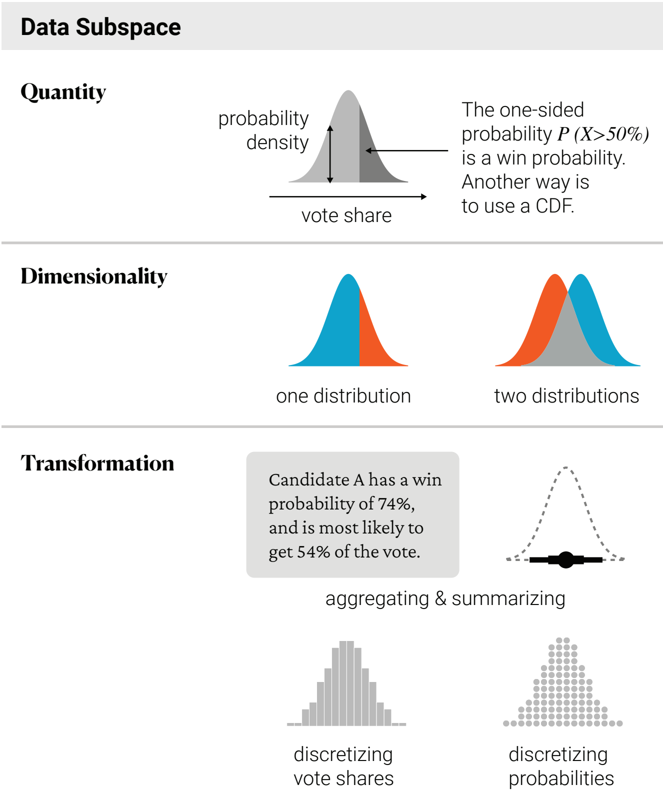


Fig. 4: **Illustration of the data subspace**.

own brainstorming and prototyping, we iteratively constructed and reconstructed our design space until we arrived at three subspaces: data, visual representation, and animation, each with its own dimensions. As for interactivity, we listed a small set of considerations.

3.1 Data Subspace (Fig. 4)

To create any data visualization, we must have the underlying data; thus, we first enumerate the data that could be visualized in a U.S. election forecast.

Quantity. Because the U.S. electoral system is dominated by the two major political parties (Democrats and Republicans), we simplified the problem and considered only the percentage of total votes from two-party voters: the vote share. Most probabilistic election forecasts estimate the vote share distribution. As such, a probability density function (PDF) describes the relative likelihood of any given vote share value; the area under the PDF within a particular range is the win/loss probability of the vote share falling within that range. The one-sided probability $P(X > 50\%)$ is the win probability. The probabilities of one candidate winning and the other candidate losing always add up to 1. Another quantity of interest may be the cumulative distribution function (CDF), $P(X \leq x)$. However, CDFs contain more information than is typically necessary to tell the winner or race competitiveness (all one-sided probabilities), and may be difficult for lay audiences to comprehend [30]. Therefore, we opted for probability density functions, variations of which are also common practices (see Sec. 2.1). As such, the core quantities in election forecasts are vote share and win probability. Both should be conveyed to readers in a probabilistic forecast.

Dimensionality. The two-party system allows for the use of one distribution to convey both parties' vote shares and win probabilities, with each side of the distribution representing one party. We can also show both candidates' vote share distributions directly (two distributions).

Transformation. We can transform a vote share distribution, by aggregating or summarizing the distribution to derive point or interval estimates. We can also discretize vote shares by binning them to generate a histogram, or discretize probabilities to generate a quantile dotplot [22, 35] (e.g., each dot represents a probability of 1%).

3.2 Visual Representation Subspace (Fig. 5)

To explore visual representations, we considered both the visual channels used to encode the data and the layout of resulting visual elements. We **prioritized frequency representations**, because a lay audience usually understands them better than a probability representation [22, 32].

Visual Channel. Among common visual channels, position, luminance, length, and angle can encode vote share, probability, or probability density. For instance, a density plot uses position to encode both vote share and probability density. The same encodings with discretization yield a histogram or a quantile dotplot. We can dual-code the probabilities in a histogram using luminance or color and flatten it to get histogram intervals, which are truncated within the 95% prediction interval, following a design by The Economist [5]. We also explored a somewhat adversarial encoding: using luminance to emphasize the bulk (shrink) or tail (pull out) of a distribution. This was inspired by our work on subjective probability correction [51]: pull out is an attempt to correct people's tendency to ignore small tail probabilities by making the distribution appear wider. Although it is difficult to apply length and angle, they both are feasible when we discretize probabilities (e.g., drawing samples based on probability density). For each draw from the distribution, we can use length or area to encode the vote share, or map the vote share to angle to get a small pie chart. We eliminated shape as it is improper for numeric data.

Layout. The most common choice for layout is to use a Cartesian coordinate system, though a polar coordinate system is also possible. More loosely, we can use a grid (icon array), a list, or a bee swarm layout [34] that nudges dot positions (used by FiveThirtyEight [1, 2]). A common practice for displaying two dotplots is to arrange them vertically, one above the other (juxtaposition [26]). However, the space constraints for a website presentation challenge this convention. To have a more compact design, we first came up with a reflection layout, inspired by Northwestern's pond fountains (see supplementary materials). Based on the insights from the interviews described in Sec. 4, we revised the reflection layout to a layout that blends two distributions using half dots. This is a refinement of superposition [26] tailored to this chart type.

3.3 Animation Subspace (Fig. 6)

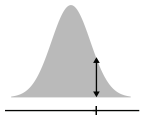
Uncertainty visualization often uses animation to convey uncertainty or randomness [27, 29, 53]. At the time, both Matt and I were passionate about animated visualizations: animation might help people experience uncertainty or engage them through a narrative. We were also fortunate to have two experts on dynamic displays at Northwestern—Ouxun Jiang, a Ph.D. student, and her advisor, Steven Franconeri. They have been working on collecting and cataloging different dynamic displays. They shared with us a few guidelines summarized from the literature, such as staging [27, 36], drawing trajectories [16], considering user control [37], and being mindful of the number of moving objects [4]. Our animation subspace was partly inspired by their hard work.

Narrative. A common form of animated uncertainty visualization without a narrative is a hypothetical outcome plot (HOP) [29, 32, 53], looping through possible outcomes. Generalizing Matt's early version of Presidential Plinko [7], animation can depict an accumulation process, implying a storyline by gradually adding more draws from the distribution. The reverse is a dissipation process, removing draws from the distribution. One possibility to convey the relation between vote share and probability is to combine accumulation with dissipation (e.g., counting dots, the second sketch in Fig. 2). A narrative alone can be abstract, but it can also be combined with an analogy to help viewers grasp the intention of the animation.

Analogy. To help viewers grasp the meaning of an animation, the animation can be designed with an analogy to reflect a real-world concept or a natural process. A first step is to relate the animation to a plausible but abstract process, such as removing or adding dots, which we term an abstract analogy. Then, a simple association between the animation and a real-world entity is a figurative analogy, such as replicating a needle or compass. A further step involves an analogy that fully connects

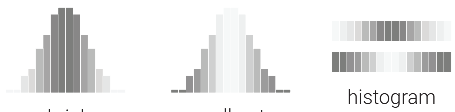
Visual Representation Subspace

Visual Channel



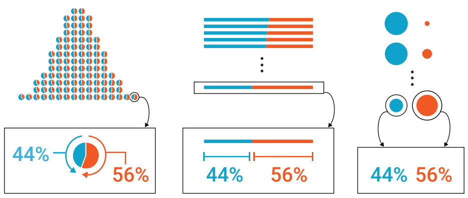
A density plot encodes vote share with **position**, probability density with **position/length**, and probability with **area**.

We can use **luminance** (or **color**) to dual code probability density. **Pull out** is an attempt to apply a subjective probability correction. The **intervals** are flattened **histograms**.



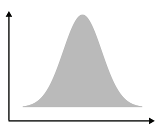
shrink | pull out | histogram intervals

We can be creative with a quantile representation. Below are different encodings for 100 draws. Each draw is a possible outcome, containing **D**'s and **R**'s vote shares (e.g., **44%** vs. **56%**).



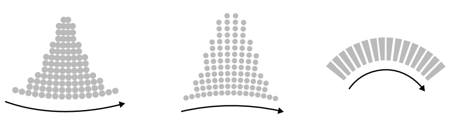
angle | length | area

Layout



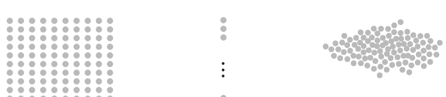
A standard density plot uses a **Cartesian** coordinate system.

Introducing a **polar** coordinate system allows a circular visualization.



centripetal | centrifugal | fan/centrifugal

Or we can reduce the complexity and use a loosely defined coordinate system.



grid | list | bee swarm

It is not obvious how to lay out two dotplots effectively. We have three ways.

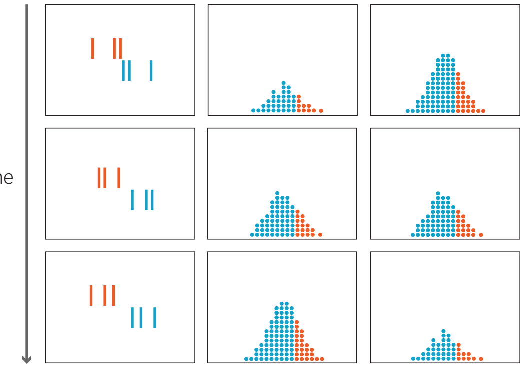
juxtaposition | reflection | blend

Fig. 5: **Illustration of the visual representation subspace.**

Animation Subspace

Narrative

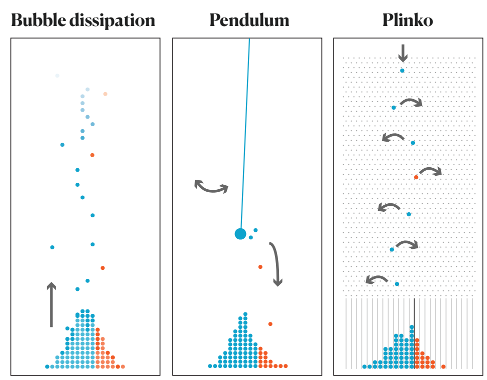
Given a discretized representation, we can present it as an **accumulation** process, and the reverse is a **dissipation** process. A narrative has to be combined with an **analogy** to be meaningful and understandable.



no narrative (HOP) | an accumulation process | a dissipation process

Analogy

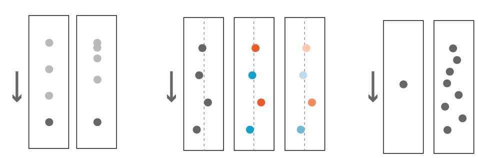
At the surface level, mapping a narrative to an abstract process could yield an **abstract analogy**. A slightly deeper level is to use a real-world concept, such as a needle, a compass, or a pendulum. Because this level matches only the visual form, we call it a **figurative analogy**. The even deeper level matches the visual form with the data generation process, called a **mechanistic analogy**. Our example here is **Plinko**, where a ball bounce represents a Bernoulli distribution, and a series of bounces resembles a Binomial distribution that approximates the Normal vote share distribution. From our experience, designing these analogies needs both creativity and technical expertise.



abstract | figurative | mechanistic

Configuration

We can configure the same narrative and analogy differently. We can add **acceleration and deceleration effects to a trajectory, change dot colors** according to where they are, and decide **how many dots to animate** at the same time.



trajectories | color changes | # of dots animated

Fig. 6: **Illustration of the animation subspace.**

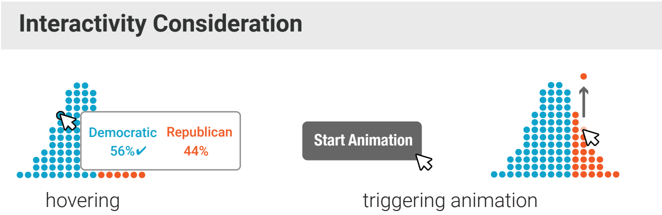


Fig. 7: **Illustration of the interactivity considerations**.

the animation to the data generation process, such that the underlying mechanism of the real-world process can give rise to the same probability distribution being depicted; we call this a mechanistic analogy. Our only attempt for a mechanistic analogy was **Plinko**, designed to reflect the data generation by using the Binomial distribution to approximate the vote share distribution. Another example of a mechanistic analogy used in election forecast visualization is dice rolling [11].

Configuration. We can configure the same narrative and analogy differently. We can vary the number of frames per second (FPS), animation duration, trajectory functions, changes in visual elements (e.g., color) over time, and the numbers of elements animated simultaneously, particularly in dotplot-like designs.

3.4 Interactivity Considerations (Fig. 7)

We decided not to have specific designs for interactivity. One reason was that the primary goal of a forecasting website is to inform viewers (i.e., it is communicative), so interactivity (more crucial for exploration) is a secondary consideration. However, a minimal level of interactivity can show more information and improve user experience. In a discretized design, a viewer can hover over an element (e.g., a dot, a bar) to gain more information about the underlying data, perhaps to make a possible election outcome more concrete. In an animated design, a viewer can choose to click on a button or mouse over an element to trigger the animation, lending them agency or control.

3.5 Design Generation & Internal Evaluation

Our design space gave us numerous possible visualization designs. We permuted different dimensions and eliminated impractical or overly complex combinations. Many of the animated designs could not (and cannot still) be implemented by the `ggdist` package [34] created and maintained by Matt. After assessing feasible combinations, I used `D3.js` to implement more than 40 visualization designs. These designs are hosted at `https://forecasts.cs.northwestern.edu/2022-initial-prototypes`, and the code is provided in supplementary materials.

As this set was infeasible for any formal quantitative or qualitative study, and it was unnecessary to evaluate all designs, we went to our colleagues for a first round of feedback. We first distributed the prototypes in our lab meeting with about 10 Ph.D. students and 1 faculty member. The lab meeting had two sessions. Each person explored 3–4 prototypes for about 25 minutes (to allow time for typing in feedback). All lab members then had a focal discussion for about 25 minutes. Both sessions had a set of seed questions to guide the thinking and discussion (e.g., How do you think these charts show the uncertainty in the outcome?). We also arranged another hour-long meeting with Ouxun Jiang and Steven Franconeri to go over each prototype and its configurations.

All discussions and feedback were incredibly valuable. One important guideline we received is "**let the animation feel natural**." We adjusted all animation configurations towards this goal and eliminated a few unnatural designs (e.g., **Bubble disspation**).

4 INTERVIEW

Although the remaining set (about 10–20) was manageable for in-lab studies, it was still too large to be deployed on a public website. Matt then suggested a qualitative study, which would help quickly narrow down our scope and gain deep insights. Nick (the third-to-last author) also made the same suggestion. I had never formally conducted a qualitative study myself, but I trusted their judgment and believed this was feasible given my experience in supervising in-person experiments. We therefore designed an interview study incorporating a think-aloud protocol complemented by a series of elicitation questions. This study gave us valuable insights into people's interpretations of uncertainty visualizations, informing our later revisions.

4.1 Methods

Prototypes. We selected nine visualizations and a text representation to cover a range of design dimensions and ensure a reasonable interview time. We felt it necessary to compare designs with both one and two distributions, so we included a single quantile dotplot (**1-Dotplot**) and dual quantile dotplots (**2-Dotplot**). The animated narrative of a quantile dotplot could be **Particle**, and with a mechanistic analogy, this turned into **Plinko**. People might confuse vote share percentage and probability of winning [48], so we wanted to include designs that explains these concepts. For this reason, we included **Playground**, an animation explaining a conversion between vote share and probability through dropping dots. We also selected **Needle** (used by New York Times [10]), and its discretized version **Compass** (shooting dots). Additionally, we selected **Mint** to test the angle encoding, and the animation of **Mint** was **Moonphase**, which also illustrated the conversion between vote share and probability through dropping dots (or moons). In summary, we had ten prototypes: **1-Dotplot**, **2-Dotplot**, **Particle**, **Plinko**, **Playground**, **Needle**, **Compass**, **Mint**, **Moonphase**, and **Text**. These prototypes covered the variations of transformation, dimensionality, visual encoding, layout, narrative, and analogy. They also contained basic annotations, as shown in Fig. 8.

Stimuli. We drew 100 samples from Normal(52.5%, 2.5%) to generate a forecast vote share distribution. This roughly corresponds to the known polling errors in the U.S. and a 74% win probability, similar to the 2016 U.S. presidential forecast [8]. We had two different forecast distributions, depending on which party's vote share was being predicted. We showed each participant all ten prototypes using a Latin square to counterbalance learning and carryover effects. We had five prototypes showing a Democratic win, while the other five showed a Republican win. The order of the favored party was also roughly balanced. The interface used in the interview is hosted at `https://forecasts.cs.northwestern.edu/2022-interview-prototypes`, and the video and code are available in supplementary materials.

Interview protocol. I conducted all interviews over Zoom. The base payment was 30 USD for an hour and 2.5 USD for every additional five minutes. Participants could choose to turn off their cameras. After consent, I asked participants to open the link to the interface, share their screen, and follow a think-aloud approach. I then instructed them to view each visualization and answer questions after each. The interview questions were, again, a reflection of all the authors' efforts. The core questions for each visualization are listed below, while the full protocol is available in supplementary materials.

- Can you describe this visualization/animation to me? What does it tell you about who might win?
- How does this animation/visualization make you feel about this race? Does it make you feel more or less uncertain about the race? Does it make you feel more or less worried about the race?
- If the [Democratic|Republican] candidate wins, what would you think about the quality of this forecast?
- What do you like or dislike about this visualization/animation? Is there anything you find confusing or distracting?

I told participants that the true winner was the predicted winner except on the fifth and tenth visualizations they saw, where I told them the forecast was "wrong". At the end, I asked them which visualizations they liked the most and least, which visualizations they would share with their friends and family, and whether they preferred to have more control over the animation. Both their screen and audio were recorded.

Participants. As we needed participants interested in and experienced with election forecasts, we recruited from the 145 participants who completed the long branch of the formative survey (see Sec. 2.2). I started with a signup form and contacted all 18 signups. In total, I interviewed 13 participants from 10 states (see Tab. 1).

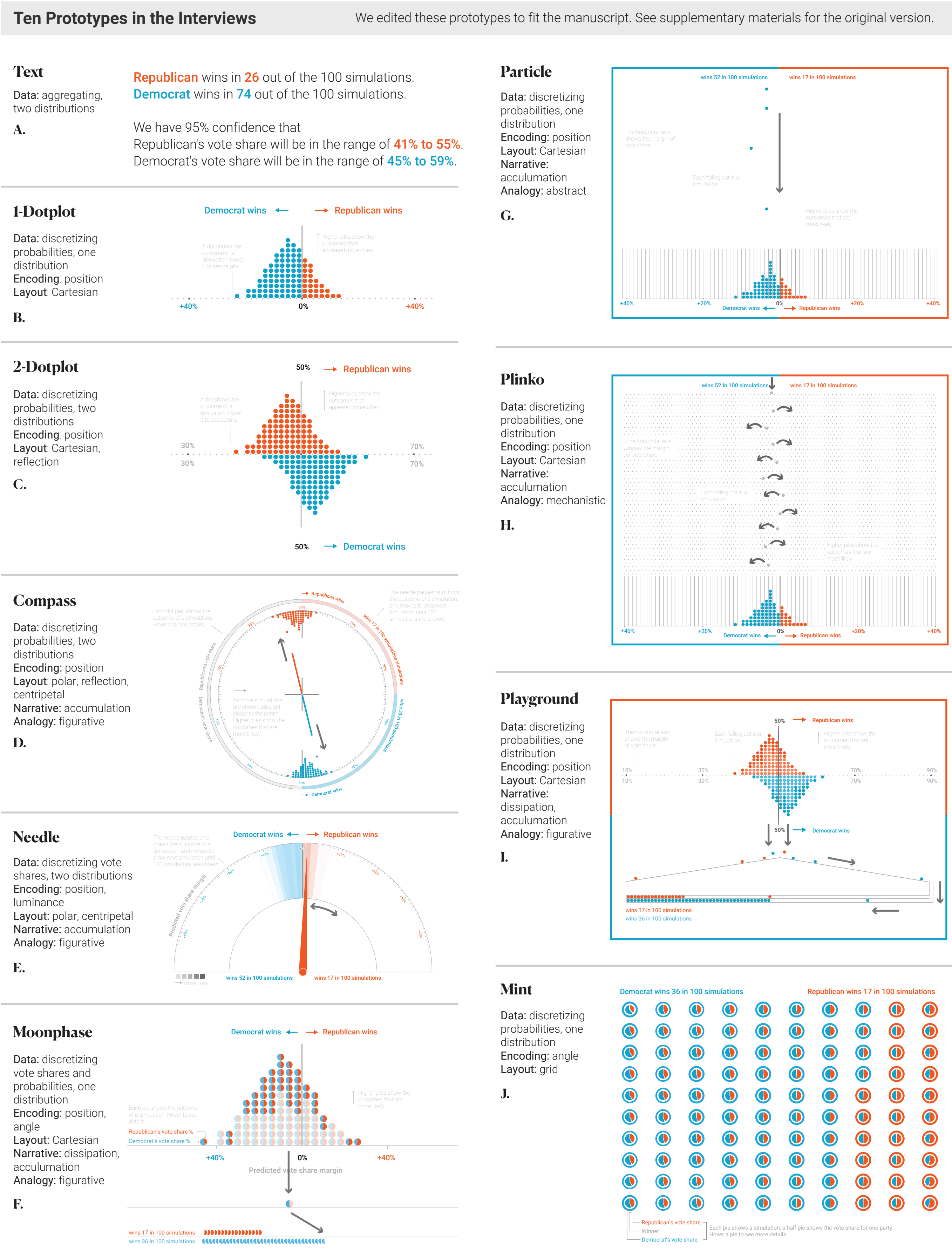


Fig. 8: The ten prototypes in the interviews, including (A) a text description, (BCJ) three static designs, and (DEFHIK) six animated designs.

Table 1: The interview participants' demographics

No.	Age	Gen.	St.	Education	Party identification
P1	39	W	TX	College	A strong Democrat
P2	32	M	TX	Masters/doctorates	Independent
P3	34	M	DC	Some college	A not very strong Democrat
P4	33	M	FL	Masters/doctorates	Independent (Democrat-leaning)
P5	37	M	TX	Some college	Independent
P6	52	W	VA	College	Independent (Democrat-leaning)
P7	38	M	IL	College	A strong Democrat
P8	32	M	MS	Some college	A strong Democrat
P9	65	M	FL	Masters/doctorates	A not very strong Democrat
P10	59	W	CA	Some college	Independent (Democrat-leaning)
P11	71	W	TX	Masters/doctorates	A strong Democrat
P12	50	M	CO	College	A strong Republican
P13	63	W	NE	College	A strong Democrat

Analysis. The 13 participants together provided about 15 hours of video. I analyzed all the videos twice. I first transcribed the interview videos and verified all the transcripts. I started with open coding on the transcripts while rewinding the videos to compile a codebook. After the first pass, I read all the transcripts and codes, and used axial coding to merge them into different axes. Though reliving the confusion and frustration the participants endured was an unforgettably painful experience, this analysis yielded invaluable insights, as shown below.

4.2 Themes Shared across Visualizations

Theme 1 Mistakenly construing visual encodings as real-world concepts

A discretized forecast visualization such as **1-Dotplot** gave participants a sense of visual concreteness. However, this concreteness (e.g., showing a dot) also reminded participants of concrete real-world entities, especially when they did not understand the visual encodings. Participants thought a dot (or a sector in **Moonphase**) could represent a vote (7), a poll (3), a person (3), a district (1), a state (1), or a simulation (1). Participants thought an animation like **Needle** or **Particle** could represent votes coming in (6) or that the model was simulating data onsite (4). For example, "It was just heavily calculating what poll and all that was that it was probably doing some of the different polls and simulations that they made had factored in." (P3) When participants didn't understand the visual encodings, they attempted to use their real-world experiences to arrive at an explanation or make up a story. For example, one participant was confused with dots disappearing in **Playground**, and used a story of voting in a town to explain it: "I don't know. I don't think it was or half the people voted in a town. People didn't vote. I'm just thinking that was just to make it look edgier or real. The ones that fell off ended up back on there or somehow. " (P10)

Theme 2 Confusion about vote share and win probability

Participants had two strategies to reason about the race presented in a visualization: comparing the two win probability numbers (7) and combining vote shares with win probabilities (5), except for 1. Comparing two win probability numbers (74 cf. 26 by subtraction or division) caused an illusion of the race being decisive: "Because I saw that, Democrats 71, 74 after the 100 and the Republicans got only 26. That's a big difference." (P5). Participants were often confused about these two quantities: they thought vote share is the win probability (5), the win probability is vote share (5), or did not understand the relation between them (5). Participants felt the two quantities were two pieces of disconnected and conflicting evidence. Still, they tended to believe the evidence they felt they understood better—most of the time, win probability. One and only one participant understood the conversion between the two quantities when seeing **Text**: "Because all the other simulations, I was assuming that at 74% meant they thought it was going to a 74 to 26 outcome. That really truly means 55 to 45 outcome." (P11)

Theme 3 Using visual cues to reason about race competitiveness

All participants relied on visual cues or heuristics to grasp a visualization and reason about race competitiveness. The most common visual cue was area/magnitude (12) (e.g., comparing the area of red and blue or the magnitude of red and blue balls). Here, a discretized design (e.g., **1-Dotplot**) invited participants to count the dots, which might be perceived as too onerous (6): "... it's just kind of you bring out your calculator and like basically average it out. You're so much a lot of people might be like, oh, we're not going to do math." (P3)

They also relied on color or gradient as a visual clue. This was most common in **Needle** (9) but also in **Moonphase** (2) and **Mint** (2). For example, "So I can clearly see the side is getting more shaded areas. I also like that the shaded areas are color-coded, so that's another easy visual that allows me to understand this graph better." (P7) Another commonly-used visual cue was height (9): "And the balls are stacked up like really nicely all over the place. And it's obvious, by the way, that the blue ones are piled up much higher." (P10)

Less commonly, participants (6) relied on a left and right separation (e.g., in **1-Dotplot** and **Particle**), and referred to outliers (5) to reason about the possible election results: "This prediction is saying that Republicans can win by as much as 10% on election night." (P7) They also looked at length (4), because in **Playground** and **Moonphase**, the winner dots were aligned on a line. Participants appealed to shape or skewness in **1-Dotplot** (1), **2-Dotplot** (1), **Playground** (1), and **Particle** (2). Very rarely, they looked at mode (1 in **2-Dotplot**), frequency (1 in **Needle**), or the spread of the two sides (1 in **Particle**): "it seems like there's more piles here on the Republican side than the Democrat side." (P12)

The cues here were different from those reported by Kale et al. [32]. While both concerns how people reason about uncertainty visualizations, all four visualizations studied by Kale et al. show two distributions. We had many one-distribution visualizations, which diminished cues like distance but emphasized cues like height and shape.

Theme 4 Trading the animation time for engagement

Visualizations designed in an animated form could be engaging (4), especially engaging with elections (2). Animations could create entertainment value (4) and capture attention and interest (8): "The animation more interesting, they showed them running up the edge, but up just to make it more eye catching or stimulating." (P10) Participants indeed felt uncertain (4). However, participants also felt the animation was unnecessary (5), distracting (3), or that they had to wait (3). An animation was perceived as a narrative, and the ending of the animation created a sense of certainty or conclusion (3): "When it's completed, the needle will stop. Or it's predicting who will win. But it's a slow process to get there." (P9) However, animations were slow in many participants' browsers (10), a confound we fixed in our later revisions.

Theme 5 Predicting the winner is most, but not all, of trust

In general, participants trusted a forecast when it correctly "predicted" the winner (all 13), and distrusted one that did not predict the winner (11). However, they might distrust a "correct" forecast for a variety of reasons: confusion and ambiguity (4), untrustworthy polls (4), low trust propensity like some believed forecasts were biased (1), complexity (1), deviation from election results (2), no transparency about the method (2), the animation or representation (2; see **Plinko** and **Playground** below), or even the 95% confidence (1 thought this was forecasters' confidence, and felt it was low).

They also used election results (5), win probability (1), and their prior knowledge (1) in judging the forecast quality: "It would make me question the quality of the forecast if the Democrat wins, because it clearly shows that the Republicans have a much more likely chance to win." (P7) They might use the vote share in their judgment, when mistakenly thinking it was the win probability, especially if the win probability matched up with the vote share coincidentally in their state: "... since California's very, very blue state, the odds are very high that the Democrats, 74%, ... a blue state have 74%. That sounds normal." (P10)

Theme 6 Reading annotations accompanies sensemaking

Participants usually expressed their thinking or understanding of a visualization upon reading the annotations (9). For example, after reading annotations, "I see a bit of an explanation, which is essentially what I was able to tell by just looking at it", P2 said, "So you got the information from like the overall just the dots." Reading annotations was occasionally followed by confusion (4). Five participants found annotations generally helpful. However, the same number of participants also indicated the annotation of **Compass** were too complex, and one commented on the annotation for **Plinko**, "I find it unnecessary." (P9)

4.3 Reactions to Specific Visualizations

1-Dotplot was very clear (10). However, this clarity also engendered a sense of certainty (all) especially if participants further contrasted between the left and right sides. For example, "I feel more certain and more confident that the Democrats will be able to win just based on how how much it is skewed in the Democrats favor." (P4)

2-Dotplot was complex and onerous to interpret (4): "It takes time to understand what the different axis means and then when the more number of piles of dots are and things like that." (P4) However, this complexity gave a bit more information (4): "I think, it's a little more what I do like about this one. I think this one makes me understand the dots better." (P12) **2-Dotplot** also caused confusion (12): magnitude was a more salient cue than the distance between the two distributions, but the magnitude was the same for the two distributions. These misguided participants (4) to conclude the election was a precise 50–50 split: "... the red dots of the Republican the way and then the blue at the bottom, they look the same, but just on different sides, like. But why does it just show you like who wins? I mean, who how would you know? ... I don't know. It's confusing." (P2) However, this confusion and misperception also mistakenly resulted in a sense of uncertainty (8): "... the 50% dividing by and having them are having one on the top and one on the bottom. It's like there's this specific couple of outcomes that make it look like it could easily come down to a handful of votes." (P8) Additionally, one participant mentioned that placement at the bottom of the plot could carry negative connotations. We fixed this later using the blend layout.

Compass was also complex (5). Likely due to confusion, it also engendered a sense of uncertainty (10) rather than certainty (4). Participants could grasp the analogy of a compass (6) and thought it showed dynamics in elections: "... the way the polls and the election works, just like on TV, when they show up, down or down fast or the numbers go up or whatever because they get the vote shares, you know, then they go at the same time." (P5)

Mint was both confusing and clear, depending on how participants made sense of its encoding. If their focus was the outer circle (showing win probability), they found it clear (9) and praised for its simplicity and clarity (9). However, this also engendered a sense of certainty (2): "So it is showing it heavily Republican by being more red and pinkish and the Democrats last." (P9) If participants started with the inner pie of vote share, they felt confused (7) and subsequently uncertain about the election (3): "So each side has a half a red side and therefore blue side. I don't know what point that serves." (P9)

Moonphase was complex (3), hard to follow (5), and confusing (9). Again, the confusion, complexity, and angle encoding prompted a sense of uncertainty (6): "Kind of close. Probably lean Republican. Probably not likely Republican. But. Okay, I was going to say, it's like halfway." (P8) However, the animation created a narrative concluding with the win probabilities, and the probabilities were clear and easy to understand. Thus, participants still found a sense of certainty despite their confusion (7): "I felt very confident that the Democrats will win. Just looking at the number of models that to the left of the 0% line." (P4)

Needle. Participants understood the analogy of a needle and made comparisons to The New York Times' needle (11). They construed the metaphor as a ruler (1) or a measure (1), and felt uncertain (8): "Come on, come on, go to the blue, you know, so basically you're waiting for it to move back the other way. So it's to you have to watch to see which way it's moving." (P10) However, the two win probability numbers on the side also created a sense of certainty (12). Though participants understood the analogy, they still found the animation confusing and hard to digest (6): "... they're here to show the light, like light blue and light red. Then some shows are dark red. I'm confused about that. Why? Why does that matter? See how good a dark red light then of you to be light? Or is the maybe the higher go, the lighter it goes? I don't know." (P5)

Particle was an animation designed to convey uncertainty, but participants perceived it as illustrating certainty (11): "... you already knew where the ball was going to land. It was either blue or red. The blue ones were on the blue side, the red on the red side, like it was predictable." (P10) This perception was again related to the clarity and simplicity of **Particle** (9): "When it started lining up you could clearly see that it was going to line up for the Republicans to win. You could just tell by the placement of the balls, though that part was easy." (P6)

Plinko. Most participants understood the design analogy of a Plinko game board (7), but one participant was unfamiliar with it, and thought it was a maze. They perceived **Plinko** as entertaining (5) and game-like (3). One participant thought it was attractive: "... the simulations, each one going into a plinko-type game. I think that it makes me smile still. It makes me smile. It's just a fun way of showing. It's just a fun way of showing a reelection forecast." (P12) However, another participant showed strong distrust: "So animation itself makes me feel untrustworthy or unreliable on the race itself because I'm like, I don't know if this data or if this result is actually correct." (P4) Again, the animation rendered a sense of uncertainty (10): "It would probably make me feel more uncertain because I'm always hopeful on the other side" (P13), but the dropping point and the narrative painted a sense of certainty (11): "I would say the Democratic side would win because that's where the ball is starting centered" (P10).

Playground was confusing (10). However, similar to **Moonphase**, even if participants did not understand the animation, they found the win probabilities at the bottom clear (6), and the end of the animation (or its narrative) strengthened this impression. Only two participants (2) understood the converting process: "It's taking information from the top and bringing to the bottom to fill. In a clearer picture like this is all together in blue." (P1) Similar to **2-Dotplot**, the two distributions appeared equal, and participants mistakenly believed the election was 50–50 and thereby felt uncertain (9): "So it factors in like equal 50-50 like over here, maybe a bunch of Republicans just stay home ..." (P8)

Text created a sense of uncertainty in the appropriate way (not due to confusion or misperception) (9). Participants understood the vote share and found the two candidates' vote shares were close. **Text** was also perceived simple and clear (5). However, if participants focused on win probabilities, they still felt inappropriately certain about election outcomes (7): "... the top part where it says it has the 74 and the 26. That to me seems like it's going to be a decisive election." (P13)

Text helped change one participant's mental model of the forecast. The participant was confused about vote share and win probability throughout the interview, but successfully connected the two concepts after reading the text description: "... the percentage we're looking to go in 74, 26. That doesn't sound realistic. But now they're saying that 74 means it's going to be 45, 59, and 26 means 41 to 55 ... that pie chart was trying to say, that 74 out of 100 is really saying 45 to 59. And now those little pies were numbers between 45 and 59. What I'm trying to sort of make the pie thing make sense." (P10) However, participants could misinterpret that 95% confidence (intervals) were forecasters' confidence: "But I would not trust the forecast only because of their 95% certain." (P11) Some thought **Text** provided enough information (7) and even more information (2): "And it's something that provides a feel more information than the visualizations ironically." (P2) But three participants thought **Text** provided less or not enough information (2) and visualizations made them feel more confident (1) and had entertainment value: "It's just not quite as fun, I guess. And you don't have the visualization, which does help." (P8)

4.4 Summarized Insights

How people interpret election forecast visualizations

- Participants rely on preexisting knowledge to interpret the visual encodings, particularly in more concrete designs. For example, they may mistakenly think a dot is a vote (**Theme 1**).
- They have deep confusion about win probability and vote share; a variety of reasons could lead them to compare the two candidates in the probability space, creating an illusion of a decisive election (**Theme 2**).
- They rely on visual cues for understanding visualizations, and use different visual cues in one and two distributions (**Theme 3**).

Design lessons we learned

- In general, participants seem averse to complexity (**2-Dotplot**, **Compass**) and prefer simplicity (**1-Dotplot**, **Particle**, **Mint**).
- They appear to seek certainty and show a reluctance towards accepting uncertainty (**Moonphase**, **Needle**, **Particle**, **Playground**, **Text**).
- Clarity can foster a sense of certainty (**1-Dotplot**, **Mint**); confusion can lead to a sense of uncertainty (**2-Dotplot**, **Moonphase**).
- Animation creates a narrative, and participants consciously and unconsciously focus on the end of the narrative (**Moonphase**, **Playground**), which can engender a sense of certainty despite the animation process itself being perceived as uncertain (**Plinko**).
- Animation has entertainment values; while some participants find it engaging, others find it unscientific or untrustworthy (**Plinko**).
- All animated designs in the interviews were somewhat confusing, but familiar and mechanistic analogies seem to mitigate confusion.
- **Text** seems to cause the least confusion and map to a different mental model, but also carries the least amount of information.

5 REVISION

Learning about what did not work was frustrating; however, we were able to use these insights to revise and improve our designs carefully.

Our key approach to addressing misinterpretation and confusion was to design **extensive annotations** [44] (see Fig. 9) surrounding how to read a forecast visualization. We first darkened and thickened all labels, axes, text, and lines to make sure people would notice (and read) them. To avoid confusion about the meaning of a dot, we informally tested a variety of text explanations, eventually landing on "1 ● = 1 election outcome", which was the most straightforward way we found to convey the visual encoding. We placed this annotation at the top to ensure it would not be ignored by viewers. It was also necessary to explain the meaning of a pile, and we annotated the most likely outcome—we experimented with a handful of participants and asked them to write down their takeaways based on annotations of different piles, but found no difference. Both anecdotally and empirically, we felt that annotations were most effective when crafted as complete sentences. We tested with different placements, and finally broke down annotations and interweaved them into visualization components to guide viewers through the components of the visualization (see Fig. 9). We also undertook multiple iterations to condense the annotations into as few words as possible, and moved secondary annotations to the bottom. Our design sketches were provided in supplementary materials.

Because all the animations were confusing, we kept only the most understandable ones, **Plinko** and **Needle**, and designed *extensive* annotations for them. We also attempted to improve **Playground** and **Needle** by simplifying the visuals and including more annotations. However, we had further conversations with Jessica Hullman (a co-author on [50] and an expert on animated uncertainty visualization) and Anna Wiederkehr (who formerly worked on forecast visualizations at FiveThirtyEight); both suggested the improved designs were still overly complicated. Ultimately, we were also unsure about the value of the non-mechanistic **Needle** analogy, and eventually decided not to deploy it, keeping only the mechanistic **Plinko** analogy.

We also made a number of other revisions. To address the negative connotation associated with being at the bottom in **2-Dotplot** (reflection), we came up with **2-Dotplot** (blend)—this was likely inspired by color blending (e.g., mixing little blue and red dots to get purple). The angle encoding (**Mint** and **Moonphase**) emerged as a possible representation to link vote share with win probability. We simplified it by deleting half of each circle, but this still resulted in a "dizzy" visualization (per Anna Wiederkehr). Finally, as participants preferred simplicity, we added back **Intervals**.

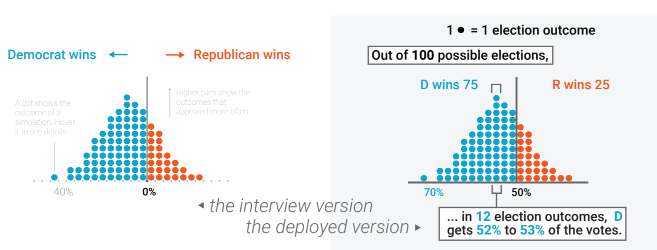


Fig. 9: **Illustration of the revisions we made.**

The deep confusion about win probability and vote share can be difficult to resolve, which motivated our work on a probability correction to directly adjust displayed probabilities [51]. Given the ethical implications of this approach, we further explored this correction in a subsequent paper that measures trust in election forecasts over time [52].

6 ENDING

It took us more than eight months to reach this point. We ultimately deployed four revised designs: **1-Dotplot**, **2-Dotplot**, **Plinko**, and **Intervals**, each with *extensive* annotations. Based on the sequel to this story [50], and the sequel to the sequel [52], which also received a best paper award at CHI, our revisions, particularly the annotations, helped address many issues in the earlier designs, though some ambiguity persisted. These sequels also revealed meaningful differences in the data and visual spaces, such as the potential for visualizations to exacerbate polarization [50], and the impact of partisanship [50, 52] and education level [52] on trust in different designs.

Designing election forecast visualizations for massive audiences was challenging and required navigating a complex research space. Looking back, our primary challenge was to connect a theoretical design space with the practice of making *real* visualizations that people actually understand. Our earlier designs were considered "standard". But our audience consisted of diverse individuals with varied cultural, educational, and political backgrounds, who would also view a visualization in different contexts—at night, on the street, or after a conversation with a friend. This variety can strongly influence how individuals interpret the visualizations. Consequently, aligning audience understanding with our intended messages became our primary goal.

There is no such thing as a universally "intuitive", broadly understandable visualization in our design space, or perhaps in any space. Even when certain visualization types work well in one domain (e.g., quantile dotplots for bus arrival predictions [22, 35]), when transplanted to a new domain, they can yield unexpected and challenging misunderstandings (e.g., misconstruing a dot as a vote instead of an election outcome). Our (partial) solution was to create extensive and high-quality annotations. Beyond captions, labels, and axes, empirically, we found other visual design parameters, such as font size, line thickness, and visual hierarchy, contribute to viewers' (mis)understanding and (dis)trust in a visualization.

No work is flawless, and ours is no exception. Given finite resources, we necessarily had to eliminate some designs we might have liked to explore, such as those based on CDFs or those tailored to different devices (e.g., mobile, tablet). Owing to our concentration on U.S. elections, much of our design space may not readily apply to multi-party systems—expanding data dimensionality could help. Additionally, the line between figurative and mechanistic analogies is blurry and worth exploring (for example, how should we classify a bingo ball blower? [3]). Future work could more formally define and assess how different types of analogies impact understanding or trust.

We are reaching the end of our journey. It is only now that we can truly see the forest for the trees. Each prior paper had its own focus, and this backstory paper allows us to step back, think, and reflect. Our methodological approach might be similar to those suggested for design studies [43, 47]. However, backed up by our sequels—our quantitative studies—we felt free to embrace a more narrative writing style than is typical. This way feels most authentic to us and more engaging for readers without sacrificing too much scientific merit. Initially, we were uncertain about how the community would react to our submission. The encouragement we received from reviewers has reassured us that this could be a promising way to share design studies and experiences. That said, we hope that our design space has illuminated new possibilities and that the insights from our interviews have left readers pondering. Beyond these, we hope that our backstory ignites readers to take on the challenge of designing uncertainty visualizations for broad audiences, and to unlock the potential of visualizations that engage, inform, and inspire.

ACKNOWLEDGEMENTS

Our deepest gratitude goes out to the extraordinary members of the MU Collective Lab (Angelos Chatzimparmpas, Yuan Cui, Lily W. Ge, Kelly Jiang, Paula Kayongo, Ziyang Guo, Maryam Hedayati, Jessica Hullman, Hyeok Kim, Taewook Kim, Sheng Long, Priyanka Nanayakkara, Abhraneel Sarma, and Dongping Zhang) as well as Ouxun Jiang and Steven Franconeri from the Visual Thinking Lab, without whom these papers would not have been possible. We would also like to extend our heartfelt thanks to Anna Wiederkehr for her invaluable feedback and guidance provided as a volunteer throughout the project. Their unwavering support and tireless efforts have been the guiding light that illuminated our journey. This research is supported by NSF 2127309 to the Computing Research Association for the CIFellows Project, and approved by Northwestern IRB.

Author contributions. All the authors conceptualized the work, designed the formative survey, crafted those interview questions, and discussed the results. Fumeng Yang prototyped the visualizations, coded the websites, collected and analyzed data, and prepared the manuscripts. Mandi Cai and Nicholas Diakopoulos helped edit the manuscripts and provided perspectives on political communication and journalism. Chloe Mortenson, Hoda Fakhari, Ayse D. Lokmanoglu, and Erik C. Nisbet provided their perspectives on political science. Matthew Kay provided supervision and prepared the manuscripts.

SUPPLEMENTAL MATERIALS

We archive all supplementary materials on Open Science Framework at `https://www.doi.org/10.17605/osf.io/ygq2v`. Our supplementary materials include (1) the original design sketches from our earlier stage; (2) the questionnaire and results of the formative study; (3) the code, screenshots, and videos of the prototypes; and (4) an example of the interview protocol. All prototypes implemented can be found at `https://forecasts.cs.northwestern.edu/2022-initial-prototypes`. The ten prototypes used in the interviews can be found at `https://forecasts.cs.northwestern.edu/2022-interview-prototypes`. The icons are designed by us or under a Flaticon license.

REFERENCES

[1] 2020 Election Forecast. https://projects.fivethirtyeight.com/2020-election-forecast. 2, 3
[2] 2022 Election Forecast. https://projects.fivethirtyeight.com/2022-election-forecast. 2, 3
[3] Canada poll tracker. https://newsinteractives.cbc.ca/elections/poll-tracker/canada/). 9
[4] Congested gulf of finland. https://tjukanov.org/gulfoffinland/. 3
[5] 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, 3
[6] Forecasting the U.S. elections. https://projects.economist.com/us-2020-forecast/president. 2
[7] Presidential plinko. https://presidential-plinko.com/. 3
[8] Who will win the presidency? https://projects.fivethirtyeight.com/2016-election-forecast/. 2, 5
[9] Who will win the senate? https://www.nytimes.com/newsgraphics/2014/senate-model/index.html. 2
[10] Why the election needle got it right but felt so wrong. https://gen.medium.com/why-the-election-needle-got-it-right-but-felt-so-wrong-e1a9c8ccb464. 2, 5
[11] Win probability visualization for election forecast models. https://electiondice.com/). 5
[12] 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
[13] 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
[14] 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
[15] 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
[16] M. Brehmer, B. Lee, P. Isenberg, and E. K. Choe. A comparative evaluation of animation and small multiples for trend visualization on mobile phones. *IEEE TVCG*, 26(1):364–374, 2020. doi: 10.1109/TVCG.2019.2934397 3
[17] M. Cai and M. Kay. Watching the election sausage get made: How data journalists visualize the vote counting process in U.S. elections. In *ACM CHI*, 2024. 1, 2
[18] 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
[19] J. Cherian and L. Bronner. An Update To The Washington Post Election Night Model, December 2020. 2
[20] 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
[21] M. Farjam. The bandwagon effect in an online voting experiment with real political organizations. *International Journal of Public Opinion Research*, 33(2):412–421, 06 2020. doi: 10.1093/ijpor/edaa008 2
[22] 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 2, 3, 9
[23] 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
[24] 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
[25] 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
[26] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011. doi: 10.1177/1473871611416549 3
[27] 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, 3
[28] 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
[29] 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, 3
[30] H. Ibrekk and M. G. Morgan. Graphical communication of uncertain quantities to nontechnical people. *Risk Analysis*, 7(4):519–529, 1987. doi: 10.1111/j.1539-6924.1987.tb00488.x 3
[31] T. W. S. Journal. Live senate election results 2022, 2022. 2
[32] 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 2, 3, 7
[33] B. Kaplan. Experimental electionomics: How election forecasts influence voter turnout, 2019. Bachelors dissertation. 2
[34] M. Kay. ggdist: Visualizations of distributions and uncertainty in the grammar of graphics. *IEEE TVCG*, 30(1):414–424, 2024. doi: 10.1109/TVCG.2023.3327195 3, 5
[35] 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, 9
[36] Y. Kim, M. Correll, and J. Heer. Designing animated transitions to convey aggregate operations. *Computer Graphics Forum*, 38(3):541–551, 2019. doi: doi.org/10.1111/cgf.13709 3
[37] R. Kosara, F. Bendix, and H. Hauser. Parallel sets: interactive exploration and visual analysis of categorical data. *IEEE TVCG*, 12(4):558–

568, 2006. doi: 10.1109/TVCG.2006.76 3
[38] 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
[39] 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 2
[40] L. Padilla, M. Kay, and J. Hullman. Uncertainty visualization. 2020. 2
[41] 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
[42] Reuters. U.S. election results, 2020. 2
[43] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE TVCG*, 18(12):2431–2440, 2012. doi: 10.1109/TVCG.2012.213 9
[44] 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 9
[45] S. Tribune. Minnesota election results 2020, 2020. 2
[46] T. W. G. van der Meer, A. Hakhverdian, and L. Aaldering. Off the Fence, Onto the Bandwagon? A Large-Scale Survey Experiment on Effect of Real-Life Poll Outcomes on Subsequent Vote Intentions. *International Journal of Public Opinion Research*, 28(1):46–72, 02 2015. doi: 10.1093/ijpor/edu041 2
[47] R. Vuillemot, P. Rivière, A. Beignon, and A. Tabard. Boundary objects in design studies: Reflections on the collaborative creation of isochrone maps. *Computer Graphics Forum*, 40(3):349–360, 2021. doi: 10.1111/cgf.14312 9
[48] 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 2, 5
[49] C. Xiong, A. Sarvghad, D. G. Goldstein, J. M. Hofman, and C. Demiralp. Investigating perceptual biases in icon arrays. In *ACM CHI*, article no. 137, 12 pages, 2022. doi: 10.1145/3491102.3501874 2
[50] F. Yang, M. Cai, C. Mortenson, H. Fakhari, A. D. Lokmanoglu, J. Hullman, S. Franconeri, N. Diakopoulos, E. C. Nisbet, and M. Kay. Swaying the public? impacts of election forecast visualizations on emotion, trust, and intention in the 2022 U.S. midterms. *IEEE TVCG*, 30(1):23–33, 2024. doi: 10.1109/TVCG.2023.3327356 1, 2, 9
[51] F. Yang, M. Hedayati, and M. Kay. Subjective probability correction for uncertainty representations. In *ACM CHI*, article no. 836, 17 pages, 2023. doi: 10.1145/3544548.3580998 1, 3, 9
[52] F. Yang, C. Mortenson, E. Nisbet, N. Diakopoulos, and M. Kay. In dice we trust: Uncertainty displays for maintaining trust in election forecasts over time. In *ACM CHI*, article no. 479, 2024. doi: 10.1145/3544548.3580998 9
[53] 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, 3