From Delays to Densities: Exploring Data Uncertainty through Speech, Text, and Visualization

Chase Stokes[1,2] , Chelsea Sanker[3] , Bridget Cogley[4] , and Vidya Setlur[1]

[1]Tableau Research, Palo Alto, USA
[2]University of California, Berkeley, USA
[3]Stanford University, Palo Alto, USA
[4]Versalytix, USA

Abstract
Understanding and communicating data uncertainty is crucial for making informed decisions in sectors like finance and healthcare. Previous work has explored how to express uncertainty in various modes. For example, uncertainty can be expressed visually with quantile dot plots or linguistically with hedge words and prosody. Our research aims to systematically explore how variations within each mode contribute to communicating uncertainty to the user; this allows us to better understand each mode's affordances and limitations. We completed an exploration of the uncertainty design space based on pilot studies and ran two crowdsourced experiments examining how speech, text, and visualization modes and variants within them impact decision-making with uncertain data. Visualization and text were most effective for rational decision-making, though text resulted in lower confidence. Speech garnered the highest trust despite sometimes leading to risky decisions. Results from these studies indicate meaningful trade-offs among modes of information and encourage exploration of multimodal data representations.

CCS Concepts
• *Human-centered computing* → *Human computer interaction (HCI); Empirical studies in HCI;* *Empirical studies in visualization;*

1. Introduction

In today's world of data-driven decision-making, effectively communicating the uncertainty inherent to the underlying information is important [Dra95]. *Data uncertainty* refers to the range of potential outcomes, variability within a dataset, or possible error in measurements or predictions [MH90]. While precise data may be ideal for making decisions, such data is uncommon in real-life decisions. Communicating uncertainty can allow for a better understanding of the true state of the data. However, recognizing and effectively communicating data uncertainty poses several challenges.

A key challenge in data uncertainty is interpretation. While experts might understand statistical nuances like confidence intervals or p-values, a lay audience might misread these indicators and make incorrect conclusions [SPS11]. This misinterpretation significantly affects decisions in vital areas like medicine, finance, and public policy [MDL08]. Conveying data uncertainty also affects trust, at times enhancing it [SSR21, SSK*16]. However, uncertain data might also lead the audience to view the information as unreliable; despite uncertainty being a normal part of data analysis [CJTD21].

Researchers have developed various methods to assist in under-

standing statistical uncertainty. Visual tools like error bars, confidence intervals, and density plots illustrate data variability, scope, and distribution [MRH*05, PKH21]. In written texts, uncertainty is indicated through the use of hedge words such as "somewhat" and "possibly" [Lak73, DH01]. In spoken communication, additional features like pitch and speech rate can signal a speaker's uncertainty or hesitation [BW95, SLW73, SK05].

There are tradeoffs, however, in the effectiveness of each mode of communication in conveying uncertainty. Viewers may not possess sufficient graphical literacy to understand complex visualizations [MRH*05], and textual explanations may be necessary to clarify visual nuances. On the other hand, readers might not follow lengthy explanations, resulting in limited comprehension [Sch97]. Finally, the transient nature of speech limits the ability to revisit information compared to text or visualizations [SW95].

Determining how speech, text, or visualizations present information requires a nuanced understanding of how readers interpret information in each mode. A careful exploration of each mode can help identify effective techniques for data communication. As multimedia representations beyond visualizations emerge, it is impor-

tant to develop a clear understanding of how other modes of information compare to visual representations.

Contributions. To this end, we explore representations of data uncertainty (speech, text, and visualization) along a spectrum of concreteness to fuzziness to help identify effective strategies and trade-offs in communicating data uncertainty for decision-making. Two pilot studies outline a detailed design space for communicating data uncertainty. We focus specifically on unimodal representations to first understand what each mode can uniquely offer, an essential step before we can understand the implications of multimodality.

We conduct a comparative analysis of the efficacy and effects of these modes and fuzziness within each one, using two experiments. While this analysis partially replicates the methods of previous work, it also allows for direct comparison across modalities, leading to new findings. Unlike prior work, we examine these modes under a unified framework and uniform experimental conditions.

The results indicate a high degree of trust in speech representations despite relatively irrational decision-making and a low rating of confidence in decisions made with text representations despite relatively rational decision-making.

2. Related Work

Our work builds on research in communicating uncertainty in data visualization, text, and speech, as well as decision-making.

2.1. Communicating Uncertainty in Data Visualization

Developing techniques for data visualization to address data uncertainty is an important research problem, as the interpretability of uncertain data can significantly impact decision-making [JS03, Hul16, PCH21] in domains such as climate and geospatial modeling [MRH*05, Pan01, KCG15, NCA*19, DM20, WLWC23], medicine [RPHL14, HSSV22], and business intelligence applications [VKKG17]. For example, employing point estimates as part of uncertainty communication has been found to improve decision-making in contexts such as weather and transit. Surveys of uncertainty visualization [BHJ*14, PKH21, SLSR08] and tools [JED*20] for exploring the various approaches underscore the extensive scope and impact of research.

Important research has explored information visualization techniques for communicating uncertainty. Thomson et al. [THM*05] present a typology of uncertainty, delineating kinds of uncertainty matched with space, time, and attribute components of data. These concepts from visual semiotics are applied to characterizing visual signification appropriate for representing different categories of uncertainty [MRO*12]. Research findings generally indicate that enhancing visualizations and data together with their uncertainty information improves users' understanding of complex information.

Some visualization techniques can better support data uncertainty. Görtler et al. [GSWD17] introduce bubble treemaps that encode uncertainty using wave-like modifications and blur effects. Sane et al. [SAJ21] extend univariate confidence isosurfaces to multivariate feature level sets and visualize regions with uncertainty in relation to the specific trait or feature. Ensemble datasets contain a collection of estimates for each simulation variable, and ensemble visualizations representing a sample of projections better support understanding of data uncertainty [LBR*16, PRCR17].

2.2. Communicating Uncertainty in Text

Communicating uncertainty in text draws from research in linguistics, human-computer interaction, and information sciences [Tou03]. Much of this work examines words and phrases that signal uncertainty in text, often called *hedges*. Hedges indicate degrees of uncertainty and concreteness, such as 'sort of,' 'perhaps,' 'might,' and 'could be' [Lak73, DH01, SVF*12, RLK06]. These words can indicate various kinds of uncertainty, e.g., probabilities for future events, matters of opinion, and information that is open to multiple interpretations [SVF*12, DH01].

Computational linguistics research has implemented techniques for identifying hedging patterns to determine their effectiveness in communicating a particular point of view to the reader [IXM20, GCBR*21]. Hedge words decrease the perceived soundness of an argument [BH05, SC93] and the credibility of the author [SAC98]. Accordingly, participants are more likely to be persuaded by an argument without hedges [DBRS08]. However, this finding may interact with how much confidence might reasonably be expected; descriptions of medical research are rated as more trustworthy when they include hedges [Jen08]. The presence of hedges is often determined by inherent data uncertainty, as occurs in predictions of probabilistic future events [SVF*12, DH01].

2.3. Communicating Uncertainty in Speech

In addition to the same lexical and semantic factors that are often studied in written text, the acoustic characteristics of speech provide an additional dimension to communicating uncertainty.

Confidence is correlated with faster speech [SLW73, SC93, BW95], higher intensity [SLW73], and fewer, shorter pauses [SLW73]. Confident utterances are also more likely to exhibit falling intonation, while uncertain utterances are more likely to exhibit rising intonation [SC93, BW95, SK05]. Some studies find a lower pitch with more confident utterances [JP15], while others find a higher pitch [SLW73]. These acoustic cues can also influence listeners' *perception* of a speaker's certainty. Perceived confidence increases with faster speech rate [BW95, PBS11, GFVJ19, KLSG22], higher intensity [JP17], lower pitch [JP17, GFVJ19, KLSG22], falling pitch [BW95, LW10, GFVJ19], fewer and shorter pauses [LW10, PBS11, JP17]. Predictors of perceived confidence have the same effects in synthetic speech [KLSG22].

Most of the existing work on the relationship between certainty and acoustic characteristics of speech addresses the *speaker's* confidence, e.g., based on instructions to speak confidently or doubtfully [SLW73, JP15] or when answering trivia questions that they feel more or less confident about [SC93, BW95]. The work in perception of these acoustic characteristics similarly asks listeners to evaluate how confident a speaker sounds [BW95, PBS11, JP17, KLSG22], rather than evaluating certainty of the information itself.

In contrast, our work examines how acoustic characteristics, such as changes in speech rate, pitch, and pauses, might influence

listeners' understanding of the uncertainty in probabilistic data and the way that they make decisions based on that data.

2.4. Decision-making and Reasoning with Uncertain Information

Perceived uncertainty and reliability of incoming information can influence the decisions that people make based on that information [HQC*19]. For instance, people are also more likely to seek more information when they are less confident in their decision about it [DBY18].

There is evidence that some uncertainty visualizations (i.e., dot plots, probability density functions) allow people to better aggregate across varying estimates [GJS*18]. Kim et al. [KWKH19, KKGMH20] looked into Bayesian reasoning scenarios, examining how different visualization techniques affect users' interpretive accuracy. Other research found that inferential uncertainty led to more overestimation of effect sizes than visualizations that show outcome variability [HGH20]. When uncertainty visualizations emphasize arithmetic means, users overlook uncertainty information and misinterpret visual distance as a proxy for effect size [KKH21].

Research has studied users' decision-making processes for specific domains. Korporaal et al. investigated how data uncertainty in maps might influence the process of spatial decision-making [KRF20]. Viewing COVID-19 visualizations with rising trends increased participants' beliefs that they and others were at risk [PHF*22]. More salient visualizations of uncertainty may lead to decreased willingness to follow COVID-19 forecasts [LM21].

In linguistically signaled uncertainty, the degree of confidence expressed in text or spoken utterance also influences decision-making. People are more likely to be persuaded by an argument that is expressed with moderate confidence than one that is expressed with very low or high confidence [LMT71]. Statements with moderate confidence are also rated as being more credible [CBD09].

Specific acoustic correlates of confidence have been found to predict how listeners handle information; arguments given at faster speech rates are perceived as more credible and are more likely to persuade listeners than arguments with slower speech rates [MMBV76, MW69]. Higher intensity and more variation in intensity are also positive predictors of how persuasive an utterance will be [VZB20]. Lower pitch produces more positive attitudes towards the message being communicated [CDRS03].

3. Research Goals

The complexity of uncertain data necessitates the careful presentation and communication of information. Different modes of data representation can present different *kinds* of information with trade-offs in their effectiveness for communicating uncertainty. For example, speech information contains signals beyond what text alone can provide - the pitch and duration of certain words can communicate nuances that text cannot. Visualizations provide more information than can be concisely represented in text or speech formats.

Our work aims to guide the effective use of these modes and provide insight for future work in multimodal uncertainty communication as well as situations where visual representations may not be feasible or useful to a given audience. We are specifically interested in the following research questions:

- **RQ1**: *How does the mode of information presentation affect decision-making with uncertain data?*
- **RQ2**: *How does decision-making change when indicators of uncertainty are intensified within each mode of communication?*

We explored the design space of uncertainty representation in each mode of information, determining factors that increase perceptions of uncertainty. We then conducted two crowdsourced experiments. The first experiment compared how three different modes (speech, text, and visualization) communicate uncertain information. The second experiment compared decisions within each mode when those indicators of uncertainty were heightened.

4. Design Space of Visualization, Text, and Speech Stimuli

As part of investigating differences *between* modes of communication, we varied the stimuli *within* each mode on a scale from concrete to fuzzy, adapted from Setlur & Cogley [SC22].

Vagueness or *fuzziness* occurs when the boundaries of meaning are not precisely defined, leading to ambiguity and multiple possible interpretations. Fuzziness is often unavoidable when distinctions are gradient rather than clear-cut categories. *Concreteness* refers to a specific, clear, and unambiguous concept. Concrete representation relies on definite and precise terms that clearly delineate their meaning, leaving little room for ambiguity. Both concepts play a crucial role in communication, with fuzziness allowing for flexibility and adaptability, while concreteness ensures clarity and precision [Sor97, SC22]. In the context of this paper, we use "precision level" to refer to the variable capturing the fuzziness of stimuli.

While there has been work on basic representations of uncertainty for each mode, as detailed in Section 2, we expand that design space to include parameters of vagueness and uncertainty to better understand their influence on uncertainty communication and decision-making. We considered three main visualization types: beeswarm plots, quantile dot plots, and density plots, based on an informal survey of commonly used uncertainty visualizations and insights from previous research [KKHM16, FWM*18, PPKH21]. We varied components of these visualizations, including fuzziness, opacity, and arrangement [PKH21]. We also included a variation in color, rendering the visualizations in orange, rather than gray (e.g., a density pot with an orange fill) [SJW00, TT14]. We analyzed a total of 21 visualizations, including two beeswarm plots, 10 density plots, five quantile dot plots, and four combinations of density and dot plots. Density plots are a common depiction of uncertainty, but we also investigated variations in frequency-based visualizations by adjusting the number of quantiles, color, and the overlap of other plot types. This resulted in a set of 21 different visualizations illustrating the design space we were interested in exploring.

For text variants, we considered the frequency and visual treatment of hedge words as well as the overall tone. Hedge words are key indicators of uncertainty in language, and this uncertainty may be emphasized by using uncertain visual appearances. We also included a "colloquial" condition, which used more informal language and everyday conversational expressions. We considered two

Table 1: *Stimulus categories and examples for Experiment 1 (2 × 3 study design). Hedge terms in the fuzzy text template are bolded to highlight the differences here but did not receive any visual treatment within the experiment.*
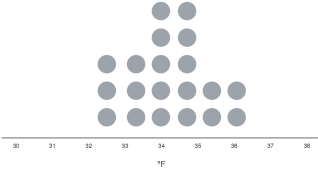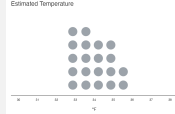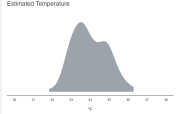
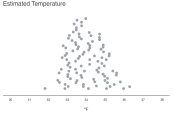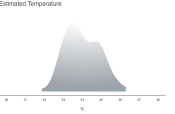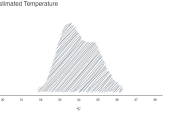|	Concrete	Fuzzy
Visual		
Text	The most likely temperature low tonight is [mean]ºF. There is a 50% chance that the temperature will fall between [50% interval min] and [50% interval max]ºF. While the range of possible lows spans [min] to [max]ºF, those extremes are less likely. It is also [significantly/no modifier/somewhat/slightly] more likely to be [warmer than cooler/cooler than warmer] within that range.	The most likely temperature low tonight **might** be around [mean]ºF. There is a 50% chance that the temperature **could** fall between [50% interval min] and [50% interval max]ºF. While the range of possible lows **could potentially** span [min] to [max]ºF, those extremes **seem** less likely. It also **appears** [significantly/no modifier/somewhat/slightly] more likely to be [warmer than cooler/cooler than warmer] within that range.
Speech	Baseline Speech (MP3 Link)	0.2s delay prior to numerical value, 65% speed on numbers, 70% speed on hedge or likelihood terms, 5% pitch decrease on numbers and hedge or likelihood terms (MP3 Link)

Table 2: *Stimulus categories and examples for Experiment 2. Cells with a light gray background indicate stimuli that overlapped with Experiment 1 (E1). A sample of the text stimuli is provided, and visual treatments shown here were applied to the stimuli in the survey.*

	Most Concrete	**Somewhat Concrete**	**Mixed**	**Somewhat Fuzzy**	**Most Fuzzy**
Visual					
Text (Sample)	The most likely temperature low tonight is [mean]ºF.	The most likely temperature low tonight might be [mean]ºF.	The most likely temperature low tonight might be around [mean]ºF.	The most likely temperature low tonight could potentially hover somewhere around [mean]ºF.	The most likely temperature low tonight *could potentially* hover *somewhere around* [mean]ºF.
Speech	Concrete Text, No Modifications (MP3 Link)	Concrete Text, E1 Modifications (MP3 Link)	Fuzzy Text, No Modifications (MP3 Link)	Fuzzy Text, E1 Modifications (MP3 Link)	Fuzzy Text, E1 Modifications and Question Contour (MP3 Link)

visual treatments: italics and gray text. This resulted in a set of 15 text variants. The speech design space was defined by variations in the duration of words and pauses, absolute pitch, and pitch contour. We selected these features as acoustic correlates of uncertainty found in earlier work (see Section 2.3). We included variations in the content of the speech, resulting in 13 total variants.

To investigate these variants and identify a manageable design space of parameters for our experiments, the two pilot studies were conducted on Prolific [PS18], the first with 20 participants and the second with 40. Participants were fluent in English, had an approval rate above 95%, and had normal color vision. They were introduced to the concepts of "concrete" and "fuzzy" information.

Participants ranked sets of stimuli on a scale from concrete to fuzzy. The first pilot tested nine sets (4 visualization, 2 text, 3 speech), and the second pilot tested seven (4 visualization, 1 text, 2 speech). Since we could only make direct comparisons between stimuli in the same ranking set, we used the second pilot to compare new stimuli combinations, primarily narrowing the visualization sets. Participants initially completed a practice ranking where

the items guided their rankings (e.g., "This item should be ranked first"). These rankings also served as a response quality check.

These pilot studies provided information to operationalize precision within individual communication modes. This method is not sufficient to standard uncertainty *between* the different modes but instead assesses the impact of varying signals of uncertainty within each mode affect perceptions of fuzziness. These pilot studies ensured that we selected experimental stimuli with perceptible variations, which effectively varied the uncertainty signals in each mode.

In selecting stimuli for the experiments, we carefully evaluated stimuli based on their average fuzziness rankings from the pilot studies. To do this, we placed the variants for each mode in their relative position on a spectrum and discussed the set of possible stimuli as a team. We chose stimuli that represented a range of most to least fuzzy, used diverse techniques in depicting uncertainty, were still within the realm of practical representations, and had basis in prior literature. This variety ensured a resulting set of stimuli with different precision levels and methods of representing fuzziness.

4.1. Visualization Stimuli

In the first pilot, participants perceived the quantile dot plot as fuzzier than the density plot. Their qualitative responses indicated they were unfamiliar with quantile dot plots. In the second pilot, after providing an explanation for quantile dot plots, participants ranked them as more concrete than density plots. The use of orange did not alter fuzziness rankings, but a lighter gray was ranked fuzzier than a darker gray. Gradient fills and sketchy styles heightened fuzziness as well, consistent with known visualization uncertainty encoding methods [MRO*12, FPS*21, WII*12].

For Experiment 1 (E1), we selected stimuli representative of the kinds of information displays often used in practice. We informally surveyed uncertainty visualizations in media and news reporting and incorporated feedback from a co-author with 15+ years of experience as a visualization designer. After this discussion, we selected the dark gray density plot, a common uncertainty representation, and the 20-quantile dot plot. While quantile dot plots required explanation to understand, prior work has indicated that they are an effective way to communicate uncertainty in a direct or frequency-based way [FWM*18, KKHM16, PPKH21]. These were representative of visualizations studied in prior work and displayed a difference in fuzziness ratings. For Experiment 2, we expand the range of fuzziness examined to include the beeswarm chart, a gradient density plot, and a density plot with a sketched style.

When creating these stimuli, we selected data from a normal distribution, using the `rnorm` function in R (v4.3.1) [R C21]. We selected 100 data points from a normal distribution with a standard deviation of 1 [PPKH21]. True standard deviations ranged from 0.89 to 1.13. Averages ranged from 29.95 to 34.18ºF. This variation allowed a more ecologically valid set of distributions while still maintaining comparability; natural data will often not be perfectly normally distributed. Unless otherwise noted, all visualization stimuli were created using the `ggdist` package in R [Kay23] with additional Figma modifications for the gradient and sketched designs. Stimuli code is included in supplementary materials.

4.2. Text Stimuli

In text stimuli, more hedge words and visual emphases, like italics or gray color, increased perceived fuzziness. Stimuli with a colloquial tone received similar rankings as variants which used a more formal tone with a medium amount of hedge words. Interestingly, a passage with fewer but emphasized hedge words was ranked on average as fuzzy as one with more hedge words.

In Experiment 1, we selected the variants without hedges and with medium hedging. In Experiment 2, we expanded this range to include minor, substantial, and italicized substantial hedging. Text stimuli contained the most likely temperature (the rounded mean value), the middle 50% range, the rounded minimum and maximum value, and a description of the distribution skew, calculated using `skewness` in the `moments` package in R [KN22].

4.3. Speech Stimuli

When ranking speech stimuli, increased and decreased pitches were viewed similarly, but a 10% pitch change was slightly fuzzier than a 5% change. However, 55% of participants in the first pilot did not rank 0% and 10% pitch changes differently. Pitch contour was the most evident fuzzy indicator, followed by delay, then average pitch. Minor delays were perceived similarly to no delays, and significant delays matched E1 delays. Layering speech modifications, such as combining pitch with delay, consistently amplified fuzziness.

For Experiment 1, we selected the default output with concrete text and a variant with delays, reduced rates, and minor pitch decreases with fuzzy text. The speech alterations applied only to number and/or hedge words. In Experiment 2, we added variants to further capture the impact of content and speech attributes, including concrete text with E1 modifications, fuzzy text with no modifications, and fuzzy text with E1 modifications and question contour. As some acoustic correlates of uncertainty are also correlates of emphasis, we used a lower pitch modification to use the acoustic cues as signaling uncertainty rather than emphasis; a higher pitch is a main correlate of emphasis [LP84, KS11], but results for its relationship to confidence are varied [JP15, SLW73]. Speech stimuli were created using Google Speech Synthesis Markup Language (SSML) [Goo23], a standardized markup language that allows adjustments in synthesized speech. Experiment 1 stimuli are shown in Table 1, and Experiment 2 stimuli are shown in Table 2.

4.4. Discussion

In this work, we mapped out a design landscape focused on the transmission of data uncertainty through various modes, examining how various techniques and visualization types influence the user's perception of uncertainty. This exploration of the design space provides two tangible contributions to the study of data uncertainty.

Signals of uncertainty. Through these pilot studies, we identified and analyzed signals of uncertainty not evaluated in prior uncertainty research, such as text color and formatting. While prior work has outlined the mechanisms for emphasizing uncertainty in visual representations [PKH21], it has not assessed how uncertain those representations *appear* to readers. These pilot studies demonstrate one approach to comparing data representations and mechanisms

of uncertainty. The findings from this design space exploration offer valuable guidance for professionals and further insight into the use of various techniques to emphasize uncertainty.

While the evaluation completed here was sufficient to select diverse, meaningful, and practical representations of uncertainty for the following experiments, future research should explore a more explicit and validated measure of how people interpret and react to uncertainty cues, including a clear mapping of perceptual thresholds across modes. Standardizing levels of fuzziness could provide more precise comparisons and a clearer understanding of how scaling uncertainty impacts decision-making across modes.

Implications for education. Our study highlights the crucial role of providing explanations for visual representations of uncertainty, underscoring the need for strategies that teach individuals how to interpret these visualizations effectively. This insight highlights the important use of clear and accessible explanations alongside visualizations and opens several avenues for future research and development in educational strategies and tools. Future research should focus on developing educational materials and methods that enhance comprehension of visual uncertainty, exploring which explanations are most effective for different audiences, and how these teachings can be integrated into various learning environments.

5. Experiment Overview

In Experiment 1, we compared six combinations of information modes (speech, text, visualization) and precision (concrete, fuzzy) for a 3×2 mixed-design study with the decision-making task. In Experiment 2, we examined the same three modes but with additional stimuli varying in precision. Understanding the unique capabilities of each mode is vital before we can assess how to merge them effectively in practical settings. This work provides a better understanding of how to approach situations where visualizations may not be effective and what alternatives could be employed in those instances. This dual focus not only maps out part of the space of unimodal representations but also guides the selection of the most appropriate communication mode based on the context.

The two experiments used a consistent decision structure from prior work [JL12, PPKH21, NGJ09, SJ13]. Participants were provided a data-driven task to decide whether to salt the road based on the evening's low-temperature forecast. This task was recognizable and relevant to participants, yet not so personal as to introduce subjective bias in their decisions. An objective, rational choice existed, determined by the cost and penalty values in the experiment.

Participants were presented with the following scenario: "In cold weather, roads may need to be treated with salt to prevent icing. This salt treatment is costly but not as costly as damages caused by ice forming on roads. You are in charge of a road maintenance company contracted to treat the roads in a U.S. town with salt to prevent icing. It is your job to apply salt to the roads when the temperature is at or below 32ºF to prevent ice from forming."

Participants were provided a fictional budget of \$12,000 to complete 12 trials. Applying salt cost \$1,000, and the penalty for road freezing was \$3,000. Based on these values, the objectively rational choice was to salt when the chance of the temperature falling below 32ºF was at or above 33%. Participants received a \$0.05 bonus for every \$1,000 left in their budget at the end of twelve trials.

These experiments evaluated decision-making within this context, including confidence in the decisions and trust in the forecast. Participants' confidence levels shed light on their personal feelings during the decision-making process, as well as some insight into the perceived quality of their decision. As trust may alleviate the hesitations and doubts inherent in uncertain scenarios, collecting this information is key to a more comprehensive understanding of the trade-offs between different information modes.

Confidence was assessed by asking participants to indicate the chance that their choice was correct, ranging from 50% to 100% [TKH08] ("How confident are you that you made the correct choice?"). We used a multi-item measure for "trust," consisting of usefulness, clarity, and accuracy [EGKX22, XPGF19, PMCO23].

5.1. Participants

For both Experiments 1 and 2, we utilized the G*Power software [FELB07, FEBL09] for power analysis, aiming to achieve a power of 0.95 with an alpha threshold of 0.05. For Experiment 1, we estimated an effect size of 0.2 based on pilot studies and previous research [JL12, PPKH21]. For Experiment 2, we estimated an effect size of 0.15 based on the results from Experiment 1. Based on the power analysis, the necessary sample size was 105 participants for Experiment 1 and 129 participants per mode for Experiment 2.

Participants were recruited via the Prolific platform [PS18]. Participants were omitted for irrelevant free-response answers. We also used an attention check for every trial, which queried the likelihood of a value falling outside the forecast's distribution range and thus with an expected likelihood of 0%. If participants estimated a likelihood above 50%, we discarded all their data. We set this threshold at 50% rather than lower in order to accommodate potential challenges in probabilistic reasoning in text and speech scenarios.

To account for potential unusable responses, we recruited 20 more participants for each study than the power analysis indicated was necessary. 130 participants were recruited for Experiment 1, and 450 participants (150 per mode) were recruited for Experiment 2. The task took about 18 minutes, and participants received a payment of \$3.60. After exclusions, this resulted in 109 participants in Experiment 1. In Experiment 2, there were 128 participants in the speech condition, 132 in the text condition, and 136 in the visualization condition. Participants shared their age range and education; the demographic patterns are provided in supplementary materials.

5.2. Task Design

Experiments 1 and 2 used the same set of tasks. Participants were randomly assigned to a condition and then introduced to both the task and their assigned forecast type. This introduction explained potential decision outcomes and specifics of the assigned forecast. For those in the visualization group, there was a brief guide on interpreting the visual. Those in the speech group were informed they could replay the forecast multiple times if necessary. After this introduction, participants went through 12 randomized trials. In each trial, participants viewed a forecast, made a decision, rated their

confidence, estimated the evening's temperature and the likelihood of freezing, and answered the attention check.

Following the 12 trials, participants evaluated the forecast type they had been exposed to, assessing its clarity, accuracy, and usefulness. They also answered open-ended questions on their decision-making strategy and what they liked/disliked about the forecast. Participants also completed a four-question measure of graphical literacy [OJGW19, BXF*20, CQHP21]. The survey concluded with demographic questions collecting age range and education level. Additionally, participants ranked their preferred way to receive information, choosing between visual (visualizations, images), written (books, articles), or audio (podcasts, radio broadcasts).

5.3. Analysis

Methods and analyses were preregistered on OSF for both Experiment 1 and Experiment 2. The analyses presented here deviate from preregistration; the initial plan included models with demographic variables that, upon further reflection and feedback, were not tied to specific hypotheses. As such, we adjusted our approach to evaluate only those models that incorporated directly relevant variables.

Both experiments investigated four dimensions of decision-making: the crossover temperature or turning point for the decision, the rationality of the decision, participants' confidence in their decision, and their overall trust in the information provided. Most of the results are based on mixed effects regression models using the `lmer` package in R, with model output generated by the `stargazer` package [Hla22]. Results for decision rationality are based on χ^2 tests. All analysis materials, including full model results and statistical tests, are reported in supplementary materials.

6. Experiment 1: Evaluation of Modes for Uncertainty Decision-making

The goal of Experiment 1 was to examine **RQ1**: *How does the mode of information presentation (speech, text, and visualization) affect decision-making with uncertain data?*, and how signals of uncertainty (precision) within each mode affect participant decisions.

6.1. Hypotheses

People process information provided in text differently than speech [FGG90, KKS*75]. Using text, probabilities expressed with hedge words produce more conservative decisions than numeric probabilities [NGJ09]. For predictions using likelihood *intervals*, text is more effective than visualizations [SJ13]. However, visualizations typically outperform text in both challenging [CBK*16] and simpler tasks [MLB*20]. Prior research found that visualizations allowing frequency reasoning, like quantile dot plots, resulted in better decisions and higher confidence than those that did not, such as density plots [KKHM16]. For each of the decision dimensions listed in Section 5.3, we assessed the effects of information mode (speech, text, or visualization) and precision (concrete vs. fuzzy).

Determining crossover temperature. First, we evaluated the *crossover temperature*. Crossover temperature is the turning point temperature at which participants were equally likely to salt or not

salt. The *optimal* crossover temperature is the point at which it becomes rational to salt the roads, determined by first identifying the freeze probability at which participants would ideally start salting, which is equivalent to the Cost:Penalty ratio (1:3 or 33%.) If the provided data suggested a freezing likelihood above 33%, salting the roads becomes the rational choice. We calculated the crossover temperature for each distribution by determining the mean value at which the distribution had a 33% chance of 32ºF or below. Optimal crossover temperatures ranged from 32.3 to 32.6ºF.

H1a: Visualization forecasts have a crossover temperature closest to optimal, followed by text, then speech. **H1b**: Concrete forecasts have a crossover temperature closer to optimal than fuzzy forecasts.

Decision rationality. H2a: Visualization forecasts have more rational decisions than text or speech. No difference between text and speech. **H2b**: Concrete forecasts have more rational decisions than fuzzy forecasts.

Decision confidence. H3a: Visualization forecasts have the highest confidence ratings, followed by text, then speech. **H3b**: Concrete forecasts have higher confidence ratings than fuzzy forecasts.

Trust in forecast. Finally, we examined trust ratings. As trust is a complex process, we examined three dimensions of trust: usefulness, clarity, and accuracy [EGKX22, XPGF19]. Visual representations afford probabilistic reasoning more than their text and speech counterparts. This increased information may increase clarity and perceived reliability. A concrete representation may also increase clarity and reliability in comparison to a fuzzy representation.

H4a: Visualization forecasts have the highest trust ratings, followed by text, then speech. **H4b**: Concrete forecasts have higher trust ratings than fuzzy forecasts.

6.2. Results

In general, participants performed well in the decision-making task, with $2,531 remaining on average. Most participants played the speech stimuli only once (72% of trials). The results indicated that both visualizations and text supported rational decision-making, but text led to lower confidence ratings. Speech, while associated with riskier decisions, led to higher trust ratings than the other two modes. Precision did not appear to have an impact on decisions.

6.2.1. H1: Decision Crossover Temperature

H1a and **H1b** were evaluated with logistic mixed effects models predicting the binary salting decision: "Do not salt" (0) or "Salt" (1). Consistent with prior work, we used these predictions to calculate the deviations from optimal crossover temperatures [PPKH21]. These calculations accounted for variations between distributions, as they were based on the distance from the mean of the distribution to the optimal crossover temperature for each distribution.

We compared models using an ANOVA test for model selection. The optimal model ($p = 0.014$) included a random effect of participant and fixed effects of the difference between the average temperature and the optimal crossover, and the mode of information.

Table 3 illustrates the differences from the optimal crossover temperature. For speech and text stimuli, people were more likely

to be risky than conservative, indicated by the negative values in each cell, and vice versa for visualization stimuli. Perfectly optimal decision-making would result in a value of 0.

We found partial support for **H1a**. Visualization forecasts had a crossover temperature closer to optimal than speech, resulting from an increased likelihood of salting the roads (2.7 - 7.6x more likely, $p = 0.004$). There was a minor difference between text and speech, with text decisions being more likely to salt (1.6 - 4.6x more likely, $p = 0.055$). There was no difference between text and visualization ($p = 0.324$). Including precision did not improve model performance, so we do not find support for **H1b**. The crossover temperature pattern was similar to the binary salting decisions.

Table 3: *Difference from optimal crossover temperature. Darker gray indicates a more optimal crossover temperature.*

	Concrete	**Fuzzy**
Speech	-0.517	-0.404
Text	-0.151	-0.038
Visualization	0.026	0.139

6.2.2. H2: Decision Rationality

The relatively close match of crossover temperatures to optimal values suggests a strong pattern of rational decision-making. Participants' decisions were grouped into three categories: rational, conservative, or risky. Conservative decisions salted the roads despite the probability of freezing less than 33%. Risky decisions did not salt despite the probability of freezing greater than 33%. We examined **H2a** and **H2b** about decision rationality using χ^2 tests.

Decisions based on visualization forecasts were more often rational compared to the more risky decisions based on speech forecasts, providing partial support for **H2a** ($\chi^2 = 30.1$, $p < 0.01$), as decisions with text forecasts exhibited intermediate rationality. **H2b** was not supported ($\chi^2 = 1.16$, $p = 0.559$), similar to with results from **H1**. Figure 1 shows the decision types by condition.
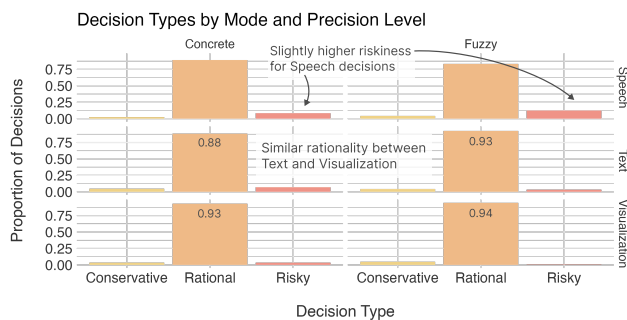

Figure 1: *Proportion of decision types for each condition in Experiment 1. Overall, decisions were mostly rational. Speech was the least rational mode, with a greater proportion of risky decisions.*

6.2.3. H3: Decision Confidence

Participant confidence ratings are shown in Figure 2. We compared relevant models using an ANOVA. The optimal model ($p = 0.014$) included a random effect for participant and fixed effects of the difference between mean temperature and the optimal crossover, decision rationality, and mode of information. Including precision did not improve model performance; we did not find support for **H3b**. We found partial support for **H3a** - visualization forecasts produced higher confidence ratings compared to text ($p = 0.005$) by 7-15%. There was no notable difference between speech and text ($p = 0.131$) nor speech and visualization ($p = 0.188$).
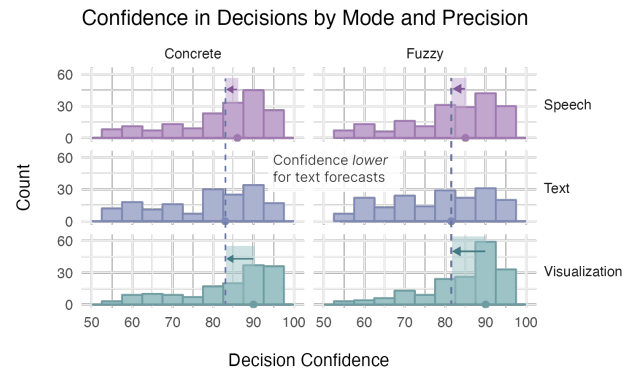

Figure 2: *Experiment 1 confidence ratings ranged from 50 to 100. Confidence was lower for text than for visualization forecasts.*

6.2.4. H4: Trust in Forecast

Finally, we conducted an analysis of trust ratings. We used the same ANOVA process to select the optimal model. The final model ($p = 0.003$) included the rationality of the decision and the mode of the forecast. Precision (concrete vs. fuzzy) did not significantly improve the model ($p = 0.984$), so it was not included as a factor. Thus, we did not find support for **H4b**. The results of this model *contradicted* **H4a**: speech forecasts were the most trusted, with 9-17% higher ratings than text forecasts ($p < 0.01$) and 6-14% higher ratings than visualization forecasts ($p = 0.013$).
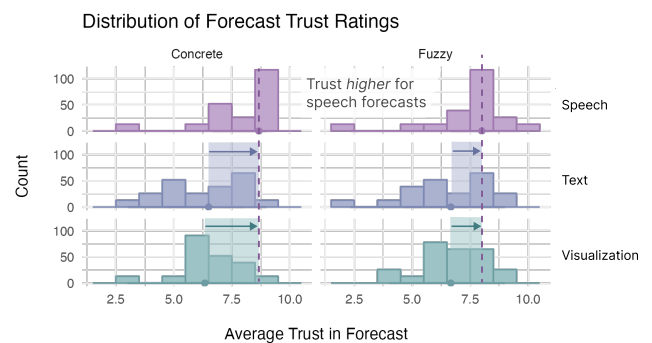

Figure 3: *Experiment 1 trust ratings. Participants trusted speech the most in comparison to both text and visualization.*

6.3. Discussion

The findings of this study offer insights into how participants interpret uncertain information presented in different modes.

High trust in speech. Speech forecasts received the highest trust ratings (Figure 3); speech also led to a decreased likelihood to salt the roads and thus riskier decisions than visualizations. This combination of results presents a complex trade-off: while speech was perceived as more trustworthy, it did not guide participants as accurately as other modes. When supported by other modes, speech may improve trust without sacrificing decision-making.

Lower confidence with text. Text forecasts were a strong mode for conveying outcomes and uncertainty. Decisions made using text were on par with visualizations in terms of crossover temperature (Table 3) and rationality (Figure 1). However, participants were less confident with text-based forecasts than with visualizations. Given that text can act as a strong support to decision-making, it may be best used when integrated with other communicative mediums.

7. Experiment 2: Effect of Mode Design Factors on Uncertainty Decision-making

We aimed to further explore variations within each mode to better encompass the design space outlined in Section 4. In Experiment 2, we asked - **RQ2**: *How does decision-making change when signs of uncertainty are intensified within each mode of communication?*

7.1. Hypotheses

Experiment 2 investigated the same decision dimensions described in Section 5.3. We compared precision levels of stimuli within each mode. Hypotheses are based on trends observed in Experiment 1.

Determining crossover temperature. H5a-b: Crossover temperature is closer to optimal as **(a) speech** and **(b) text** forecasts increase in fuzziness. **H5c**: Crossover temperature is further from optimal as **visualization** forecasts increase in fuzziness.

Decision rationality. H6a-b: Decisions are *more* frequently rational as **(a) speech** and **(b) text** forecasts increase in fuzziness. **H6c**: Decisions are *less* frequently rational as **visualization** forecasts increase in fuzziness.

Decision confidence. H7a-c: Confidence in the decision increases as **all three modes** increase in fuzziness.

Trust in forecast. H8a-c: Trust in the forecast decreases **all three modes** increase in fuzziness.

7.2. Results

As in Experiment 1, participants generally performed well at the task, with \$1,737 remaining on average. Most participants played the speech stimuli only once (79% of trials). Figures for rationality, confidence, and trust can be found in supplemental materials.

7.2.1. H5: Determining Crossover Temperature

As in Experiment 1, we used ANOVA for model selection. Including precision as a factor did not improve model performance for the speech ($p = 0.957$), text ($p = 0.884$), or visualization models ($p = 0.133$). We did not find support for an effect of precision on the likelihood of salting and crossover temperatures (**H5a-c**).

Participants make riskier decisions in this experiment than in Experiment 1, as almost all crossover temperatures were positive (Table 4). This may be related to the differences in the participant pool or specific distributions, as they differed from Experiment 1. Additionally, decisions were closer to optimal with speech forecasts than in the prior experiment, especially in comparison to text forecasts.

7.2.2. H6: Decision Rationality

Decisions were again categorized as rational, conservative, or risky and evaluated using χ^2 tests. There were no significant differences in the rationality of decisions based on precision for speech ($\chi^2 = 4.07$, $p = 0.850$) or text stimuli ($\chi^2 = 5.40$, $p = 0.714$); there was no support for **H6a-b**. However, there was a significant effect of precision on rationality with visualization stimuli ($\chi^2 = 15.9$, $p = 0.0443$). Participants made more conservative decisions in the somewhat fuzzy condition (density plot with gradient) than in the most concrete condition (quantile dot plot). However, there was no consistent effect of precision, providing unclear support for **H6a**.

7.2.3. H7: Decision Confidence

As in Experiment 1, we examined participants' confidence in their decisions. We used ANOVA testing for model selection; Including precision as a factor did not improve model performance for the speech ($p = 0.656$), text ($p = 0.244$), or visualization model ($p = 0.509$). As such, we did not find support for **H7a-c**. We did not find a connection between decision confidence and forecast fuzziness.

7.2.4. H8: Trust in Forecast

Finally, we conducted an analysis of participants' trust ratings for the forecast. Precision did not significantly improve model performance for speech ($p = 0.133$) or visualization ($p = 0.080$); we did not find support for **H8a** or **H8c**. For text stimuli, the best model ($p = 0.012$) included decision rationality and stimulus precision. Most concrete was the most trusted, while most fuzzy was less trusted than other variants. While the presence of hedge words influenced trust, there was no evidence for a continuous pattern based on the hedge frequency. However, emphasizing these hedge words by italicizing them *decreased* trust. This finding is contrary to **H8b**.

7.3. Effects of Mode

The Experiment 1 conditions were also present in Experiment 2, so we ran the analyses from Experiment 1 on the Experiment 2 data to test the consistency of findings across the experiments. We consider concrete and fuzzy categories from Experiment 1; results largely support our initial findings, with variation in decision quality.

We did *not* find that participants were less likely to apply salt with speech forecasts than with text ($p = 0.209$) or visualizations ($p = 0.884$). Likewise, we did not find that speech led to riskier decision-making ($\chi^2 = 7.00$, $p = 0.136$). We did observe a difference by precision; fuzzy forecasts across modes led to more conservative decisions than concrete ($\chi^2 = 7.22$, $p = 0.027$).

Table 4: *Difference from optimal crossover temperature. Darker gray indicates a more optimal crossover temperature.*

	Most concrete	Somewhat concrete	Mixed	Somewhat fuzzy	Most fuzzy
Speech	0.054	0.029	−0.011	0.089	0.113
Text	0.281	0.439	0.380	0.338	0.367
Visualization	−0.015	0.148	0.110	0.312	0.209

We again found a lower confidence level for text, this time in comparison to speech ($p = 0.002$) rather than visualization. ($p = 0.186$). Despite relatively similar decision quality, text again received lower confidence ratings. We also found that speech forecasts had higher trust ratings than visualization ($p = 0.043$).

7.4. Discussion

When comparing the decisions made with these variants within each mode, we found little effect of fuzziness. There was some effect of hedging on trust and a connection between visualization fuzziness and conservative decision-making. However, these results were not consistent across variants and did not provide a clear effect of precision. These results may reflect a true lack of effect of precision on decision-making; they may also be due to the selected manipulations of fuzziness; more extreme variants (e.g., significant pitch changes) may result in different effects. We constrained our stimuli set to realistic variants for ecological validity.

Between the two experiments, we found that, despite similar success at the task, participants were riskier in Experiment 1 and more conservative in Experiment 2. This may explain some of the differences between the results for each study; the impact of fuzziness may vary based on how participants were approaching the task. However, the findings in trust and confidence were consistent.

8. Discussion and Future Work

Our exploration of speech, text, and visualization for conveying data uncertainty revealed several variations in decision-making.

RQ1: *Comparison of Modes - How does the mode of information presentation affect decision-making with uncertain data?* Visualization and text forecasts supported rational decision-making, while speech introduced a slightly increased level of risk. Decisions using text received lower confidence ratings than visualizations. Speech received higher trust ratings than the other modes.

RQ2: *Comparison of Precision - How does decision-making change when indicators of uncertainty are intensified within each mode of communication?* Increasing indicators of uncertainty did not have a measurable impact on decisions. The confidence and trust findings from Experiment 1 were supported in Experiment 2.

8.1. Limitations and Future Directions

Explore multimodal representations. In this work, we uncovered key trade-offs for individual modes. Speech representations can lead to higher levels of trust but may also increase irrational decision-making. Text representations provide support for rational decisions but lower confidence in these decisions. Visualizations are useful for decision-making overall but can be difficult to interpret and require additional explanation. While we did not investigate multimodal representations, our results suggest a need for further research on how people make decisions when presented with uncertainty data in multimodal contexts.

Expand the design space. The primary focus of our work was to understand the impact of different modes on the perception of data uncertainty. By keeping elements like visualization style, font, and speaker characteristics constant, the study aimed to isolate the effects of the mode of information itself. Exploring additional elements of each mode would substantially increase the complexity of the study design space and is outside the scope of our study. Future research could explore how additional design factors such as the speaker's (perceived) gender or accent [Gil70, MR77, SRG85] might affect the perception of uncertainty.

Explore additional uncertainty tasks. Our work considered one type of decision-making task as described in [PCRHS18]. While other work in uncertainty has explored other kinds of decisions (e.g., transportation [KKHM16]), these studies typically examine only one mode of information rather than comparing different representations. Future studies could diversify scenarios to include personal experiences with related activities, such as living in cold climates or financial decisions (e.g., funding a startup) to better understand how modes affect different categories of decision-making.

Supporting accessibility. Future research should address how to support accessibility when designing multimodal interfaces for data uncertainty. While the current study did not specifically delve into evaluation based on accessibility, designing equitable interfaces is an important research direction to consider and one that these findings support; accessible interfaces will provide flexibility for users to choose modes that best align with their abilities and preferences.

9. Conclusion

Effectively communicating and understanding data uncertainty is essential for making everyday decisions and in domains such as finance, healthcare, and policy-making. This work explores how different modes - speech, text, and visualization, can be used to convey uncertainty. From hedge words and prosody variants to quantile dot and density plots, we explore a design space within each mode. Through two crowdsourced studies, we identify how the mode of information influences decision-making and trust when reasoning with uncertain data. While visualization and text best support decision-making, users found speech more trustworthy and had lower confidence in decisions made with text. These findings suggest future research directions that explore a more nuanced approach to uncertainty communication.

References

[BH05] BLANKENSHIP K. L., HOLTGRAVES T.: The Role of Different Markers of Linguistic Powerlessness in Persuasion. *Journal of Language and Social Psychology 24*, 1 (2005), 3–24. 2

[BHJ*14] BONNEAU G.-P., HEGE H.-C., JOHNSON C. R., DE OLIVEIRA NETO M. M., POTTER K. C., RHEINGANS P., SCHULTZ T.: Overview and State-of-the-Art of Uncertainty Visualization. In *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization* (London, 2014), Hansen C. D., Chen M., Johnson C. R., Kaufman A. E., Hagen H., (Eds.), Springer, pp. 3–27. 2

[BW95] BRENNAN S. E., WILLIAMS M.: The Feeling of Another's Knowing: Prosody and Filled Pauses as Cues to Listeners About the Metacognitive States of Speakers. *Journal of Memory and Language 34* (1995), 383–398. 1, 2

[BXF*20] BURNS A., XIONG C., FRANCONERI S., CAIRO A., MAHYAR N.: How to Evaluate Data Visualizations across Different Levels of Understanding. In *2020 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)* (Piscataway, NJ, 2020), IEEE, pp. 19–28. 7

[CBD09] CRAMER R. J., BRODSKY S. L., DECOSTER J.: Expert Witness Confidence and Juror Personality: Their Impact on Credibility and Persuasion in the Courtroom. *Journal of the American Academy of Psychiatry and the Law Online 37*, 1 (2009), 63–74. 3

[CBK*16] CHEONG L., BLEISCH S., KEALY A., TOLHURST K., WILKENING T., DUCKHAM M.: Evaluating the Impact of Visualization of Wildfire Hazard upon Decision-making Under Uncertainty. *International Journal of Geographical Information Science 30*, 7 (2016), 1377–1404. 7

[CDRS03] CHATTOPADHYAY A., DAHL D. W., RITCHIE R. J., SHAHIN K. N.: Hearing Voices: The Impact of Announcer Speech Characteristics on Consumer Response to Broadcast Advertising. *Journal of Consumer Psychology 13*, 3 (2003), 198–204. 3

[CJTD21] CAPURRO G., JARDINE C. G., TUSTIN J., DRIEDGER M.: Communicating Scientific Uncertainty in a Rapidly Evolving Situation: A Framing Analysis of Canadian Coverage in Early Days of COVID-19. *BMC Public Health 21*, 1 (2021), Article 2181. 1

[CQHP21] CASTRO S. C., QUINAN P. S., HOSSEINPOUR H., PADILLA L.: Examining Effort in 1d Uncertainty Communication Using Individual Differences in Working Memory and NASA-TLX. *IEEE Transactions on Visualization and Computer Graphics 28*, 1 (2021), 411–421. 7

[DBRS08] DURIK A. M., BRITT M. A., REYNOLDS R., STOREY J.: The Effects of Hedges in Persuasive Arguments: A Nuanced Analysis of Language. *Journal of Language and Social Psychology 27*, 3 (2008), 217–234. 2

[DBY18] DESENDER K., BOLDT A., YEUNG N.: Subjective Confidence Predicts Information Seeking in Decision Making. *Psychological Science 29*, 5 (2018), 761–778. 3

[DH01] DE HAAN F.: The Relation between Modality and Evidentiality. *Linguistische Berichte 9* (2001), 201–16. 1, 2

[DM20] DING Q., MILLET B.: Visualizing Uncertainty in Weather Forecasts. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 64* (12 2020), 1064–1068. doi:10.1177/1071181320641255. 2

[Dra95] DRAPER D.: Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society, Series B (Methodological) 57* (1995), 45–97. doi:10.2307/2346087. 1

[EGKX22] ELHAMDADI H., GABA A., KIM Y.-S., XIONG C.: How Do We Measure Trust in Visual Data Communication? In *2022 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)* (Piscataway, NJ, 2022), IEEE, pp. 85–92. 6, 7

[FEBL09] FAUL F., ERDFELDER E., BUCHNER A., LANG A.-G.: Statistical Power Analyses using G* Power 3.1: Tests for Correlation and Regression Analyses. *Behavior Research Methods 41*, 4 (2009), 1149–1160. 6

[FELB07] FAUL F., ERDFELDER E., LANG A.-G., BUCHNER A.: G* Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behavior Research Methods 39*, 2 (2007), 175–191. 6

[FGG90] FURNHAM A., GUNTER B., GREEN A.: Remembering Science: The Recall of Factual Information as a Function of the Presentation Mode. *Applied Cognitive Psychology 4*, 3 (1990), 203–212. 7

[FPS*21] FRANCONERI S. L., PADILLA L. M., SHAH P., ZACKS J. M., HULLMAN J.: The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest 22*, 3 (2021), 110–161. 5

[FWM*18] FERNANDES M., WALLS L., MUNSON S., HULLMAN J., KAY M.: Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), Association for Computing Machinery, pp. 1–12. 3, 5

[GCBR*21] GOODLUCK CONSTANCE T., BAJAJ N., RAJWADI M., MALTBY H., WALL J., MONIRI M., WOODRUFF C., LAIRD T., LAIRD J., GLACKIN C., CANNINGS N.: Resolving Ambiguity in Hedge Detection by Automatic Generation of Linguistic Rules. In *International Conference on Artificial Neural Networks* (2021), Springer-Verlag, p. 369–380. 2

[GFVJ19] GUYER J. J., FABRIGAR L. R., VAUGHAN-JOHNSTON T. I.: Speech Rate, Intonation, and Pitch: Investigating the Bias and Cue Effects of Vocal Confidence on Persuasion. *Personality and Social Psychology Bulletin 45*, 3 (2019), 389–405. 2

[Gil70] GILES H.: Evaluative Reactions to Accents. *Educational Review 22*, 3 (1970), 211–227. 10

[GJS*18] GREIS M., JOSHI A., SINGER K., SCHMIDT A., MACHULLA T.: Uncertainty Visualization Influences How Humans Aggregate Discrepant Information. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), CHI '18, Association for Computing Machinery, p. 1–12. doi:10.1145/3173574.3174079. 3

[Goo23] GOOGLE: Speech Synthesis Markup Language (SSML) Reference, 2023. Google Cloud Documentation. URL: https://cloud.google.com/text-to-speech/docs/ssml. 5

[GSWD17] GÖRTLER J., SCHULZ C., WEISKOPF D., DEUSSEN O.: Bubble Treemaps for Uncertainty Visualization. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (08 2017), 719–728. doi:10.1109/TVCG.2017.2743959. 2

[HGH20] HOFMAN J. M., GOLDSTEIN D. G., HULLMAN J.: How Visualizing Inferential Uncertainty Can Mislead Readers About Treatment Effects in Scientific Results. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2020), CHI '20, Association for Computing Machinery, p. 1–12. doi:10.1145/3313831.3376454. 3

[Hla22] HLAVAC M.: *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Social Policy Institute, Bratislava, Slovakia, 2022. R package version 5.2.3. URL: https://CRAN.R-project.org/package=stargazer. 7

[HQC*19] HULLMAN J., QIAO X., CORRELL M., KALE A., KAY M.: In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (jan 2019), 903–913. doi:10.1109/TVCG.2018.2864889. 3

[HSSV22] HOEFER M. J. D., SCHUMACHER B. E., SZAFIR D. A., VOIDA S.: Visualizing Uncertainty in Multi-Source Mental Health Data. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2022), CHI EA '22, Association for Computing Machinery. doi:10.1145/3491101.3519844. 2

[Hul16] HULLMAN J.: Why Evaluating Uncertainty Visualization is

Error Prone. In *BELIV '16: Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (10 2016), Association for Computing Machinery, pp. 143–151. doi:10.1145/2993901.2993919. 2

[IXM20] ISLAM J., XIAO L., MERCER R. E.: A Lexicon-Based Approach for Detecting Hedges in Informal Text. In *Proceedings of the 12th Language Resources and Evaluation Conference* (Marseille, France, 2020), European Language Resources Association, pp. 3109–3113. 2

[JED*20] JENA A., ENGELKE U., DWYER T., RAIAMANICKAM V., PARIS C.: Uncertainty Visualisation: An Interactive Visual Survey. In *2020 IEEE Pacific Visualization Symposium (PacificVis)* (2020), IEEE, pp. 201–205. doi:10.1109/PacificVis48177.2020.1014. 2

[Jen08] JENSEN J. D.: Scientific Uncertainty in News Coverage of Cancer Research: Effects of Hedging on Scientists' and Journalists' Credibility. *Human Communication Research 34*, 3 (2008), 347–369. 2

[JL12] JOSLYN S. L., LECLERC J. E.: Uncertainty Forecasts Improve Weather-Related Decisions and Attenuate the Effects of Forecast Error. *Journal of Experimental Psychology: Applied 18*, 1 (2012), 126–140. 6

[JP15] JIANG X., PELL M. D.: On How the Brain Decodes Vocal Cues About Speaker Confidence. *Cortex 66* (2015), 9–34. 2, 5

[JP17] JIANG X., PELL M. D.: The Sound of Confidence and Doubt. *Speech Communication 88* (2017), 106–126. 2

[JS03] JOHNSON C., SANDERSON A.: A Next Step: Visualizing Errors and Uncertainty. *Computer Graphics and Applications, IEEE 23* (09 2003), 6–10. doi:10.1109/MCG.2003.1231171. 2

[Kay23] KAY M.: *ggdist: Visualizations of Distributions and Uncertainty*. Northwestern University, 2023. R package version 3.3.0. URL: https://mjskay.github.io/ggdist/, doi:10.5281/zenodo.3879620. 5

[KCG15] KOO H., CHUN Y., GRIFFITH D. A.: Geovisualization of Attribute Uncertainty. *Journal of Visual Languages & Computing 44* (2015), 89–96. 2

[KKGMH20] KIM Y.-S., KAYONGO P., GRUNDE-MCLAUGHLIN M., HULLMAN J.: Bayesian-assisted Inference from Visualized Data. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2020), 989–999. 3

[KKH21] KALE A., KAY M., HULLMAN J.: Visual Reasoning Strategies for Effect Size Judgments and Decisions. *IEEE Transactions on Visualization & Computer Graphics 27*, 02 (Feb 2021), 272–282. doi:10.1109/TVCG.2020.3030335. 3

[KKHM16] KAY M., KOLA T., HULLMAN J. R., MUNSON S. A.: When (Ish) is My Bus? User-Centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2016), Association for Computing Machinery, pp. 5092–5103. 3, 5, 7, 10

[KKS*75] KINTSCH W., KOZMINSKY E., STREBY W. J., MCKOON G., KEENAN J. M.: Comprehension and Recall of Text as a Function of Content Variables. *Journal of Verbal Learning and Verbal Behavior 14*, 2 (1975), 196–214. 7

[KLSG22] KIRKLAND A., LAMERIS H., SZÉKELY E., GUSTAFSON J.: Where's the Uh, Hesitation? The Interplay Between Filled Pause Location, Speech Rate and Fundamental Frequency in Perception of Confidence. In *Proceedings of Interspeech* (Incheon, Korea, 2022), Interspeech, pp. 4990–4994. 2

[KN22] KOMSTA L., NOVOMESTKY F.: *Moments, Cumulants, Skewness, Kurtosis and Related Tests*. CRAN, 2022. R package version 0.14.1. URL: https://cran.r-project.org/web/packages/moments/moments.pdf. 5

[KRF20] KORPORAAL M., RUGINSKI I. T., FABRIKANT S. I.: Effects of Uncertainty Visualization on Map-Based Decision Making Under Time Pressure. *Frontiers of Computer Science 2* (2020), Article 32. 3

[KS11] KATZ J., SELKIRK E.: Contrastive Focus vs. Discourse-new: Evidence From Phonetic Prominence in English. *Language 87*, 4 (2011), 771–816. 5

[KWKH19] KIM Y.-S., WALLS L. A., KRAFFT P., HULLMAN J.: A Bayesian Cognition Approach to Improving Data Visualization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–14. 3

[Lak73] LAKOFF G.: Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic 2*, 4 (1973), 458–508. doi:10.1007/BF00262952. 1, 2

[LBR*16] LIU L., BOONE A., RUGINSKI I., PADILLA L., HEGARTY M., CREEM-REGEHR S., THOMPSON W., YUKSEL C., HOUSE D.: Uncertainty Visualization by Representative Sampling from Prediction Ensembles. *IEEE transactions on visualization and computer graphics PP* (09 2016). doi:10.1109/TVCG.2016.2607204. 2

[LM21] LEFFRANG D., MÜLLER O.: Should I Follow this Model? The Effect of Uncertainty Visualization on the Acceptance of Time Series Forecasts. In *2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX)* (2021), IEEE, pp. 20–26. doi:10.1109/TREX53765.2021.00009. 3

[LMT71] LONDON H., MCSEVENEY D., TROPPER R.: Confidence, Overconfidence and Persuasion. *Human Relations 24*, 5 (1971), 359–369. 3

[LP84] LIBERMAN M., PIERREHUMBERT J.: Intonational Invariance Under Changes in Pitch Range and Length. In *Language Sound Structure. Studies in phonology presented to Morris Halle*, Aronoff M., Oehrle R. T., (Eds.). MIT Press, Cambridge, MA, 1984, pp. 157–233. 5

[LW10] LASARCYK E., WOLLERMANN C.: Do Prosodic Cues Influence Uncertainty Perception in Articulatory Speech Synthesis? In *Seventh ISCA Workshop on Speech Synthesis* (Kyoto, Japan, 2010), nternational Symposium on Computer Architecture (ISCA), pp. 230–235. 2

[MDL08] MORSS R., DEMUTH J. L., LAZO J. K.: Communicating Uncertainty in Weather Forecasts: A Survey of the U.S. Public. *Weather and Forecasting 23* (2008), 974–991. doi:10.1175/2008WAF2007088.1. 1

[MH90] MORGAN M. G., HENRION M.: *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, 1990. URL: https://api.semanticscholar.org/CorpusID:57858998. 1

[MLB*20] MULDER K. J., LICKISS M., BLACK A., CHARLTON-PEREZ A. J., MCCLOY R., YOUNG J. S.: Designing Environmental Uncertainty Information for Experts and Non-experts: Does Data Presentation Affect Users' Decisions and Interpretations? *Meteorological Applications 27*, 1 (2020), Article e1821. 7

[MMBV76] MILLER N., MARUYAMA G., BEABER R. J., VALONE K.: Speed of Speech and Persuasion. *Journal of Personality and Social Psychology 34*, 4 (1976), 615–624. 3

[MR77] MULAC A., RUDD M. J.: Effects of Selected American Regional Dialects upon Regional Audience Members. *Communications Monographs 44*, 3 (1977), 185–195. 10

[MRH*05] MACEACHREN A., ROBINSON A., HOPPER S., GARDNER S., MURRAY R., GAHEGAN M., HETZLER E.: Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science 32* (07 2005), 139–160. doi:10.1559/1523040054738936. 1, 2

[MRO*12] MACEACHREN A. M., ROTH R. E., O'BRIEN J., LI B., SWINGLEY D., GAHEGAN M.: Visual Semiotics & Uncertainty Visualization: An Empirical Study. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2496–2505. 2, 5

[MW69] MEHRABIAN A., WILLIAMS M.: Nonverbal Concomitants of Perceived and Intended Persuasiveness. *Journal of Personality and Social psychology 13*, 1 (1969), 37–58. 3

[NCA*19] NAGEL F., CASTIGLIA G., ADEMAJ G., BUCHMLLER J., SCHLEGEL U., KEIM D. A.: cpmViz: A Web-Based Visualization Tool

for Uncertain Spatiotemporal Data. In *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2019), IEEE, pp. 140–141. doi:10.1109/VAST47406.2019.8986941. 2

[NGJ09] NADAV-GREENBERG L., JOSLYN S. L.: Uncertainty Forecasts Improve Decision Making Among Nonexperts. *Journal of Cognitive Engineering and Decision Making 3*, 3 (2009), 209–227. 6, 7

[OJGW19] OKAN Y., JANSSEN E., GALESIC M., WATERS E. A.: Using the Short Graph Literacy Scale to Predict Precursors of Health Behavior Change. *Medical Decision Making 39*, 3 (2019), 183–195. 7

[Pan01] PANG A. T.: Visualizing Uncertainty in Geospatial Data. In *Proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology* (2001). URL: https://api.semanticscholar.org/CorpusID:16285011. 2

[PBS11] PON-BARRY H., SHIEBER S. M.: Recognizing Uncertainty in Speech. *EURASIP Journal on Advances in Signal Processing 2011* (2011), Article 251753. 2

[PCH21] PADILLA L., CASTRO S. C., HOSSEINPOUR H.: Chapter Seven - A Review of Uncertainty Visualization Errors: Working Memory as an Explanatory Theory. In *The Psychology of Learning and Motivation*, Federmeier K. D., (Ed.), vol. 74 of *Psychology of Learning and Motivation*. Academic Press, Cambridge, MA, 2021, pp. 275–315. URL: https://www.sciencedirect.com/science/article/pii/S0079742121000074, doi:10.1016/bs.plm.2021.03.001. 2

[PCRHS18] PADILLA L. M., CREEM-REGEHR S. H., HEGARTY M., STEFANUCCI J. K.: Decision Making with Visualizations: A Cognitive Framework Across Disciplines. *Cognitive Research: Principles and Implications 3*, 1 (2018), Article 29. 10

[PHF*22] PADILLA L., HOSSEINPOUR H., FYGENSON R., HOWELL J., CHUNARA R., BERTINI E.: Impact of COVID-19 Forecast Visualizations on Pandemic Risk Perceptions. *Scientific Reports 12*, 1 (Feb. 2022), Article 2014. Funding Information: The National Science Foundation supported this work with Grant 2028374. Publisher Copyright: © 2022, The Author(s). doi:10.1038/s41598-022-05353-1. 3

[PKH21] PADILLA L., KAY M., HULLMAN J.: Uncertainty Visualization. In *Wiley StatsRef: Statistics Reference Online*. Wiley, 02 2021, pp. 1–18. doi:10.1002/9781118445112.stat08296. 1, 2, 3, 5

[PMCO23] PANDEY S., MCKINLEY O. G., CROUSER R. J., OTTLEY A.: Do You Trust What You See? Toward A Multidimensional Measure of Trust in Visualization. *IEEE Transactions on Visualization and Computer Graphics* (2023). 6

[PPKH21] PADILLA L. M., POWELL M., KAY M., HULLMAN J.: Uncertain about Uncertainty: How Qualitative Expressions of Forecaster Confidence Impact Decision-making with Uncertainty Visualizations. *Frontiers in Psychology 11* (2021), 579267. 3, 5, 6, 7

[PRCR17] PADILLA L., RUGINSKI I., CREEM-REGEHR S.: Effects of Ensemble and Summary Displays on Interpretations of Geospatial Uncertainty Data. *Cognitive Research: Principles and Implications 2* (10 2017). doi:10.1186/s41235-017-0076-1. 2

[PS18] PALAN S., SCHITTER C.: Prolific.ac — A Subject Pool for Online Experiments. *Journal of Behavioral and Experimental Finance 17* (2018), 22–27. 4, 6

[R C21] R CORE TEAM: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. https://www.R-project.org/. 5

[RLK06] RUBIN V., LIDDY E., KANDO N.: Certainty Identification in Texts: Categorization Model and Manual Tagging Results. In *Computing Attitude and Affect in Text: Theory and Applications. The Information Retrieval Series*, Shanahan J. G., Qu Y., Wiebe J., (Eds.), vol. 20. Springer, Dordrecht, 01 2006, pp. 61–76. doi:10.1007/1-4020-4102-0_7. 2

[RPHL14] RISTOVSKI G., PREUSSER T., HAHN H. K., LINSEN L.: Uncertainty in Medical Visualization: Towards a Taxonomy. *Computers &*

Graphics 39 (2014), 60–73. URL: https://www.sciencedirect.com/science/article/pii/S0097849313001568, doi:10.1016/j.cag.2013.10.015. 2

[SAC98] SPARKS J. R., ARENI C. S., COX K. C.: An Investigation of the Effects of Language Style and Communication Modality on Persuasion. *Communications Monographs 65*, 2 (1998), 108–125. 2

[SAJ21] SANE S., ATHAWALE T. M., JOHNSON C. R.: Visualization of Uncertain Multivariate Data via Feature Confidence Level-Sets. In *Eurographics Conference on Visualization* (2021), The Eurographics Association, pp. 43–47. 2

[SC93] SMITH V. L., CLARK H. H.: On the Course of Answering Questions. *Journal of Memory and Language 32* (1993), 25–38. 2

[SC22] SETLUR V., COGLEY B.: *Functional Aesthetics for Data Visualization*. John Wiley & Sons, 2022. 3

[Sch97] SCHRIVER K. A.: *Dynamics in Document Design: Creating Text for Readers*. John Wiley & Sons, Inc., USA, 1997. 1

[SJ13] SAVELLI S., JOSLYN S.: The Advantages of Predictive Interval Forecasts for Non-expert Users and the Impact of Visualizations. *Applied Cognitive Psychology 27*, 4 (2013), 527–541. 6, 7

[SJW00] SMITH-JACKSON T. L., WOGALTER M. S.: Users' Hazard Perceptions of Warning Components: An Examination of Colors and Symbols. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (2000), vol. 44, SAGE Publications, pp. 6–55. 3

[SK05] SWERTS M., KRAHMER E.: Audiovisual Prosody and Feeling of Knowing. *Journal of Memory and Language 53* (2005), 81–94. 1, 2

[SLSR08] SKEELS M., LEE B., SMITH G., ROBERTSON G.: Revealing Uncertainty for Information Visualization. In *AVI '08: Proceedings of the working conference on advanced visual interfaces* (05 2008), vol. 9, pp. 376–379. doi:10.1057/ivs.2009.1. 2

[SLW73] SCHERER K. R., LONDON H., WOLF J. J.: The Voice of Confidence: Paralinguistic Cues and Audience Evaluation. *Journal of Research in Personality 7* (1973), 31–44. 1, 2, 5

[Sor97] SORENSEN R.: Vagueness. In *Stanford Encyclopedia of Philosophy*, Zalta E. N., (Ed.). Stanford University, 1997. 3

[SPS11] SPIEGELHALTER D., PEARSON M., SHORT I.: Visualizing Uncertainty About the Future. *Science 333*, 6048 (2011), 1393–1400. doi:10.1126/science.1191181. 1

[SRG85] STEWART M. A., RYAN E. B., GILES H.: Accent and Social Class Effects on Status and Solidarity Evaluations. *Personality and Social Psychology Bulletin 11*, 1 (1985), 98–105. 10

[SSK*16] SACHA D., SENARATNE H., KWON B. C., ELLIS G., KEIM D. A.: The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 240–249. doi:10.1109/TVCG.2015.2467591. 1

[SSR21] STEIJAERT M. J., SCHAAP G., RIET J. V.: Two-sided Science: Communicating Scientific Uncertainty Increases Trust in Scientists and Donation Intention by Decreasing Attribution of Communicator Bias. *Communications 46*, 2 (2021), 297–316. 1

[SVF*12] SZARVAS G., VINCZE V., FARKAS R., MÓRA G., GUREVYCH I.: Cross-genre and Cross-domain Detection of Semantic Uncertainty. *Computational Linguistics 38*, 2 (2012), 335–367. 2

[SW95] SPERBER D., WILSON D.: *Relevance: Communication and Cognition*. Blackwell, Oxford, 1986/1995. 1

[THM*05] THOMSON J. R., HETZLER E. G., MACEACHREN A. M., GAHEGAN M., PAVEL M.: A Typology for Visualizing Uncertainty. In *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE) 5669* (2005), Erbacher R. F., Roberts J. C., Grohn M. T., Borner K., (Eds.), SPIE, pp. 146–157. 2

[TKH08] TSAI C. I., KLAYMAN J., HASTIE R.: Effects of Amount of Information on Judgment Accuracy and Confidence. *Organizational Behavior and Human Decision Processes 107*, 2 (2008), 97–105. 6

[Tou03] TOULMIN S. E.: *The Uses of Argument*, second ed. Cambridge University Press, Cambridge, 2003. `doi:10.1017/CBO9780511840005`. 2

[TT14] TAK S., TOET A.: Color and Uncertainty: It is Not Always Black and White. In *Eurographics Conference on Visualization (EuroVis)* (2014), The Eurographics Association, pp. 55–59. 3

[VKKG17] VOSOUGH Z., KAMMER D., KECK M., GROH R.: Visualizing Uncertainty in Flow Diagrams: A Case Study in Product Costing. In *Proceedings of the 10th International Symposium on Visual Information Communication and Interaction* (New York, NY, USA, 2017), VINCI '17, Association for Computing Machinery, pp. 1–8. `doi:10.1145/3105971.3105972`. 2

[VZB20] VAN ZANT A. B., BERGER J.: How the Voice Persuades. *Journal of Personality and Social Psychology 118*, 4 (2020), 661–682. 3

[WII*12] WOOD J., ISENBERG P., ISENBERG T., DYKES J., BOUKHELIFA N., SLINGSBY A.: Sketchy Rendering for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2749–2758. 5

[WLWC23] WITT J. K., LABE Z. M., WARDEN A. C., CLEGG B. A.: Visualizing Uncertainty in Hurricane Forecasts with Animated Risk Trajectories. *Weather, Climate, and Society 15*, 2 (2023), 407 – 424. `doi:10.1175/WCAS-D-21-0173.1`. 2

[XPGF19] XIONG C., PADILLA L., GRAYSON K., FRANCONERI S.: Examining the Components of Trust in Map-based Visualizations. In *1st EuroVis Workshop on Trustworthy Visualization, TrustVis 2019* (Genova, Italy, 2019), The Eurographics Association, pp. 19–23. 6, 7