

Mixing Modes: Active and Passive Integration of Speech, Text, and Visualization for Communicating Data Uncertainty

Chase Stokes^{1,2} , Chelsea Sanker³ , Bridget Cogley⁴ , and Vidya Setlur¹ 

¹Tableau Research, Palo Alto, USA

²University of California, Berkeley, USA

³Stanford University, Palo Alto, USA

⁴Versalytix, USA

Abstract

Interpreting uncertain data can be difficult, particularly if the data presentation is complex. We investigate the efficacy of different modalities for representing data and how to combine the strengths of each modality to facilitate the communication of data uncertainty. We implemented two multimodal prototypes to explore the design space of integrating speech, text, and visualization elements. A preliminary evaluation with 20 participants from academic and industry communities demonstrates that there exists no one-size-fits-all approach for uncertainty communication strategies; rather, the effectiveness of conveying uncertain data is intertwined with user preferences and situational context, necessitating a more refined, multimodal strategy for future interface design. Materials for this paper can be found on [OSF](#).

CCS Concepts

• **Human-centered computing** → **Visualization; Interaction techniques;**

1. Introduction

Communicating data uncertainty can be challenging, especially for audiences with limited statistical expertise [MRH*05, Hul19]. However, accurately communicating uncertainty is critical for decision-making, as it could impact risk assessments.

Prior work has explored visual representations for uncertain data [KKHM16, FWM*18, PPKH21]. Other modes of communication, such as text or speech explanations, can mitigate some of these issues but have limitations. For instance, audiences with limited graphical literacy more easily interpret text, but text cannot convey statistical information. [Sch97]. Speech conveys tone not present in written text, but is ephemeral and poses challenges for conveying complex information [SW95]. Integrating visual, textual, and speech elements may use one method's strengths to offset another's limitations, facilitating more effective data communication.

Depending on the user's context, different modes may be used for decision-making. For example, text-to-speech features can be useful when the user's hands are occupied, such as while driving. However, these passive contexts limit users' ability to further interact with the data. We implemented two multimodal interface types: passive, which provided an integrated presentation of information while minimizing user effort, and active, which emphasized user-driven interaction and in-depth exploration [EFN12]. We then completed a preliminary evaluation of the prototypes, focusing on design implications and future research directions.

2. Related Work

2.1. Representations of Uncertainty

A range of studies have examined the use of uncertainty visualizations, from surveys [BHJ*14, PKH21, SLSR08] to specialized tools [JED*20] and libraries [Kay23]. This research includes applications such as weather forecasting [PPKH21] and public transportation [FWM*18, KKHM16]. Different kinds of uncertainty visualizations have been evaluated for decision-making [PPKH21, FWM*18, KKHM16].

Text representations of uncertainty often use hedges to signal uncertainty, such as 'might' or 'possibly' [Lak73, SVF*12]. These hedges can reflect either inherent uncertainty in the data being described or the author's uncertainty about it. Text descriptions can also convey statistics about uncertain outcomes.

Hedges are also used in speech along with phonetic patterns like speech rate, pauses, and pitch [SLW73, JP15]. Uncertain speech often has rising intonation [SC93], slower speech rate [SLW73, SC93], and more frequent pauses [SLW73]. Some studies find a higher pitch associated with lower confidence [JP15], while others find a lower pitch associated with lower confidence [SLW73]. These acoustic characteristics can also impact how confident or uncertain a listener deems the speaker [KLSG22].

Building upon this rich body of research, our work further explores the integration of speech, text, and visualization through the

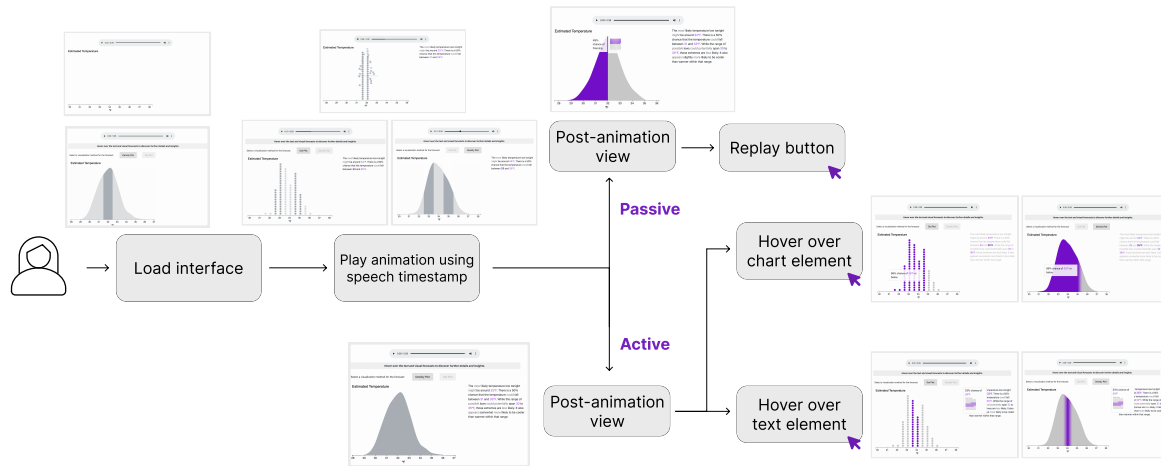


Figure 1: Diagram showing multimodal components and possible user interactions with passive and active interfaces.

implementation of multimodal prototypes for both passive and active consumption of uncertainty communication.

2.2. Multimodal Techniques for Communicating Information

Existing research has explored how multimodal interaction can blend various modes of communication, such as visual, auditory, and haptic cues, to effectively convey information. Bromley and Setlur [BS23] explored the use of language with visualizations by employing hedges to describe trends (e.g., “gradual fall”).

Multimodal techniques can be useful for conveying uncertainty to users who may be blind or have low vision. Sharif et al. [SZW23] provide empirical findings from semi-structured interviews with 16 screenreader users to identify user preferences in uncertainty visualizations. Several studies explore how non-speech audio can be used to represent data, either on its own or in combination with visualizations [LWS96, LBH*97, Bea11]. Previous work has also explored differences between passive and active communication of data insights. Both variants result in similar levels of comprehension [RSK*21]; interactivity can increase user enjoyment [DAV08], while static visualizations offer a more direct message [MG18].

3. Prototype Implementation

Our work builds upon this research to explore the research question, “**What are user preferences for multimodal presentations of uncertain data?**” We implemented two prototypes with passive and active interface behaviors. Design choices were made in consultation with a UX designer at a visual analytics company with 10+ years of experience and one of the co-authors who is an interactive visualization designer with 15+ years of experience. Prototype design was iterative, with feedback provided on four occasions.

3.1. Decision-making Context

The interfaces studied here were situated in a decision-making context. Users assumed the role of a road maintenance company contracted to treat roads with salt to prevent icing. Users were tasked

with applying salt to the roads when the temperature was at or below 32°F (0°C) to prevent ice from forming. They were informed that salt supplies were limited, so maintenance companies must balance cost (and supply) with damage prevention. To make this decision, they viewed and listened to a forecast depicting predicted temperature lows for a given evening.

The appropriateness of a user’s decision to salt the roads depends on the costs of salting and potential damages; information not provided in this scenario. The decision context was used to provide a *situation* of use for the interface rather than to assess rationality. Prior work examined each mode of information in decision-making, finding that speech led to higher trust in information, while text was associated with lower decision confidence [SSCS24]. This related paper studied uncertainty data in the same decision context as our work, but differed in research questions and methodology.

3.2. System Components

The interface flow for both prototypes is shown in Figure 1. Each prototype employs visualizations, text descriptions with hedges, and speech elements with pauses and rate changes to convey data uncertainty. The speech forecast is displayed at the top of the page. A D3 visualization [BOH11] and a text description of uncertainty are shown below the MP3 player. An event listener in the interface monitors time progression in the speech module, triggering animated updates in the visualization and text modules. In addition, the active prototype leverages interaction with the visual and textual modes, including additional linking between the two.

3.2.1. Data Module

Raw distribution data is integrated into a Node.js [Ope23] application using an Express framework, serving the dataset as a static file. Upon loading the prototype, a random trial is selected from the dataset. This selection process links the columns of the .csv file to the unique trial number. Dataset values in the .csv are derived by selecting 100 points from a normal distribution, such that the resulting distributions may not have been normally distributed. This

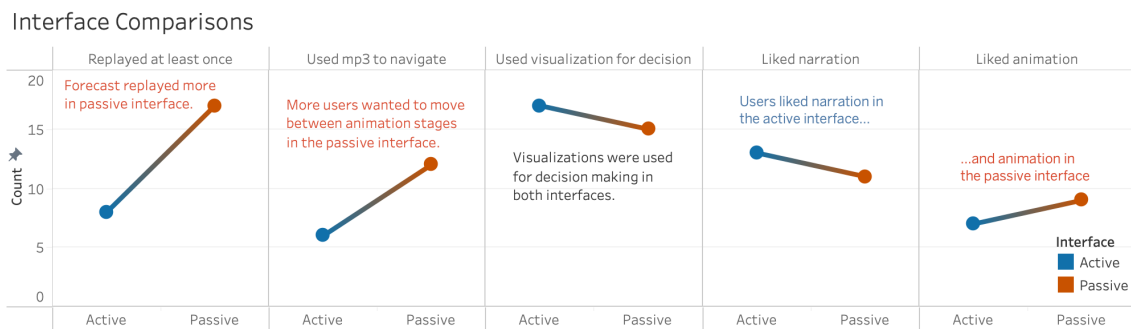


Figure 2: Insights from comparing active and passive interfaces. Different features may be better suited for specific interfaces, while some user interactions are consistent.

variation provided more ecologically valid distributions, as natural data tends to not always follow a normal distribution.

Text templates and timings also serve as a static JavaScript file, including each sentence and the timestamp it occurs in the speech element. The text-to-speech synthesis engine computes the timing information while generating speech from the SSML syntax. These timings are available as static files.

3.2.2. Multimodal Modules

Visualization module. To communicate data uncertainty in visualizations, we employ density and 100-quantile dot plots [KKHM16]. Density plots are created using a kernel density estimator to determine the continuous probability curve from discrete data points. The quantile dot plot is created using a histogram, with 20 evenly spaced bins based on the range of the x-axis scale.

Text module. The text descriptions are generated with natural language templates containing hedges and summary statistics from the visualized dataset [SSCS24]. These statistics include the mean of the distribution, the range of the middle 50% of data, the full range of the data, and a verbal representation of distribution skew. The skew value is computed using the skewness function in R [KN22], then mapped to magnitudes (“slightly” to “significantly”) and a positive or negative direction. Hedges were included in each sentence to communicate uncertainty (e.g., ‘might,’ ‘could’). A standard black color is used to render the text, with a gray color (#757575) applied to hedges to indicate a higher level of uncertainty. Colors employed in the prototypes are compliant with the WCAG AA guidelines for color contrast [W319].

Speech module. The text templates are translated into Google Speech Synthesis Markup Language (SSML) [Goo24] to provide adjustments in pitch, rate of speech, and pauses for communicating uncertainty. For hedges in the text, we modify the speech to 65% of the original rate and lower the pitch by 5%. We apply the same pitch treatment to numerical values, slowing the speech to 70% and adding a 0.2 second break before the values.

3.2.3. Interaction Module

In both prototypes, the forecast can be replayed to view and listen to the information. During the replay, the animated sequence for the

text and visualization also repeats. The active prototype includes detail-on-demand tooltips and linking between modes, shown in Figure 1. There are three distinct hover interactions (hedges, numerical values, visualization) that provide dynamic feedback and detailed insights about the data. A video demonstrating these behaviors is included in the supplementary material.

When a user hovers over a hedge in the text description, the word visually responds by “wobbling” at 3° angles and applying a 0.5px blur effect. Hovering over a numerical value in the text shows a tooltip featuring an icon array visualizing the likelihood of the hovered number occurring within the dataset. Simultaneously, the corresponding section of the visualization that represents this number is highlighted and wobbles to draw attention and convey uncertainty. Upon hovering over a mark in the visualization, a tooltip appears with a text description of the cumulative likelihood of achieving the corresponding value or lower.

4. Preliminary User Study

Using the passive and active prototypes as a design probe, we conducted a preliminary evaluation to qualitatively assess the overarching idea of combining speech, text, and visualization modes for communicating data uncertainty information.

4.1. Method

We employed a within-subjects design to elicit subjective user feedback for both interfaces. Participants engaged with both prototypes. Half of the participants viewed the active prototype first; half viewed the passive. Sessions lasted approximately 45 minutes and were recorded with the participant’s consent. One author coded the transcripts in close consultation with a second author.

For each interface, participants viewed the speech-linked animation, accompanied by the narration and linked text transcript. They could replay the animation and narration or interact with available features. They were encouraged to think aloud to share their observations and reactions to the interface. The facilitator then asked for specifics about the potential use of the information and possible changes to the interface. After participants had viewed both interfaces, the facilitator asked about their usage of the different modes and their perception of each interface’s strengths and weaknesses.

4.2. Participants

The 20 participants (8 female, 12 male) were fluent in English. Half were graduate students from a nearby university, and half were professionals employed at a technology enterprise with a variety of backgrounds (e.g., technical engineering, product management). Participants were recruited via internal mailing lists and Slack channels. Graduate students were compensated with a \$20 Amazon gift card. Professionals were not compensated due to company policy prohibiting payment. We refer to participants using [P#].

4.3. Study Findings

Code counts from the interview transcripts can be found in Figure 2, along with supplementary and recruitment materials. Overall, participants used the speech and text modes as supplementary to the visualization for both interfaces. Interacting with the active interface was helpful for decision-making and providing further detail on event likelihoods. We observed a slight difference in interaction behaviors: passive-first participants replayed the speech element more (1.5 times on average) than active-first (0.6 times).

Replay and navigation. Replay frequency and interaction with the speech element were much higher for the passive interface (17 participants) than the active interface (8 participants). When viewing the passive interface, 12 participants used the speech element to navigate between animation stages, compared to 6 for the active interface. Some of the participants felt that there were too many elements in the interface (13 active, 11 passive).

Trends in interaction. In the active interface, participants could choose the visualization displayed (dot plot or density plot), switching an average of 11.5 times. 14 participants preferred the density plot, citing familiarity with the chart type. Each participant interacted with the density plot an average of 6.9 times, taking an average of 14.5 seconds (s) per interaction. Similarly, they interacted with the dot plot an average of 6.6 times with an average of 10.2s per interaction. The text component had fewer interactions, an average of 5.2 interactions per participant, but each interaction was longer, averaging 30.0s.

Decision-making. Participants frequently reported that they used visualizations in the decision-making process (17 active, 15 passive). The interactions with the visualizations were particularly useful, with 13 participants reporting that they relied primarily on information from the visualization tooltips. According to P5, “Some of the interactions, if not all of them, I feel, were more valuable... I got to the decision much faster.”

The text tooltips were less useful overall, with 9 participants finding them redundant and not decision-relevant. Rather than offering the likelihood of going *below* a given temperature, as in the visualization tooltips, the text tooltips offered the likelihood of being *at* a given temperature. 9 participants found this relatively unhelpful, as this information did not greatly inform or change their decisions. However, 9 participants still found the interaction beneficial for overall understanding - “the information from decision-making perspective doesn’t seem super helpful... but I absolutely love seeing the vis.” [P10].

5. Discussion and Future Work

The exploration of these two interfaces provided useful insights for future multimodal system design. The active interface provided users with a rich set of interactions, and participants often switched between visualizations, suggesting that providing more than one way to represent the data visually was useful. However, for both interfaces, the narrative and animation straddled the line between distracting and engaging. These varied responses underscore the challenges in designing clear multimodal data representations.

Benefits of each mode. Participants most often used visualizations in the decision-making process. Many of the participants also liked the narration as a source of information - “it’s very useful if you’re too lazy to read the text... Audio complemented with text is very powerful” [P14]. Speech also improved comprehension of the data display; “the audio was helpful to understand the overall instructions of how to read the view” [P20]. These findings are consistent with prior results in decision-making with unimodal data representations for visualization and speech: **decision quality was highest with visualization; trust was highest with speech** [SSCS24].

The staged chart composition illustrated how to read the chart, which was useful for some users: “The animation is quite cool, and I understand how to read the graph if I don’t know how” [P14]. However, several participants admitted that the multimodality was overwhelming. P1 commented, “I couldn’t keep track of all of that information... there was just too much going on.” Future work should investigate how to balance multiple modes but not overwhelm users with too many interface components. **Different speakers, text descriptions, and visualizations could also affect user experiences and preferences.**

Context of use. Participants noted crucial differences between the potential usage context of the interfaces. The passive interface was viewed as better suited for a more casual setting - “I think if it was not a serious decision... I would actually prefer the [passive]... I could just see it and then just draw insight from it quickly” [P16], or may also be better suited for users with less experience in data; P6 suggested, “That would be something that would probably appeal to a lot of folks who don’t have a desire to go in and be analysts themselves.” The active interface seemed a better fit for an audience with some expertise in similar types of data interpretation, particularly those who are “proficient in the tool already” [P12].

Future work should explore multimodality in different settings and with users of different levels of expertise. **While casual users might appreciate a guided, narrative experience with high-level insights (the passive interface), expert users may seek detailed, interactive tools for data exploration (the active interface).**

6. Conclusion

This work discusses two interfaces that explore the paradigms of expressing data uncertainty in both active and passive consumption contexts. Preliminary user feedback illuminated use cases and challenges when presenting and interpreting multimodal data uncertainty across these contexts. This research identifies the need for refined, context-specific multimodal strategies in uncertainty communication, different situational contexts (e.g., for analysis vs. during travel), and user groups (e.g., data experts vs. casual users).

References

- [Bea11] BEARMAN N.: Using Sound to Represent Uncertainty in Future Climate Predictions for the UK. In *17th International Conference on Auditory Display* (2011), International Community for Auditory Display. 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. 1
- [BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2301–2309. 2
- [BS23] BROMLEY D., SETLUR V.: What Is the Difference Between a Mountain and a Molehill? Quantifying Semantic Labeling of Visual Features in Line Charts. In *IEEE Transactions on Visualization and Computer Graphics* (2023). 2
- [DAV08] DYER S., ADAMO-VILLANI N.: Animated Versus Static User Interfaces: A Study of Mathsigner™. *International Journal of Human and Social Sciences* 3, 6 (2008). 2
- [EFN12] ENDERT A., FIAUX P., NORTH C.: Semantic Interaction for Visual Text Analytics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), CHI '12, Association for Computing Machinery, p. 473–482. 1
- [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. 1
- [Goo24] GOOGLE: Speech Synthesis Markup Language (SSML) Reference, 2024. Google Cloud Documentation. URL: <https://cloud.google.com/text-to-speech/docs/ssml>. 3
- [Hul19] HULLMAN J.: Why authors don't visualize uncertainty. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 130–139. 1
- [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. 1
- [Jia15] JIANG X., PELL M. D.: On How the Brain Decodes Vocal Cues About Speaker Confidence. *Cortex* 66 (2015), 9–34. 1
- [Kay23] KAY M.: *ggdist: Visualizations of Distributions and Uncertainty*. Northwestern University, 2023. R package version 3.3.0. URL: <https://mjskay.github.io/ggdist/>. 1
- [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. 1, 3
- [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. 1
- [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>. 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. 1
- [LBH*97] LODHA S. K., BEAHAN J., HEPPE T., JOSEPH A., ZANE-ULMAN B.: MUSE: A Musical Data Sonification Toolkit. In *Proceedings of the 4th International Conference on Auditory Display, Palo Alto, California, November 2–5, 1997* (1997), Georgia Institute of Technology. 2
- [LWS96] LODHA S., WILSON C., SHEEHAN R.: LISTEN: Sounding Uncertainty Visualization. In *Proceedings of Seventh Annual IEEE Visualization '96* (1996), IEEE, pp. 189–195. 2
- [MG18] MAHAJAN K. N., GOKHALE L. A.: Comparative Study of Static and Interactive Visualization Approaches. *International Journal on Computer Science and Engineering (IJCSE), e-ISSN* (2018), 0975–3397. 2
- [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. 1
- [Ope23] OPENJS FOUNDATION: Node.js, 2023. <https://nodejs.org>. 2
- [PKH21] PADILLA L., KAY M., HULLMAN J.: Uncertainty Visualization. In *Wiley StatsRef: Statistics Reference Online*. Wiley, 02 2021, pp. 1–18. 1
- [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. 1
- [RSK*21] RÖBBELEN A., SCHMIEDING M. L., KOPKA M., BALZER F., FEUFEL M. A.: Interactive Versus Static Decision Support Tools for COVID-19: Randomized Controlled Trial. *JMIR Public Health and Surveillance* 8 (2021). URL: <https://api.semanticscholar.org/CorpusID:248240389>. 2
- [SC93] SMITH V. L., CLARK H. H.: On the Course of Answering Questions. *Journal of Memory and Language* 32 (1993), 25–38. 1
- [Sch97] SCHRIEVER K. A.: *Dynamics in Document Design: Creating Text for Readers*. John Wiley & Sons, Inc., USA, 1997. 1
- [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. 1
- [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
- [SSCS24] STOKES C., SANKER C., COGLEY B., SETLUR V.: From Delays to Densities: Exploring Data Uncertainty through Speech, Text, and Visualization. In *Eurographics Conference on Visualization (EuroVis)* (2024), The Eurographics Association. to appear. 2, 3, 4
- [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. 1
- [SW95] SPERBER D., WILSON D.: *Relevance: Communication and Cognition*. Blackwell, Oxford, 1986/1995. 1
- [SZW23] SHARIF A., ZHONG R., WANG Y.: Conveying Uncertainty in Data Visualizations to Screen-Reader Users Through Non-Visual Means. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23)* (New York, NY, USA, 2023), Association for Computing Machinery. 2
- [(W319) (W3C) W. W. W. C.: WAI Web Accessibility Tutorials: Complex Images. World Wide Web Consortium website, 2019. <https://www.w3.org/WAI/tutorials/images/complex/>. 3