Towards Visualization-Supported Uncertainty Elicitation

Nikolaus Piccolotto* Fatih Öztank† Silvia Miksch‡ Markus Bögl§

TU Wien

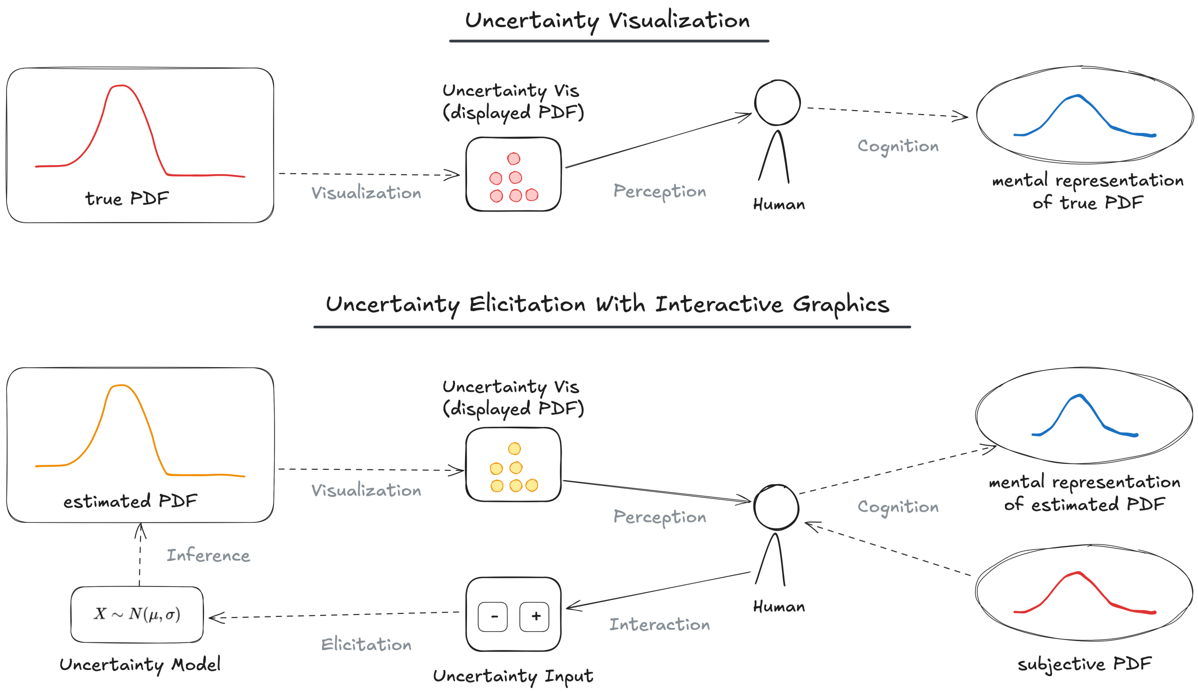

Figure 1: *Uncertainty Visualization* vs. *Uncertainty Elicitation* using interactive graphics. The former (top row) features an unidirectional information flow where the true probability density function (PDF) is known (left) and finds their way into the expert's mind via visualization (right). In the latter (bottom row), the expert communicates their subjective PDF to the computer via elicitation, thus forming a cyclic interaction.

Abstract

Expert knowledge in visual analytics (VA) informs design processes, is required to gain insights from data, and may steer computational inference of patterns and trends in the data. Many fields and industries, such as food safety or civil engineering, rely on experts' subjective probabilities of future events and uncertain quantities to make rational and informed decisions given particular risks. For example, subjective probabilities may be put into Bayesian models as prior distributions of variables. In VA, too, experts are asked to provide (elicit) subjective probabilities, but our field has not yet developed good practices on how visual-interactive interfaces to elicit subjective uncertainties should be designed and evaluated. In an attempt to divide and conquer, this paper provides relevant research directions and opportunities after reviewing the literature on uncertainty elicitation (UE) and uncertainty visualization.

Index Terms: Input visualization, uncertainty, subjective probability, knowledge-assisted visual analytics.

*e-mail: nikolaus.piccolotto@tuwien.ac.at
†e-mail: fatih.oeztank@tuwien.ac.at
‡e-mail: silvia.miksch@tuwien.ac.at
§e-mail: markus.boegl@tuwien.ac.at

1 Introduction

Expert knowledge is indispensable to visual analytics (VA), which "combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets" [45]. Domain-specific knowledge is, e.g., required for modeling the problem: Only a human can identify, relate to, and understand subjects interacting in the real world. Therefore, only a human can decide how this reality should be simplified, while remaining a useful generalization, so that a computer can appropriately infer relationships and patterns from the available data [2]. On the other hand, expert knowledge can steer or support the inference process itself [28]. E.g., in a classification problem, an expert may provide the categories themselves (e.g., in active machine learning) or the rules to classify them. Domain knowledge is, therefore, expected to improve the VA process and the insights it may yield.

In turn, uncertainty is an integral factor of information itself. It can be, among other things, inconsistent, partial, subjective, or imprecise [31, 59, 70]. An important aspect of expert knowledge, therefore, is their opinion on its uncertainty. For example, uncertainty may present as missingness, a data quality problem that needs fixing [30] before continuing the analysis. In that case, expert opinions may take the form of estimations of missing values [8]. Subject-matter experts may be able to articulate reasonable value ranges (expressed as credible intervals) or whether some values are more likely to have been measured than others (expressed as a probability distribution function (PDF)). The kind of uncertainty is then mostly aleatoric (random noise) [67]. In other contexts, such as architecture, uncertainty might be conceptualized as vagueness of temporal/spatial primitives (such as the undetermined position of a wall [5]) or of categorical data (such as materials or colors to be used). Vagueness in designs [1] then accounts for epistemic uncertainty [67], i.e., that floor plan, materials, and colors will be known eventually, but are not yet fixed. These two examples highlight how it can be impossible not to deal with uncertainty, although in very different ways.

If expert knowledge is essential to VA but that knowledge is associated with uncertainty, then we require well-thought-out ways for human experts to externalize their knowledge *along with their uncertainty about it*. The authors of this paper are by no means the first to point that out [33, 54, 57]. *Input visualization* [11] has been recently proposed as "visual representations that are designed to collect (and represent) new data rather than encode preexisting datasets". While thematically fitting, as the expert knowledge is generally not part of the dataset can thus be considered new data, uncertainty is only of minor importance in the referenced approaches. Even before that, the visualization field proposed interfaces to externalize uncertainty (cf. Sec. 3), but these were not the primary focus of the respective paper. To the best of our knowledge, the problem of externalizing uncertainty is not yet systematically researched within our field. If we continue to avoid doing so, it hinders, e.g., our ability to develop tested solutions, encapsulate them into reusable software artifacts, and effectively use expert knowledge in our VA approaches.

Located in the intersection of psychology and statistics, *Uncertainty Elicitation* (UE, also known as *expert elicitation*) is a long-studied field of research with a rich body of literature going back at least to the 1970s. UE studies how judgments of uncertain quantities may be drawn out from human experts as reliably and objectively as possible. To that end, it investigates, e.g., how alternative question or response formats influence expert answers. UE also devises analysis methods and protocols. UE methods find broad application in important domains such as food safety [26] and ecological modeling [49]. Therefore, visualization researchers need to consider insights from UE. While domain knowledge elicitation [46] captures a broader view of expert knowledge that may aid visualization researchers in a design study [66], we focus on eliciting just uncertainty here.

Contributions. In this paper, we aim to give a very brief overview of UE (Sec. 2) along with references that may prove helpful to visualization researchers. We contrast that information with our field's perspective on uncertainty and present some efforts towards uncertainty externalization (Sec. 3). The remainder of the document will interchangeably refer to uncertainty externalization and elicitation and shorten both to UE. The practice of UE with the support of interactive visualization will be called UE inputs. Finally, we develop research questions for our field regarding responsible research of UE inputs (Sec. 4). In doing so, we aim to break it into smaller, more manageable problems and contribute to an input visualization research agenda. The effect of our efforts, we hope, is to initiate a discussion of how expert uncertainty can be externalized with interactive visualizations.

2 Uncertainty Elicitation

This section aims to give a non-exhaustive overview of uncertainty elicitation (UE). We refer the reader to textbooks [21,57] and recent publications [56] for more detailed discussions.

The goal of UE is to obtain the expert's knowledge/uncertainty of an unknown quantity as best as possible [57, Chp. 1]. Their knowledge/uncertainty will be subjective, so they are referred to as *subjective probability*, and it will be expressed imprecisely. Neither is considered a problem in practice as long as the elicitation procedure is well-reasoned, thoroughly planned, and carefully conducted [13]. The person responsible for that is the *facilitator*. The elicitation output generally comprises a credible interval or a complete probability density function (PDF) of the unknown quantity's value.

Why UE. We allured to reasoning why this is useful for VA in Sec. 1. Generally, expert opinions become an attractive alternative when one is constrained by money, time, or circumstances to determine the unknown quantity empirically. Such is the case in many complex engineering projects or strategic industry decisions. For example, putting tunnels underneath Stuttgart's train station is a unique challenge, as no one has ever put tunnels underneath that particular station. Nor will one do so a second time. Similarly, a wood company deliberating whether to plant a new tree species in a given area cannot exactly test it out and postpone the decision for three to four decades until the test trees mature. However, similar tunnels/trees would have been constructed/planted elsewhere, so some experience will apply to the new problem. The expectation is that an expert will be able to judge its extent. Hence, real case studies can be found across domains such as food safety, medicine, ecology, agriculture, or sales [13,57].

Psychology Background. Experts do not have probability distributions readily available on some mental shelf, but construct them after being asked to assess them. For this reason, memory, psychological biases [23] and mental heuristics, such as the availability bias or base rate fallacy, may influence the expert's answer. Some can be counteracted by particular question phrasing, but warning or informing experts about a bias is often ineffective in that regard [3]. More potentially confounding factors include the question format, e.g., whether a probability is stated as $1/100$ (frequency), $1\,000/100\,000$ (frequency with unnecessarily big numbers), $1\,\%$ (percent), $1:99$ (odds), or $432/4319$ (natural frequency). In addition, how the expert can respond, e.g., by entering a number or selecting from a provided scale, can impact the outcome. As a simple example, a response of 50 % probability for an event may be intended as "I don't know" if no such option exists. As such, much of what visualization researchers know from designing their own surveys [18, 74] is applicable in UE as well. See [56] for a brief introduction and [57] for deeper discussions.

Eliciting a Distribution. Often, one is not just interested in a single number (e.g., how much will buildings around the train station sink because of the tunnels), but also in the (un)certainty of the expert's estimate, i.e., we look for a PDF fitting to the expert's subjective probabilities. In practice, the facilitator assumes, per the expert, a suitable family of PDFs (such as Normal or Binomial) for the random variable in question, as many probability distributions could fit a given set of elicited subjective probabilities. The goal is then to estimate its parameters. To do so, Kadane and Wolfson [39] distinguish two approaches. *Structural* approaches ask the expert directly about the parameters, e.g., the mean of the Normal distribution. O'Hagan et al. [57, Sec. 5.2.3] suggest that this approach is not very promising as people tend not to be great at judging means or variances of a sample. On the other hand, *predictive* approaches ask the expert to predict either the probability p of a given value x of a random variable X (x in $P(X < x) = p$) or the value x for a given probability p (p in the same equation). The latter is sometimes preferred as it is easier to choose sensible values for p than x. E.g., by choosing $p \in [.5, .25, .75]$, the experts elicit the median and lower/upper quantiles. O'Hagan et al. recommend overfitting, i.e., eliciting more data points than necessary to fit a distribution. Doing so allows for identifying inappropriate distributions and extreme assessments. Multivariate distributions where random variables depend on each other are too complex a topic for this paper, and we

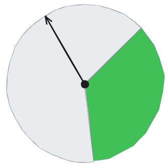

Figure 2: Probability Wheel design recreated from [61].

refer the reader to the aforementioned textbooks.

Elicitation Formats (UE Inputs). Assuming that a facilitator determined what and how to ask, the elicitation format determines how an expert may respond to inquiries. O'Hagan et al. [57, Sec. 4.4.2] and Renooij [61] distinguish two broad categories. *Direct* methods allow to provide the sought value (such as p, x in $P(X < x) = p$) directly, e.g., in a form-based survey. On the other hand, *indirect* methods often take the form of *bets* as they force the expert to choose between their assessment and a fair game of random chance to win an imaginary or real benefit. A notable example, as [57] lists it in both categories, is the probability wheel (Fig. 2). The elicitation format metaphorically represents a "wheel of fortune". It has two sectors, so it looks like a pie chart. The expert can adjust the sectors' sizes to define the probability. They can either submit the resulting probability directly (*direct method*) or adjust the probability of the alternative game such that they would be indifferent in choosing their assessment or the alternative (*indirect*). As far as we can tell, more common graphical designs involve popular uncertainty visualizations [19, 34, 38], like density curves or dot plots.

Measures of Quality. As the subjective PDF cannot be observed directly, we cannot ascertain the quality of a given elicitation by computing the deviation from measurement to ground truth. However, the literature discusses concepts researchers may use to evaluate UE methods or particular elicitations. The *reliability* of a method was defined as the error in repeated measurements [77]. Presumably because UE often involves "one-off" assessments, such as the particular tunnel to drill, it seems not much discussed in more recent works on UE. Researchers may also not agree with what reliability implies, i.e., that the elicitation process distorts an otherwise (in the expert's mind) perfectly defined PDF. It is generally not expected that subjective probabilities are *coherent*, i.e., that they collectively follow the laws of probability. E.g., if an expert elicited $P(A) = p_A$ and $P(B) = p_B$, their assessment of the probability $P(A,B)$ is not necessarily equal or close to $p_A * p_B$ (assuming independent variables). However, coherence is a requirement if we are to reason rationally. Thus, the facilitator is expected to double-check such cases during the inquiry and point out discrepancies to the expert so that they may revise earlier assessments. Ideally, the elicited uncertainties account properly for tails and center of a distribution, i.e., events with very low frequency are judged as very unlikely. When this is not the case, the participants exhibit *over- or underconfidence* in their assessments. As a special case, they may not *discriminate* between likely and unlikely events and judge everything having roughly the same probability. UE literature refers to this topic as *calibration* [57, Sec. 4.2]. Researchers may measure calibration when the true frequency of events is known. However, a well or badly calibrated result may be related to the experts as much as to the UE method. Among others, bad calibration may be caused by social relationships (e.g., overconfident judgments to demonstrate expertise), honest lack of knowledge, or anchoring biases. Good calibration may happen due to appropriate question formats, the problem context (experts in some domains seem generally better calibrated), or effective training. Either way, calibration does not account for the order of judgments, which may be an issue for some applications [53, Sec. 7.2]. If a ground truth is available, the expert's judgment can be assessed with *scoring rules*. Such rules assign a score (hence the name) to the elicited uncertainty based on their proximity to or association with the ground truth. As such, they are similarity measures of probabilities. The Kullback-Leibler Divergence would be an option if complete PDFs are available; O'Hagan et al. [57, Sec. 8.2] list many alternatives for discrete and continuous probabilities.

Multiple Experts. In order not to rely on single opinions and to average over individual errors in judgment, it is generally advised to ask multiple experts, if possible. Importantly, they are to be asked individually first to avoid unwanted group dynamics that may lead to an opinion not being adequately considered (e.g., [32]). Multiple opinions must be reconciled into one, for which two strategies exist: *Pooling* aggregates them mathematically, while *behavioral aggregation* asks experts to discuss and find a consensus of anonymized results. The catch with the former neutral-appearing option is that several pooling functions exist, and choosing one is a subjective choice by the facilitator, as they inevitably incur a weighting of opinions. A possible issue with the latter can be, e.g., that group dynamics prevent a real consensus from being found or that such a thing was unattainable to begin with, and the outcome is instead a compromise.

Summarizing insights and recommendations from a whole research field in a few paragraphs inevitably leads to oversimplifying some and excluding other topics. However, we hope to have provided a reasonable starting point with useful references for visualization researchers. For practitioners, it is recommended to follow broadly accepted protocols [56], such as IDEA, SHELF, or guidelines by the European Food & Safety Authority.

3 Uncertainty in Visualization

Similar to the previous section, we give an overview of uncertainty concepts and research in visualization. The field recognizes that uncertainty may stem from various sources [9] and comes in many forms at different stages of the analysis process: E.g., missing and erroneous data [45], trustworthiness of the data source [52], imprecise measurements [70], multiple interpretations [59], risk assessment and decision-making [25, 47], or ambiguities in visual representations ("uncertainty of visualization") [12]).

Given the importance of the topic for visualization, researchers added it to accepted theoretical models. Correa et al. [15] proposed an addition to the VA model [45] where uncertainty propagates from the data source to the final image. Sacha et al. [64] subsequently considered the relation of uncertainty in their knowledge-generation model [65]. Recently, Andrienko et al. [2] suggested that VA equates to interactive model building and therefore also model checking [34, 41], which is a process that has to account for decision uncertainty (i.e., is this or the other model better?).

Maybe the bulk of visualization's literature on the topic considers effective and faithful representations of quantifiable uncertainty, i.e., visual representations that induce accurate probability judgments in viewers ("visualization of uncertainty"). Several surveys exist already [35, 37, 42, 48]. Approaches to visualize uncertainty can be categorized into different groups. Some show the distribution in a discrete/frequency or continuous/probability idiom, e.g., [6, 44, 63], usually in separate views. Additional marks may be added to an existing visualization, such as error bars [16], variance bands [69], or boundaries [5]. The visualization itself can be visually scaled to make more uncertain values less distinguishable from others [17]. Representative random samples from the distribution can be shown simultaneously or in sequence (hypothetical outcome plots [36]). Finally, it was recently investigated to correct the visualized uncertainty based on subjective probabilities [78] or ways to handle qualitative uncertainty [51, 52].

Regarding "uncertainty of visualization", research investigated cognitive heuristics and perceptual biases. Dimara et al. [23] recently gave a task-based overview of cognitive biases in visualization, noting that they found few studies on that topic in visualization. Quadri et al. [60] provided a survey of perception-based studies in visualization. Examples include for instance anchoring effects, where the judgment of cluster separability depended on what images were seen previously [72]. Heuristics include those used in ensemble coding [68], where so-called visual proxies determine judgments about the visualized distribution (e.g., the hull of a point cloud may be used to judge variance). Dimara et al. [22] found that an attraction effect in visualizations could be reduced with interaction, suggesting that similar strategies might work for other cognitive biases.

In some instances, data uncertainty is seen as a quality problem that needs to be fixed, e.g., because other required computational machinery cannot deal with it. Interactive visualizations to, e.g., impute, de-trend, or generally preprocess data have been proposed for tabular data [43] and time series [4, 8, 30].

Existing UE Inputs. The visualization community already proposes UE inputs, and we present here a non-exhaustive set of examples. In their experiment, Newburger et al. [55] used sliders for mean and variance of a normal distribution, where participants fitted bell curves to an existing visualization. Dhanoa et al. [20] proposed a fuzzy spreadsheet that allows complex calculations with random variables. Users must define distribution parameters directly, e.g., specifying cells containing mean and variance (Normal) or selecting the binomial p parameter from a provided discrete scale (Bernoulli). In the experiment by Hullman et al. [34], participants used several interfaces with both discrete and continuous visual representations to construct a distribution curve graphically. Borrelli et al. [10] proposed a parallel coordinates plot (PCP) with probabilistic brushing. They implemented a spline editor so that users may define their subjective probability for an attribute range to be representative for a category. E.g., in the context of cars one could consider the attribute *motor power* and the category *sports car*. An option would be to assign no probability up to 100 kW and an exponentially rising probability after that. The final selection certainty is then computed from the individual attribute probabilities. As for a spatiotemporal example, Ribičić et al. [62] allowed users to sketch, e.g., uncertain breach positions into a flood simulation model. Users would select a point along a spline representing barriers to do so. This position is the center of a Normal distribution, while additional handles control the standard deviation. Then, k samples from this distribution are drawn and fed into the simulation model so users can assess potential breach impacts. The time when the breach happens is fixed and selected by the user. Other works come close to *predictive* elicitation methods, though they do not explicitly account for uncertain outcomes. E.g., in Podium [76], users move options in a list up and down to obtain attribute weights for a scoring function consistent with their ranking. Similarly, Explainers [29] allowed users to construct a distance function interactively by moving data points on a 2D plane. Both these approaches fall into the category of *semantic interaction* [27], where the model (e.g., Podium's weighting function) is inferred by interactions with the visual representation.

To summarize, uncertainty in visualization research focuses a lot on the top row of Fig. 1, i.e., how we can get people to perceive accurately and thus act rationally upon visualized probabilities. Some works propose UE inputs, but these proposals seem to favor structural and direct UE approaches, which are generally seen as inferior to their predictive counterparts. Of course, there are design tradeoffs at play here, i.e., it may be unwarranted to submit the expert to multiple lengthy elicitations when they use the fuzzy spreadsheet [20] to estimate the cost of an upcoming holiday. We will get to those in the next section.

4 Discussion & Research Directions

After having discussed the practice of uncertainty elicitation (Sec. 2) and summarized the uncertainty visualization research field (Sec. 3), we will derive research directions and opportunities from their overlaps and gaps.

Explicit Uncertainty Modeling. An UE procedure is always conducted in the context of a previously developed Bayesian model. In the simplest case, it can be one normally distributed random variable, but much more specialized models, including multiple variables, may be necessary depending on the occasion [19, 40, 71]. As "all models are wrong," the statistical modeling of the uncertain quantities presents an important dimension of the design process for uncertainty inputs. As the model itself may be implicit, e.g., when the experts classify data points [75], it might first be necessary to develop its underlying mechanics with user research and participatory design methods. Model selection [7] is an iterative and subjective process, informed by computed quality metrics. In the context of UE inputs, the choice of model impacts the other VA components. More complex models will require longer and more sophisticated elicitation methods, while simulations based on those models (e.g., Markov Chain Monte-Carlo) will also take more time, thus impacting usability and interactivity. Eventually, experts must accept a model as useful, and their familiarity with Bayesian statistics might also influence this design choice. Further, complications arise should a model not be set in stone but be malleable. E.g., the expert may already have partially elicited some facts about their subjective PDF before they realize a necessary change to the model. Changing models is common in the Bayesian research cycle [73], and interactive modeling has recently received research attention [41]. Further advances in these directions will also benefit the design methodology around UE inputs.

Uncertainty Input Design Space. The probability wheel from Sec. 2 is just one example of an UE input. Many more variants can be thought of, e.g., involving changed metaphors (static handle and rotating wheel), alternative encodings (1D instead of 2D proportions), or different visual channels (e.g., color). It is unlikely that they will perform equally well. Similarly, many options exist for how the uncertainty display can be changed, e.g., all forms of direct and indirect manipulation. In addition, we get the necessary choices from UE, i.e., when and how to give experts feedback about their elicitation, and whether the elicitation happens directly or indirectly. Finally, the structure of the uncertainty model and the problem context, i.e., variables, their meaning, and assumed distributions, will impact possible UE input designs. In order to understand the properties of effective UE inputs, this design space needs to be logically structured so that, subsequently, we may empirically test the effect of individual design choices. Defining its dimensions will also assist practitioners and researchers in the design process.

Variety of Data Types. Special attention must be drawn to the fact that visualization designers deal with various data types: Tables, networks, hierarchies, time-oriented, geospatial, and others, in various combinations. How such data will be modeled is a task for the visualization designers. One can expect that UE inputs for attribute uncertainty [70] in a time series model will work differently than UE inputs for a classification problem or those for modeling uncertainty of geospatial locations. Tentative evidence suggests that using maps to elicit subjective probabilities with geographic datasets leads to differences in priors and is preferred by experts [19, 58]. Eliciting other kinds of uncertainty, like temporal intervals or spatial locations, seems not explored yet. Therefore, researching UE inputs in various application scenarios will inform our uncertainty modeling and UE design options.

Variety of Uncertainty Concepts. A subset of uncertainty concepts in visualization (cf. Sec. 3) is well expressed by Bayesian models and statistics, such as imprecision or error. Other kinds of uncertainty seem less easily mapped, like recency or trustworthiness. While they could be integrated into some weighting function for data points, the question is how it should be done systematically and whether that is useful in the first place.

Composition of Uncertainty Inputs. It stands to reason that eventually we will find appropriate uncertainty inputs for certain kinds of random variables, e.g., normally or beta-distributed. Once that is achieved, the question becomes what effects the model's structure has on uncertainty inputs while keeping the UE input interface usable. How do UE inputs adapt to, e.g., dependent variables or hierarchical models? Do we need entirely different inputs, or can we compose them into more complex interfaces, much like the models themselves? Accomplishing the latter would make building reusable software components, which can then be deployed to practitioners, easier.

Risk of Interactivity. Interactivity [24], i.e., natural and intuitive interactions with visual representations and their timely update to those interactions, is highly regarded in our field. Visualization researchers strive to couple view and model tightly through interactivity. While interactivity generally comes with costs [50], in the context of UE inputs specifically, it also poses risks. Consider the bottom row of Fig. 1, displaying a possible UE process with visualizations. For the sake of argument, a subjective PDF exists in the expert's mind and they wish to express it to the computer through a highly interactive UE input. In a naive interactive scenario, they manipulate the UE input, perceive updates of the displayed estimated PDF, compare that to their subjective PDF, and restart the cycle until subjective, perceived, and displayed PDF are in sync. However, this process carries some risks that stem from distortions along the pictured arrows. As outlined in Sec. 2, humans often understand their subjective PDF poorly. Overconfident or incoherent judgments addled by biases and misconceptions distort the elicitation and comparison arrows. Next, too coarse or inappropriate uncertainty visualization idioms may distort the visualization and perception arrows. Incorporating real observations into the process risks anchoring biases, where experts hesitate to deviate too much from existing trends. The PDF elicited by an expert who manipulates the UE input until an uncertainty visualization looks "good enough" is unlikely to be precise unless we can model and correct all the mentioned distortions, biases, and errors. The question then becomes how to maximize interactivity and tight model/view coupling while retaining reliable elicitation.

Empirical Comparison of Uncertainty Inputs. To identify unreliable elicitation, we need an empirical way to compare approaches. A significant challenge for UE inputs is that there are so many moving parts: There is the problem context, which somewhat goes hand in hand with the uncertainty model, there are the test subjects and their knowledge about the problem, and the UE input options we would like to test. In an ideal world, we could quantitatively compare the subjective with the estimated PDF. In reality, researchers are likely forced to use some observable quantity that may be used as a ground truth. UE literature leverages predictions of quantities that will be known in the future (e.g., on how many days of the following three weeks will it rain?) or so-called almanac questions that are knowable today (e.g., is Tokyo or Algiers more North?). Hullman et al. [34], in an experiment about graphical prediction, used performance on a statistical reasoning task after the true distribution was shown as a discrete or continuous visualization, which might not be a generally feasible strategy for UE inputs. Therefore, we ask by which methods and measures we shall best investigate the effectiveness of UE inputs.

Facilitator Role and Automatic Feedback. There is a discrepancy between the main proposed UE methods, which involve a human conductor (the facilitator), and common usage scenarios in VA, where the expert is generally on their own. Having a facilitator present for a VA session at all times is not feasible. However, the facilitator performs important tasks for UE, such as giving assessment feedback, testing their coherence, or highlighting extreme judgments. At the same time, the cost of wrong uncertainties differs depending on circumstances. The safety of a novel edible item to be marketed in the EU will impact the health of 450 million people. A breach location that is too narrowly considered in a spatiotemporal flood simulation model may lead to emergency personnel preparing for ineffective mitigation measures. On the other hand, it may be of comparably minor consequence if someone misjudged the budget for a birthday present. As such, the risk context of the analysis seems to dictate a design tradeoff between the interaction cost for elicitation [50] and trustworthiness of the outcome. For low-risk situations it could be beneficial to not consider some or all facilitator tasks, or prefer structural UE approaches even if they are known to be inferior to predictive approaches. To that end, we could investigate which facilitator tasks could be transferred to a guidance subsystem [14].

5 Conclusion

In this paper, we introduced research in uncertainty elicitation and contrasted it with uncertainty-related research in the field of visualization, especially considering how experts interact with the uncertainty model via UE inputs. In doing so, we derived several directions for future research in visualization, thus contributing to an input visualization research agenda.

Acknowledgments

This research was funded in whole or in part by the Austrian Science Fund (FWF) [10.55776/PAT6437224]. For open access purposes, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

References

[1] J. Abualdenien and A. Borrmann. Vagueness visualization in building models across different design stages. *Advanced Engineering Informatics*, 45:101107, 2020. doi: 10.1016/j.aei.2020.101107 2

[2] N. Andrienko, T. Lammarsch, G. Andrienko, G. Fuchs, D. Keim, S. Miksch, and A. Rind. Viewing Visual Analytics as Model Building. *Computer Graphics Forum*, 37(6):275–299, 2018. doi: 10.1111/cgf.13324 1, 3

[3] S. Belia, F. Fidler, J. Williams, and G. Cumming. Researchers Misunderstand Confidence Intervals and Standard Error Bars. *Psychological Methods*, 10(4):389–396, 2005. doi: 10.1037/1082-989X.10.4.389 2

[4] J. Bernard, M. Hutter, H. Reinemuth, H. Pfeifer, C. Bors, and J. Kohlhammer. Visual-Interactive Preprocessing of Multivariate Time Series Data. *Computer Graphics Forum*, 38(3):401–412, 2019. doi: 10.1111/cgf.13698 4

[5] G. Berseth, B. Haworth, M. Usman, D. Schaumann, M. Khayatkhoei, M. Kapadia, and P. Faloutsos. Interactive Architectural Design with Diverse Solution Exploration. *IEEE Transactions on Visualization and Computer Graphics*, 27(1):111–124, 2021. doi: 10.1109/TVCG.2019.2938961 2, 3

[6] M. Blumenschein, L. J. Debbeler, N. C. Lages, B. Renner, D. A. Keim, and M. El-Assady. V-plots: Designing Hybrid Charts for the Comparative Analysis of Data Distributions. *Computer Graphics Forum*, 39(3):565–577, 2020. doi: 10.1111/cgf.14002 3

[7] M. Bögl, W. Aigner, P. Filzmoser, T. Lammarsch, S. Miksch, and A. Rind. Visual Analytics for Model Selection in Time Series Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2237–2246, 2013. doi: 10.1109/TVCG.2013.222 4

[8] M. Bögl, P. Filzmoser, T. Gschwandtner, S. Miksch, W. Aigner, A. Rind, and T. Lammarsch. Visually and statistically guided imputation of missing values in univariate seasonal time series. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 189–190, 2015. doi: 10.1109/VAST.2015.7347672 1, 4

[9] G.-P. Bonneau, H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz. Overview and State-of-the-Art of Uncertainty Visualization. In C. D. Hansen, M. Chen, C. R. Johnson, A. E. Kaufman, and H. Hagen, eds., *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*, Mathematics and Visualization, pp. 3–27. Springer, London, 2014. doi: 10.1007/978-1-4471-6497-5_1 3

[10] G. Borrelli, T. Ittermann, and L. Linsen. Mapping Mental Models of Uncertainty to Parallel Coordinates by Probabilistic Brushing. *Computer Graphics Forum*, p. e70103, 2025. doi: 10.1111/cgf.70103 4

[11] N. Bressa, J. Louis, W. Willett, and S. Huron. Input Visualization: Collecting and Modifying Data with Visual Representations. In *CHI 2024*, p. Article No.: 499. ACM, Honolulu HI USA, 2024. doi: 10.1145/3613904.3642808 2

[12] K. Brodlie, R. Allendes Osorio, and A. Lopes. A Review of Uncertainty in Data Visualization. In J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong, eds., *Expanding the Frontiers of Visual Analytics and Visualization*, pp. 81–109. Springer London, London, 2012. doi: 10.1007/978-1-4471-2804-5_6 3

[13] N. C. Brownstein, L. , Thomas A., O. , Anthony, and J. and Pendergast. The Role of Expert Judgment in Statistical Inference and Evidence-Based Decision-Making. *The American Statistician*, 73(sup1):56–68, 2019. doi: 10.1080/00031305.2018.1529623 2

[14] D. Ceneda, T. Gschwandtner, T. May, S. Miksch, H. Schulz, M. Streit, and C. Tominski. Characterizing Guidance in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):111–120, 2017. doi: 10.1109/TVCG.2016.2598468 5

[15] C. D. Correa, Y.-H. Chan, and K.-L. Ma. A framework for uncertainty-aware visual analytics. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pp. 51–58, 2009. doi: 10.1109/VAST.2009.5332611 3

[16] M. Correll and M. Gleicher. Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151, 2014. doi: 10.1109/TVCG.2014.2346298 3

[17] M. Correll, D. Moritz, and J. Heer. Value-Suppressing Uncertainty Palettes. In *CHI 2018*, pp. 1–11. ACM Press, Montreal QC, Canada, 2018. doi: 10.1145/3173574.3174216 3

[18] C. Courage, K. Baxter, and K. Caine. *Understanding Your Users: A Practical Guide to User Research Methods*. Elsevier, Morgan Kaufmann, Amsterdam Boston, 2nd edition ed., 2015. 2

[19] R. Denham and K. Mengersen. Geographically assisted elicitation of expert opinion for regression models. *Bayesian Analysis*, 2(1):99–135, 2007. doi: 10.1214/07-BA205 3, 4

[20] V. Dhanoa, C. Walchshofer, A. Hinterreiter, E. Gröller, and M. Streit. Fuzzy Spreadsheet: Understanding and Exploring Uncertainties in Tabular Calculations. *IEEE Transactions on Visualization and Computer Graphics*, 29(2):1463–1477, 2023. doi: 10.1109/TVCG.2021.3119212 4

[21] L. C. Dias, A. Morton, and J. Quigley, eds. *Elicitation: The Science and Art of Structuring Judgement*, vol. 261 of *International Series in Operations Research & Management Science*. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-319-65052-4 2

[22] E. Dimara, G. Bailly, A. Bezerianos, and S. Franconeri. Mitigating the Attraction Effect with Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):850–860, 2019. doi: 10.1109/TVCG.2018.2865233 4

[23] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. A Task-Based Taxonomy of Cognitive Biases for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 26(2):1413–1432, 2020. doi: 10.1109/TVCG.2018.2872577 2, 3

[24] E. Dimara and C. Perin. What is Interaction for Data Visualization? *IEEE Transactions on Visualization and Computer Graphics*, pp. 119–129, 2019. doi: 10.1109/TVCG.2019.2934283 5

[25] E. Dimara and J. Stasko. A Critical Reflection on Visualization Research: Where Do Decision Making Tasks Hide? *IEEE Transactions on Visualization and Computer Graphics*, pp. 1128–1138, 2021. doi: 10.1109/TVCG.2021.3114813 3

[26] EFSA. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. *EFSA Journal*, 12(6):3734, 2014. doi: 10.2903/j.efsa.2014.3734 2

[27] A. Endert, R. Chang, C. North, and M. Zhou. Semantic Interaction: Coupling Cognition and Computation through Usable Interactive Analytics. *IEEE Computer Graphics and Applications*, 35(4):94–99, 2015. doi: 10.1109/MCG.2015.91 4

[28] P. Federico, M. Wagner, A. Rind, A. Amor-Amorós, S. Miksch, and W. Aigner. The Role of Explicit Knowledge: A Conceptual Model of Knowledge-Assisted Visual Analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 92–103, 2017. doi: 10.1109/VAST.2017.8585498 1

[29] M. Gleicher. Explainers: Expert Explorations with Crafted Projections. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2042–2051, 2013. doi: 10.1109/TVCG.2013.157 4

[30] T. Gschwandtner and O. Erhart. Know Your Enemy: Identifying Quality Problems of Time Series Data. In *2018 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 205–214, 2018. doi: 10.1109/PacificVis.2018.00034 1, 4

[31] A. Hamdi, K. Shaban, A. Erradi, A. Mohamed, S. K. Rumi, and F. D. Salim. Spatiotemporal data mining: A survey on challenges and open problems. *Artificial Intelligence Review*, 55(2):1441–1488, 2022. doi: 10.1007/s10462-021-09994-y 1

[32] V. Hemming, M. A. Burgman, A. M. Hanea, M. F. McBride, and B. C. Wintle. A practical guide to structured expert elicitation using the IDEA protocol. *Methods in Ecology and Evolution*, 9(1):169–180, 2018. doi: 10.1111/2041-210X.12857 3

[33] J. Hullman and A. Gelman. Designing for Interactive Exploratory Data Analysis Requires Theories of Graphical Inference. *Harvard Data Science Review*, 3(3), 2021. doi: 10.1162/99608f92.3ab8a587 2

[34] J. Hullman, M. Kay, Y.-S. Kim, and S. Shrestha. Imagining Replications: Graphical Prediction & Discrete Visualizations Improve Recall & Estimation of Effect Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):446–456, 2018. doi: 10.1109/TVCG.2017.2743898 3, 4, 5

[35] J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay. In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):903–913, 2019. doi: 10.1109/TVCG.2018.2864889 3

[36] J. Hullman, P. Resnick, and E. Adar. Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering. *PLOS ONE*, 10(11):e0142444, 2015. doi: 10.1371/journal.pone.0142444 3

[37] A. Jena, U. Engelke, T. Dwyer, V. Raiamanickam, and C. Paris. Uncertainty Visualisation: An Interactive Visual Survey. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 201–205, 2020. doi: 10.1109/PacificVis48177.2020.1014 3

[38] G. Jones and W. O. Johnson. Prior Elicitation: Interactive Spreadsheet Graphics With Sliders Can Be Fun, and Informative. *The American Statistician*, 68(1):42–51, 2014. doi: 10.1080/00031305.2013.868828 3

[39] J. Kadane and L. J. Wolfson. Experiences in Elicitation. *Journal of the Royal Statistical Society Series D: The Statistician*, 47(1):3–19, 1998. doi: 10.1111/1467-9884.00113 2

[40] J. B. Kadane, J. M. Dickey, R. L. Winkler, W. S. Smith, and S. C. Peters. Interactive Elicitation of Opinion for a Normal Linear Model. *Journal of the American Statistical Association*, 75(372):845–854, 1980. doi: 10.2307/2287171 4

[41] A. Kale, Z. Guo, X. L. Qiao, J. Heer, and J. Hullman. EVM: Incorporating Model Checking into Exploratory Visual Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):208–218, 2024. doi: 10.1109/TVCG.2023.3326516 3, 4

[42] A. Kamal, P. Dhakal, A. Y. Javaid, V. K. Devabhaktuni, D. Kaur, J. Zaientz, and R. Marinier. Recent advances and challenges in uncertainty visualization: A survey. *Journal of Visualization*, 24(5):861–890, 2021. doi: 10.1007/s12650-021-00755-1 3

[43] S. Kandel, A. Paepcke, J. Hellerstein, and J. Heer. Wrangler: Interactive Visual Specification of Data Transformation Scripts. In *CHI 2011*, CHI '11, pp. 3363–3372. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1979444 4

[44] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson. When (ish) is

My Bus?: User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In *CHI 2016*, pp. 5092–5103. ACM, San Jose California USA, 2016. doi: 10.1145/2858036.2858558 3

[45] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann. *Mastering the Information Age: Solving Problems with Visual Analytics*. Eurographics Association, 2010. doi: handle/10.2312/14803 1, 3

[46] D. Kerrigan, J. Hullman, and E. Bertini. A Survey of Domain Knowledge Elicitation in Applied Machine Learning. *Multimodal Technologies and Interaction*, 5(12):73, 2021. doi: 10.3390/mti5120073 2

[47] C. Kinkeldey, M. , Alan M., R. , Maria, and J. and Schiewe. Evaluating the effect of visually represented geodata uncertainty on decision-making: Systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science*, 44(1):1–21, 2017. doi: 10.1080/15230406.2015.1089792 3

[48] C. Kinkeldey, A. M. MacEachren, and J. Schiewe. How to Assess Visual Communication of Uncertainty? A Systematic Review of Geospatial Uncertainty Visualisation User Studies. *The Cartographic Journal*, 51(4):372–386, 2014. doi: 10.1179/1743277414Y.0000000099 3

[49] P. M. Kuhnert, T. G. Martin, and S. P. Griffiths. A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters*, 13(7):900–914, 2010. doi: 10.1111/j.1461-0248.2010.01477.x 2

[50] H. Lam. A Framework of Interaction Costs in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1149–1156, 2008. doi: 10.1109/TVCG.2008.109 5

[51] H. Lin, D. Akbaba, M. Meyer, and A. Lex. Data Hunches: Incorporating Personal Knowledge into Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):504–514, 2022. doi: 10.1109/TVCG.2022.3209451 3

[52] N. McCurdy, J. Gerdes, and M. Meyer. A Framework for Externalizing Implicit Error Using Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):925–935, 2019. doi: 10.1109/TVCG.2018.2864913 3

[53] R. McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Texts in Statistical Science. Chapman & Hall/CRC, Boca Raton, second edition ed., 2020. 3

[54] P. Mikkola, O. A. Martin, S. Chandramouli, M. Hartmann, O. A. Pla, O. Thomas, H. Pesonen, J. Corander, A. Vehtari, S. Kaski, P.-C. Bürkner, and A. Klami. Prior Knowledge Elicitation: The Past, Present, and Future. *Bayesian Analysis*, 19(4):1129–1161, 2024. doi: 10.1214/23-BA1381 2

[55] E. Newburger, M. Correll, and N. Elmqvist. Fitting Bell Curves to Data Distributions Using Visualization. *IEEE Transactions on Visualization and Computer Graphics*, pp. 5372–5383, 2022. doi: 10.1109/TVCG.2022.3210763 4

[56] A. O'Hagan. Expert Knowledge Elicitation: Subjective but Scientific. *The American Statistician*, 73(sup1):69–81, 2019. doi: 10.1080/00031305.2018.1518265 2, 3

[57] A. O'Hagan, C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. *Uncertain Judgements: Eliciting Experts' Probabilities*. Wiley, 1 ed., 2006. doi: 10.1002/0470033312 2, 3

[58] R. A. O'Leary, S. L. Choy, J. V. Murray, M. Kynn, R. Denham, T. G. Martin, and K. Mengersen. Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby *Petrogale penicillata*. *Environmetrics*, 20(4):379–398, 2009. doi: 10.1002/env.935 4

[59] G. Panagiotidou, H. Lamqaddam, J. Poblome, K. Brosens, K. Verbert, and A. V. Moere. Communicating Uncertainty in Digital Humanities Visualization Research. *IEEE Transactions on Visualization and Computer Graphics*, pp. 635–645, 2022. doi: 10.1109/TVCG.2022.3209436 1, 3

[60] G. J. Quadri and P. Rosen. A Survey of Perception-Based Visualization Studies by Task. *IEEE Transactions on Visualization and Computer Graphics*, pp. 5026–5048, 2021. doi: 10.1109/TVCG.2021.3098240 4

[61] S. Renooij. Probability elicitation for belief networks: Issues to consider. *The Knowledge Engineering Review*, 16(3):255–269, 2001. doi: 10.1017/S0269888901000145 3

[62] H. Ribičić, J. Waser, R. Gurbat, B. Sadransky, and M. E. Gröller. Sketching Uncertainty into Simulations. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2255–2264, 2012. doi: 10.1109/TVCG.2012.261 4

[63] N. Rodrigues, C. Schulz, S. Döring, D. Baumgartner, T. Krake, and D. Weiskopf. Relaxed Dot Plots: Faithful Visualization of Samples and Their Distribution. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):278–287, 2023. doi: 10.1109/TVCG.2022.3209429 3

[64] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):240–249, 2016. doi: 10.1109/TVCG.2015.2467591 3

[65] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge Generation Model for Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, 2014. doi: 10.1109/TVCG.2014.2346481 3

[66] M. Sedlmair, M. Meyer, and T. Munzner. Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012. doi: 10.1109/TVCG.2012.213 2

[67] D. Spiegelhalter. Risk and Uncertainty Communication. *Annual Review of Statistics and Its Application*, 4(1):31–60, 2017. doi: 10.1146/annurev-statistics-010814-020148 2

[68] D. A. Szafir, S. Haroz, M. Gleicher, and S. Franconeri. Four types of ensemble coding in data visualizations. *Journal of Vision*, 16(5):11–11, 2016. doi: 10.1167/16.5.11 4

[69] S. Tak, A. Toet, and J. van Erp. The Perception of Visual Uncertainty Representation by Non-Experts. *IEEE Transactions on Visualization and Computer Graphics*, 20(6):935–943, 2014. doi: 10.1109/TVCG.2013.247 3

[70] J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel. A typology for visualizing uncertainty. In R. F. Erbacher, J. C. Roberts, M. T. Grohn, and K. Borner, eds., *Electronic Imaging 2005*, p. 146. San Jose, CA, 2005. doi: 10.1117/12.587254 1, 3, 4

[71] P. N. Truong, G. B. M. Heuvelink, and J. P. Gosling. Web-based tool for expert elicitation of the variogram. *Computers & Geosciences*, 51:390–399, 2013. doi: 10.1016/j.cageo.2012.08.010 4

[72] A. C. Valdez, M. Ziefle, and M. Sedlmair. Priming and Anchoring Effects in Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):584–594, 2018. doi: 10.1109/TVCG.2017.2744138 4

[73] R. van de Schoot, S. Depaoli, R. King, B. Kramer, K. Märtens, M. G. Tadesse, M. Vannucci, A. Gelman, D. Veen, J. Willemsen, and C. Yau. Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1):1, 2021. doi: 10.1038/s43586-020-00001-2 4

[74] D. L. Vannette and J. A. Krosnick, eds. *The Palgrave Handbook of Survey Research*. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-319-54395-6 2

[75] M. Wagner, D. Slijepcevic, B. Horsak, A. Rind, M. Zeppelzauer, and W. Aigner. KAVAGait: Knowledge-Assisted Visual Analytics for Clinical Gait Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 25(3):1528–1542, 2019. doi: 10.1109/TVCG.2017.2785271 4

[76] E. Wall, S. Das, R. Chawla, B. Kalidindi, E. T. Brown, and A. Endert. Podium: Ranking Data Using Mixed-Initiative Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):288–297, 2018. doi: 10.1109/TVCG.2017.2745078 4

[77] T. S. Wallsten and D. V. Budescu. State of the Art—Encoding Subjective Probabilities: A Psychological and Psychometric Review. *Management Science*, 29(2):151–173, 1983. doi: 10.1287/mnsc.29.2.151 3

[78] F. Yang, M. Hedayati, and M. Kay. Subjective Probability Correction for Uncertainty Representations. In *CHI 2023*. ACM, 2023. doi: 10.1145/3544548.3580998 3