EuroVis Workshop on Visual Analytics (2015)
E. Bertini and J. C. Roberts (Editors)

Visual Analytics and Uncertainty: Its Not About the Data

A.M. MacEachren[1]

[1]The Pennsylvania State University, University Park, PA, USA

Abstract

Uncertainty visualization research has a long history, with contributions from scientific, information, geo-graphic and other visualization perspectives as well as from cognitive and HCI perspectives. But we still do not have generally accepted strategies for leveraging visualization to cope with uncertainty. Here, I argue that taking a visual analytics rather than visualization perspective can overcome this inertia. While uncertainty visualization research has focused on visually signifying and interacting with data uncertainty, taking a visual analytics approach recognizes that the challenge is about much more than uncertain data. The larger challenge is to enable reasoning under uncertainty (in all its forms). In this short paper, I sketch elements of what we know and outline some key challenges for developing visual analytics methods and tools that enable users to cope with uncertainty throughout the processes of sensemaking, decision-making, and action-taking.

Categories and Subject Descriptors (according to ACM CCS): H.1.2 Use/Machine Systemshuman information processing; H.5.2 User Interfaces-Theory and methods/Graphical user interfaces (GUI)

1 Introduction

Within the visualization fields, uncertainty has generally been treated as an attribute of data, with the research focus on developing methods to signify and interact with data uncertainty, assess method communication effectiveness and/or efficiency, and assess impacts of uncertainty visualization methods on interpretations and decisions. In previous papers, colleagues and I have reviewed this work, with an emphasis on uncertainty related to geographical information [KMRSed, KMS14, THM*05, MRH*05], thus I will only summarize a few key outcomes of those reviews here.

As outlined in Kinkeldey [KMS14], three primary distinctions have been made in research to develop and assess uncertainty visualization methods: (a) intrinsic/extrinsic signification: where signification of data and data uncertainty are either integrated (e.g., bi-variate color) or visually distinct (e.g., glyphs depicting uncertainty on a heatmap depicting data); (b) coincident/adjacent display: where both data and data uncertainty are in the same versus different views; and (c) static/dynamic: where unchanging visual signification contrasts with either/both data driven or user driven changes (e.g., animated depiction of changing uncertainty over time, user probing to focus attention on uncertainty for selected features, or dynamic linking of views).

Lack of consistency in empirical research about effectiveness of uncertainty visualization strategies makes studies difficult to compare and generalizations speculative [KMS14]. In relation to the first distinction (a above), evidence supports a contention that extrinsic signification that is visually separable (e.g., texture for uncertainty superimposed on a color-coded heat map for data) enables users to interpret both data and uncertainty depictions more accurately than intrinsic signification (e.g., bi-variate color for data + uncertainty) [Ret12, MBP98]. But, recent evidence suggests that extrinsic signification supporting the ability to interpret data and uncertainty independently can cause users to ignore uncertainty entirely when faced with a decision-making task [Ret12].

Results of empirical comparison of coincident and adjacent display (b above) has produced similar mixed results [KMS14]. Adjacent display takes extrinsic signification to the extreme, putting data and uncertainty in different views. If the task is to interpret each independently, this simplifies the perceptual task; on the other hand it makes it more likely that uncertainty will be ignored when a complex task is carried out. Success of coincident display will vary depending on whether signification is extrinsic or intrinsic and will be increasingly less successful as display complexity increases.

For the third distinction (c, which includes dynamism, both via animation and interaction), there has been less em-

DOI: 10.2312/eurova.20151104

pirical research and little evidence thus far that animated display of uncertainty has advantages over static display. Various authors have proposed interactive interfaces to support signification of uncertainty in ways that provide the user with control of its prominence in the display (e.g., [HM96, Fis94, SZD*10]) along with potential support for adjacent display through dynamic linking. There has been a surprising paucity of empirical research to assess such systems in spite of the long history of dynamic linking as a strategy to support effective multi-view display [Rob05]; Sanyal, et al [SZD*10] do present results of a qualitative assessment of a highly interactive system with multiple methods for depicting numerical weather predictions and their uncertainty.

Past research on *communication effectiveness* of uncertainty visualization has been complemented by research on uncertainty visualization *effects*; this work emphasizes effects for decisions [BCJ*11, CHL13, FFK14, LB00, RHFL14] (e.g., are resulting decisions different and, if so, better) and for risk assessment in contexts ranging from medical diagnosis (e.g., [MDF12]) to climate change impacts (e.g., [Ret13]). Colleagues and I have reviewed more than 40 empirical (mostly geospatial) assessments of uncertainty visualization effects [KMRSed]. Key findings are that: (a) these studies have conflicting results due to inconsistent methodologies, even on the question of whether uncertainty depiction helps or hinders decisions; and (b) the judgment/decision tasks involved are typically extremely simple ones, unrepresentative of real-world decisions. Thus, even when we assume that the uncertainty *communication* is successful, understanding the *effect* of uncertainty visualization for risk assessment, decisions, action-taking, or other objectives that the visualization supports remains a largely open research challenge, as is how to design uncertainty visualization to achieve positive effects.

My contention here is that our lack of progress in addressing the challenges of uncertainty stems from two primary impediments. *First*, empirical research has been ad hoc, with inconsistent, uncomparable methodologies applied. Thus, results are less than the sum of the parts due to contradictions. The review papers cited above provide some suggestions for addressing this impediment, thus, I will not address it further here. *Second*, the field has devoted almost all of its attention to *data uncertainty* and its signification. Almost no attention has been given to *reasoning/decision-making under uncertainty*. While it is understandable that research in information/scientific visualization has focused on visual signification of data uncertainty, it is surprising that there has been so little attention to the challenge of reasoning with uncertainty in the visual analytics. Here, I focus the remainder of the paper on sketching some ideas toward addressing this second impediment. The goal is to provide a base for developing, subsequently, a conceptual framework to understand and facilitate visually-enabled reasoning/decision-making under uncertainty and to use that framework to develop visual analytics methods to achieve this objective.

2 Conceptualizing Uncertainty

There have been a range of approaches to conceptualizing uncertainty in the cognitive and decision sciences that are relevant to developing visual analytics methods and tools that support analytical reasoning under uncertainty. Here, it is impossible to cover these approaches comprehensively. Instead, I highlight three approaches that address core perspectives: (1) understanding the basis for uncertainty; (2) understanding levels of uncertainty; and (3) understanding the role of information and knowledge in relation to uncertainty.

2.1 Basis for uncertainty

In an important early paper, Kahneman and Tversky introduce a model of the basis of uncertainty in human reasoning that underpins much of their subsequent work on judgment and decision-making [KT82]. The key contention (Figure 1) is that uncertainty can be attributed to both the human thinkers internal knowledge and to the external world (which can be difficult or impossible to predict due to dispositions outside the control of the decision-maker; dispositions here are the tendency of something to act in a certain manner under given circumstances). Externally they distinguish two bases for uncertainty: distributional (where the case in question is an instance of a class of similar caseswith relative outcome frequencies known or estimateable) and singular (where probabilities are assessed by propensities of a particular case at hand). Internal bases are partitioned into: reasoned (that could reflect a process of sifting and weighing evidence and arguments) and introspective (that involve judgment of the strength of an association).

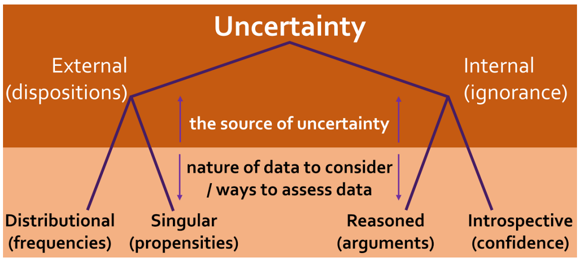

Figure 1: *Variants of uncertainty proposed by Kahneman and Tversky; extended from Figure 2 in [KT82].*

2.2 Levels of uncertainty

Courtney [Cou03] identified four levels of uncertainty, each of which requires different strategies (and decision aids) to address: (a) *clear enough future*: where point forecasts exist that are 'close enough' for the decision at hand (e.g., location decision certainty: will a Segafredo Café opened at Kärntner Str and Philharmoniker Str in Vienna make a profit?); (b) *choice among alternate futures*: with a limited set of possible outcomes defined, one of which will occur (e.g., given a choice of two intersections vs. a location inside the train station for a Segafredo Café, which will be the most successful?); (c) *range of futures*: with a possible range definable within some bounds (e.g., given an estimate that 35-55% of Starbucks patrons will switch allegiance, will a Segafredo Café be successful across the street?) and (d) *true ambiguity*: given which it is impossible to even define the range of possible outcomes uncertainties may be both unknown and unknowable (e.g., how will Star Trek like food replicators that bring a cup of 'real' Segafredo espresso and pastry to you impact the success of physical Segafredo Café locations?).

This continuum ranges from situations in which visual, decision-support tools should focus on helping the decision-maker understand the range of outcomes and consequences that may result from the decision to situations in which a more iterative, speculative what-if process needs to be supported (in which a wide range of uncertainties can come into play). For the latter, analytical tools need to support: gathering more information, working backwards from potential options to develop scenarios that could support them, and testing scenarios and their assumptions by comparison to analogies, references cases, and other past experiences.

2.3 Information & knowledge in relation to uncertainty

In relation to organizational decision-making, Zack [Zac07] characterizes the information and knowledge-based problems faced by organizations. He identifies four categories as part of a 2x2 typology, one cell of which is labeled as uncertainty (Figure 2). But, in relation to the external-internal basis for uncertainty outlined above (as it relates to decision-making), it seems reasonable to consider all four categories as reasons for 'uncertainty' about which decision is best (with uncertainty about information being one component).

processing problem / what is processed	Information	Knowledge
Lack of …	uncertainty	ambiguity
Diversity of …	complexity	equivocality

→ analysis → interpretation

↳ acquire ↳ restrict

Figure 2: *Four reasons for uncertainty about which decision is best; derived from Figure 1 in [Zac07, p. 1665]*

The categories focus on information input to decisions and knowledge available through which to interpret the information. Specifically, they address the challenges of lack versus diversity of inputs along with the distinction between information and the knowledge required for its interpretation. To some extent, this distinction echoes that made by Kahneman and Tversky [KT82, p. 151] between external and internal uncertainty. Both components related to information require tools that support analysis (to turn data into information in the first place and then to support development of decision options and identification of potential outcomes and consequences. Lack of *information* is labeled as *uncertainty*: defined as insufficient information to accomplish a task. Addressing 'uncertainty' in this narrow sense requires visual analytics tools that support information foraging along with analysis. Alternatively, particularly in the era of big data, situations arise with information volumes and diversity rather than limits. Zack calls this problem *complexity*: more information than can be processed. To address complexity, visual analytics tools need to support analysis combined with selection and abstraction (e.g., tools for filtering, topic modeling and clustering, dimension-reduction).

The *knowledge* column, described by Zack [Zac07, p. 1665], relates to context needed for information interpretation. Lack of knowledge is labeled ambiguity. This can be considered as the lack of an appropriate "frame of "reference through which to interpret the information. Visual analytics tools can to build the frame needed to create context. An example of an effort to develop such visual analytics tools is found in [TM12]. Just as with information, there can be a diversity of reference frames leading to a diversity of possible interpretations. Uncertainty about which one to accept as a basis for decisions is termed equivocation: competing/contradictory reference frames. Equivocation yields confusion and requires visual analytics tools for re-framing, analysis of competing hypotheses, deliberation, negotiation, etc.

3 Reasoning under uncertainty

Reasoning under uncertainty is a complex topic that has been addressed over many decades from many disciplinary perspectives; Google Scholar lists over 7,700 entries for the phrase. But, few papers in visualization or visual analytics have directed specific attention to reasoning under uncertainty. This situation is surprising given the 2005 definition of visual analytics as the science of analytical reasoning facilitated by visual interfaces [TC05, p. 4]. This emphasis on analytical reasoning suggests that visual analytics is the ideal conceptual framework for connecting advances in uncertainty visualization with the complex challenge of enabling reasoning under uncertainty. This should be particularly true for real-world problems in which multiple forms of uncertainty, both external and internal in Kahneman and Tverskys [KT82] terms, are involved and in which the reasoning and decision-making process is often based upon large amounts of messy, inconsistent, and changing data.

One basis for understanding reasoning is the classic work on heuristics reported in the Tversky and Kahneman [KT82] paper cited above. Specifically, they identify three heuristics used in judgments under uncertainty.

• *Representativeness*: When asked to judge whether an entity is a member of a class, people consider how 'representative' the entity is of the class. A common flaw with this heuristic (one of six flaws outlined) is that base rates (i.e., the overall probability that any entity will be part of the class) are frequently neglected. One example is the task of judging how likely it is that a particular social media post was by someone from a particular country. Using a representativeness heuristic, an analyst might consider word use, sentence structure, places mentioned to decide without considering the proportion of all tweets that are issued from the country in question. Another example is found in a study by Micallef, et al [MDF12] focused on estimating the probability of having cancer based on a positive test for cancer. Results suggest that with no visual aids, participants ignored base rates and that graphic depictions without numerical information work better than those with numerical information as a strategy to overcome this bias. Overall, including visual depictions that remind the analyst to consider a range of statistical factors important in weighing evidence can help to counter misapplication of the representative heuristic.

• *Availability*: Individual knowledge and memory underlie the availability heuristic; here, ability to recall concrete instances / occurrences is used as a heuristic to judge the probability of an event or frequency of a class. For example, a coastal vacationer who knew a Hurricane Katrina victim is more likely to decide that the risk of an impending hurricane is high enough to evacuate than someone who once evacuated a vacation site only to have the hurricane fizzle out and do no damage. While availability is in part a function of probability or frequency, it is also affected by a wide range of other factors underlying long-term knowledge held by an individual as well as short term knowledge and recent experiences. Kahnemahn and Tversky [KT82] cite four biases that impact availability: (a) retrievability of information (frequency judgments are biased by familiarity of entities); (b) effectiveness of search set (frequency judgments are biased by the number of instances of a set that can be thought of); (c) imagability (the ability to conceive of alternative situations or outcomes positively biases estimates of frequency); and (d) illusory correlation (the propensity to assume that if one feature exists others will too, biases probability and frequency judgments whenever one feature is present). In all cases, visual analytics methods can help counter availability bias by providing access to representative evidence.

• *Adjustment from an anchor*: People have a propensity to make any estimate in relation to some, often arbitrary, initial value. Initial values can be suggested by almost anything; experiments have shown that randomly generated numbers can act to anchor estimates to which those numbers have absolutely no relation [MS00]. Anchoring results in a numerical estimate being insufficiently different from the anchor; any anchor can prompt biases in evaluating likelihood of conjunctive or disjunctive events (i.e., the chance of multiple specific events in a set number of tries versus the chance of any one event across multiple tries) and in estimates of confidence intervals that specific outcomes will be achieved. For example, anchoring based on current global temperature is likely to produce overly conservative max/min estimates of expected global temperature in 100 years.

The ideas above just scratch the surface of research on reasoning, judgment, and decision-making under uncertainty. Integrating insights from this work is a core challenge for visual analytics in support of reasoning with uncertain data about uncertain outcomes.

Zuk and Carpendale [ZC07] report one of the few steps in this direction. One component of this work was a reconceptualization of the typology of uncertainty that colleagues and I presented in [THM*05]. Zuk and Carpendale use this typology as a base to delineate the types of reasoning uncertainty that visualization needs to support in relation to each of the 9 uncertainty categories: (1) *Currency/Timing*—temporal gaps between assumptions and reasoning steps; (2) *Credibility*—heuristic accuracy & bias of analyst; (3) *Lineage*—conduit of assumptions, reasoning, revision, and presentation; (4) *Subjectivity*—amount of private knowledge or heuristics utilized; (5) *Accuracy/Error*—difference between heuristic & algorithm (e.g. Bayesian); (6) *Precision*—variability of heuristics and strategies; (7) *Consistency*—extent to which heuristic assessments agree; (8) *Interrelatedness*—heuristic & analyst independence; (9) *Completeness*—extent to which knowledge is complete.

4 Challenges

Past uncertainty visualization research has focused almost exclusively on developing and evaluating methods to depict numerical (or ordinal) measures of uncertainty (or probability, variance, etc). Even when the role of uncertainty in decisions has been the focus, research has been restricted to assessing whether methods to visualize data uncertainty lead to different/better (usually trivial) decisions. Missing (almost entirely) is research on visual methods to enable reasoning and decisions under uncertainty.

In the context of real-world decisions, uncertainty related to reasoning with information in context is more important than uncertainty related to data. Thus, in visualization/visual analytics, we have missed more than half of the picture. Emphasis has been on (a) developing methods to depict data uncertainty visually and (b) determining how effectively and efficiently they can be used. Attention is needed to the role of visual interfaces in reasoning under uncertainty – our focus needs to switch from depicting uncertain data to supporting reasoning and decisions in situations for which not only data but the problem itself, options, potential outcomes, implications of outcomes, etc. are uncertain. Doing so requires addressing (at least) these following challenges:

• understand components of uncertainty and their relationships to: use domains, information needs, and expertise;
• understand how knowledge of uncertainty (or lack of it) influences reasoning, decision making, and outcomes;
• understand how (or whether) uncertainty visualization aids / hinders exploratory analysis, reasoning, and decisions;
• leverage understanding to develop useful/usable methods/tools to: signify multiple kinds of uncertainty; interact with uncertainty depictions; support reasoning/decisions under uncertainty; capture/encode analysts uncertainty;
• assess usability and utility of the methods/tools – design studies for reproducibility and comparability.

References

[BCJ*11] BISANTZ A., CAO D., JENKINS M., PENNATHUR P., FARRY M., ROTH E., POTTER S., PFAUTZ J.: Comparing uncertainty visualizations for a dynamic decision-making task. *Journal of Cognitive Engineering and Decision Making 5*, 3 (2011), 277–293. 2

[CHL13] COX J., HOUSE D., LINDELL M.: Visualizing uncertainty in predicted hurricane tracks. *International Journal for Uncertainty Quantification 3*, 2 (2013), 143–156. 2

[Cou03] COURTNEY H.: Decision-driven scenarios for assessing four levels of uncertainty. *Strategy Leadership 31*, 1 (2003), 14–22. 2

[FFK14] FERREIRA N., FISHER D., KONIG A. C.: Sample-oriented task-driven visualizations: allowing users to make better, more confident decisions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (2014), ACM, pp. 571–580. 2

[Fis94] FISHER P.: Hearing the reliability in classified remotely sensed images. *Cartography and Geographic Information Systems 21*, 1 (1994), 31–36. 2

[HM96] HOWARD D., MACEACHREN A. M.: Interface design for geographic visualization: Tools for representing reliability. *Cartography and Geographic Information Systems 23*, 2 (1996), 59–77. 2

[KMRSed] KINKELDEY C., MACEACHREN A. M., RIVEIRO M., SCHIEWE J.: Evaluating the effect of visually represented geodata uncertainty on decision making: Systematic review, lessons learned and recommendations. *Cartography Geographic Information Science* (submitted). 1, 2

[KMS14] KINKELDEY C., MACEACHREN A. M., SCHIEWE J.: How to assess visual communication of uncertainty? a systematic review of geospatial uncertainty visualization user studies. *Cartographic Journal 51*, 4 (2014), 372–386. 1

[KT82] KAHNEMAN D., TVERSKY A.: Variants of uncertainty. *Cognition 11*, 2 (1982), 143–157. 2, 3, 4

[LB00] LEITNER M., BUTTENFIELD B.: Guidelines for the display of attribute certainty. *Cartography and Geographic Information Science 27*, 1 (2000), 3–14. 2

[MBP98] MACEACHREN A. M., BREWER C. A., PICKLE L. W.: Visualizing georeferenced data: Representing reliability of health statistics. *Environment and Planning: A 30* (1998), 1547–1561. 1

[MDF12] MICALLEF L., DRAGICEVIC P., FEKETE J.: Assessing the effect of visualizations on bayesian reasoning through crowdsourcing. *Visualization and Computer Graphics, IEEE Transactions on 18*, 12 (2012), 2536–2545. 2, 4

[MRH*05] MACEACHREN A. M., 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*, 3 (2005), 139–160. 1

[MS00] MUSSWEILER T., STRACK F.: Numeric judgments under uncertainty: The role of knowledge in anchoring. *Journal of Experimental Social Psychology 36*, 5 (2000), 495–518. 4

[Ret12] RETCHLESS D.: Mapping climate change uncertainty: Effects on risk perceptions and decision making, paper gc43b-1024, Dec. 3-7, 2012 2012. 1

[Ret13] RETCHLESS D. P.: Communicating climate change: spatial analog versus color-banded isoline maps with and without accompanying text. *Cartography and Geographic Information Science 41*, 1 (2013), 55–74. 2

[RHFL14] RIVEIRO M., HELLDIN T., FALKMAN G., LEBRAM M.: Effects of visualizing uncertainty on decision-making in a target identification scenario. *Computers Graphics* (2014). 2

[Rob05] ROBERTS J.: Exploratory visualization with multiple linked views. In *Exploring Geovisualization*, Dykes J., MacEachren A. M., Kraak M.-J., (Eds.). Elsevier Science, Amsterdam, 2005. 2

[ST92] SHAFIR E., TVERSKY A.: Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology 24*, 4 (1992), 449–474.

[SZD*10] SANYAL J., ZHANG S., DYER J., MERCER A., AMBURN P., MOORHEAD R. J.: Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *Visualization and Computer Graphics, IEEE Transactions on 16*, 6 (2010), 1421–1430. 2

[TC05] THOMAS J. J., COOK K. A.: *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, Los Alamos, CA, 2005. 3

[THM*05] THOMSON J., HETZLER B., MACEACHREN A., GAHEGAN M., PAVEL M.: Typology for visualizing uncertainty, January 16-20, 2005 2005. 1, 4

[TK74] TVERSKY A., KAHNEMAN D.: Judgement under uncertainty: Heuristics and biases. *Science 185* (1974), 1124–1131.

[TM12] TOMASZEWSKI B., MACEACHREN A. M.: Geovisual analytics to support crisis management: Information foraging for geo-historical context. *Information Visualization invited extension of paper originally published in Proceedings of IEEE VAST 2010 11*, 4 (2012), 339–359. 3

[Zac07] ZACK M. H.: The role of decision support systems in an indeterminate world. *Decision Support Systems 43*, 4 (2007), 1664–1674. 3

[ZC07] ZUK T., CARPENDALE S.: Visualization of uncertainty and reasoning. In *Smart Graphics, LNCS 4569*, Butz A. e. a., (Ed.). Springer-Verlag, Berlin, 2007, pp. 164–177. 4