Original Article

Revealing uncertainty for information visualization

Meredith Skeels[a]
Bongshin Lee[b,*]
Greg Smith[b] and
George G. Robertson[b]

[a]Biomedical and Health Informatics, University of Washington, Seattle, Washington 98195, USA.
E-mail: mskeels@u.washington.edu
[b]Microsoft Research, One Microsoft Way, Redmond, Washington 98052, USA.
E-mails: bongshin@microsoft.com; gregsmi@microsoft.com; ggr@microsoft.com

*Corresponding author.

Received: 21 December 2008
Accepted: 8 January 2009

Abstract Uncertainty in data occurs in domains ranging from natural science to medicine to computer science. By developing ways to include uncertainty in our information visualizations, we can provide more accurate depictions of critical data sets so that people can make more informed decisions. One hindrance to visualizing uncertainty is that we must first understand what uncertainty is and how it is expressed. We reviewed existing work from several domains on uncertainty and created a classification of uncertainty based on the literature. We empirically evaluated and improved upon our classification by conducting interviews with 18 people from several domains, who self-identified as working with uncertainty. Participants described what uncertainty looks like in their data and how they deal with it. We found commonalities in uncertainty across domains and believe our refined classification will help us in developing appropriate visualizations for each category of uncertainty.
Information Visualization (2010) **9,** 70–81. doi:10.1057/ivs.2009.1; published online 21 May 2009

Keywords: uncertainty visualization; uncertainty classification; qualitative research; user-centered design

Introduction

When information is shown in a computer interface, it often appears absolute. The native machine or language data types used to store numerical data enforce a very high level of precision. There is no sense of the level of certainty in that data or the degree to which the data are only possibly true. However, in reality, data are rarely absolutely certain. By developing ways to make the uncertainty associated with data more visible, we can help users better understand and use their data. Making uncertainty visible will also promote the appropriate use of data by helping to avoid forgetting about or hiding uncertainty and will improve the way people communicate certainty in data to others. Thomas and Cook recommend development of visual representations that allow the analyst to understand the uncertainty inherent in a visual analytics application.[1] Johnson and Sanderson suggested possible future directions for visualizing errors and uncertainty.[2]

Before we can visualize uncertainty we need to understand what it is. There has been a significant amount of research on uncertainty in fields such as information theory,[3] Bayesian theory,[4] probabilistic reasoning,[5] and fuzzy set theory.[6] However, they mainly focused on how to compute uncertainty by developing a formal mathematical method rather than identifying what kinds of uncertainty people encounter in their data and how to show uncertainty to others. Unfortunately, uncertainty is not always expressed as a quantifiable probability and there has also been research on the problem of quantifying textual or verbal descriptions of certainty.[7] Much of the previous work on both the visualization of uncertainty and the classification of uncertainty occurs within isolated domains. Cartography and geography, intelligence analysis and work in weather modeling are all domains where uncertainty has been discussed. Uncertainty work done in disparate domains shares some common themes, but it does not reach agreement. The predominant consensus among papers on uncertainty appears to be that uncertainty has been defined in many ways and is referred to inconsistently in a variety of fields. For example, MacEachren *et al* state, 'Information uncertainty is a complex concept with many interpretations across knowledge domains and application contexts'[8] and Norton *et al* warn us that, 'An important barrier to achieving a common understanding or interdisciplinary framework is the diversity of meanings associated with terms such as "uncertainty" and "ignorance," both within and between disciplines'.[9]

Work on uncertainty within domains can inform the design of visualizations, but unfortunately uncertainty is not referred to consistently. Within this amorphous concept there are many types of uncertainty that almost certainly warrant different visualization techniques. In order to create techniques for visualizing uncertainty that span domains, we need a better understanding of what it is we are trying to visualize. To that end, we have reviewed existing work on uncertainty within a number of domains and created a classification of uncertainty for use in information visualization. We empirically evaluated and improved upon the classification with an interview-based user study. This classification has consistencies with previous work; however, unlike much of the previous work, it spans multiple domains and is focused on differentiating types of uncertainty for the purposes of visualization.

We begin by reviewing previous work on visualizing and classifying uncertainty. We then describe the domain-spanning classification we developed and used as a starting point for our evaluation. The remainder of the paper is devoted to discussing the refinement of our classification based on the interview data.

Previous Work

Previous research on visualizing uncertainty

There is a limited amount of research on visualizing uncertainty and much of it is in the field of geographic visualization, geographic information science and scientific visualization. The main techniques developed include adding glyphs,[10,11] adding geometry, modifying geometry,[12] modifying attributes, animation,[10,13] and sonification.[14] These techniques have been applied to a variety of applications such as fluid flow, surface interpolants and volumetric rendering. The visualization by Pang and Freeman[15] shows differences between 3D surfaces generated by various rendering algorithms. CandidTree shows two types of structural uncertainty based on the differences between two tree structures.[16] LISTEN visualizes geometric uncertainty using sound, which represents the difference between geometric quantities obtained by two interpolants.[14] DaVis visualizes the differences between two data sets containing sensor readings, such as traffic speed, volume and occupancy.[17] It also shows the missing values with an empty cell. MANET uses a complementary display to show the proportion of missing data,[18] while Restorer uses grayscale to indicate missing data on a color map.[19] Djurcilov and Pang discussed several visualizations to deal with gridded datasets with a large number of missing or invalid data.[20] Collins and Penn developed a visualization that shows translation uncertainty using color saturation, size and transparency of the node border in the translated message graph.[21] Olston and Mackinlay introduced visualizations to address two forms of uncertainty: error bars for showing statistical uncertainty and ambiguation for showing bounded uncertainty.[22]

Unfortunately, most of these uncertainty visualizations were isolated efforts designated for a specific purpose. To move forward with the challenges of visualizing uncertainty and creating interfaces for interacting with uncertainty in data, we need to have a model of uncertainty that can cover the needs of users across different domains.

Previous classifications of uncertainty

We examined the definitions and classifications of uncertainty within domains in order to define a classification composed of the overlap among several existing schemes of uncertainty. Within the domain of geography, MacEachren *et al* published a review of models of information uncertainty and imperfect knowledge with the goal of informing visualizations for geospatial information analysis.[8] They present a number of conceptualizations of uncertainty that could be useful for beginning to develop visualizations of uncertainty. We will not replicate their review, but will mention some of the most significant work below. The result of MacEachren *et al*'s work is an outline of challenges, including 'understanding the components of uncertainty and their relationships to domains, users, and information needs,' 'developing methods for depicting multiple kinds of uncertainty' and 'developing methods and tools for interacting with uncertainty depictions.'

Thomson *et al* propose a typology of categories of uncertainty within the domain of intelligence information analysis.[23] In the domain of decision support and policy making, Walker *et al* describe a way to convey uncertainty in a model to decision makers.[24]

In response to Walker *et al*'s work, Norton *et al*[9] described concerns about the preference for deterministic knowledge and structured ordering of levels proposed by Walker *et al*. They also argued that instead of seeing uncertainty as something additive we should view all the aspects of uncertainty associated with any decision as a whole.[9] The epistemological differences between fields add to the difficulty of creating a unified classification of uncertainty, and we share Norton's concern about the non-additive nature of different types of uncertainty. Part of the motivation for visualizing different types of uncertainty should be to help people see the multifaceted uncertainty in their data and make informed decisions.

Pang and his colleagues have worked extensively in the area of information visualization of modeling weather data. In Pang *et al*,[25] the authors define uncertainty 'to include statistical variations or spread, errors and differences, minimum–maximum range values, and noisy or missing data.' They also describe the process through which uncertainty is introduced into data. Uncertainty can be introduced at 'acquisition,' including issues with measurement or statistical variation; at 'transformation,' including any manipulation of data; or at 'visualization.'

One attempt to describe uncertainty outside any specific domain is a taxonomy created by Gershon of imperfect information.[26] The taxonomy incorporates uncertainty, but separates out other aspects that Thomson *et al*, for example, include within uncertainty. In Gershon's taxonomy, causes of imperfect knowledge include 'corrupt data/info,' 'imperfect presentation,' 'uncertainty,' 'info too complicated,' 'inconsistency,' and 'incomplete info.' This taxonomy was created earlier than most other uncertainty classifications and does not incorporate the commonalities within discussions of uncertainty from different domains.

These models of uncertainty have not been empirically evaluated and it is not obvious how to select between the models or integrate them into a single model. We were able to identify some common concepts between these models, so we started our work by describing these concepts in our own integrated classification. To empirically evaluate and improve our classification, we conducted interviews with 18 people from several domains.

Our Initial Classification

We created a classification designed to describe uncertainty for information visualization. Our initial classification is based on areas of consistency and overlaps we found between previous classifications. We discuss the five types of uncertainty we identified.

Approximation

Approximation is often necessary in science and other domains, but it leads to uncertainty. Various techniques are used to attempt to come close to measuring or describing a phenomenon when we cannot measure or describe it with perfect precision. The extent to which data are an approximation of the actual phenomena can be described in a number of ways. Sometimes this imprecision is due to expected random variations in the phenomena being measured or is due to necessary simplifying assumptions being made in order to do a study. Approximations also happen when we measure samples instead of entire populations. This is usually described by the work experimenters do to determine the generalizability of their findings.

Prediction

There are a variety of reasons for developing predictions. Predictions can be projections of future events, which may or may not happen. A common example of this is a weather prediction. We also include in prediction any explanations developed to explain something that has already happened when the true explanation is not known. Fields that are involved in model building often create models to make inferences about data; sometimes these models predict the future and sometimes they make inferences explaining what has happened in the past. These automated methods of prediction based on inference and model-building often have probabilities associated with them.

Disagreement or inconsistency

We consider disagreement or inconsistency between experts in a field to indicate uncertainty. The amount of disagreement necessary for something to be considered uncertain can be a source of contention. For example, in science, there are often situations where there are multiple plausible explanations for a phenomena and no single explanation has been proven to the satisfaction of all. This often indicates uncertainty surrounding a topic and would be helpful to distinguish from topics that are well accepted within a community of experts.

Completeness

We view missing or incorrect data as a type of uncertainty. Data sets missing large amounts of data may have far more uncertainty than those that are nearly complete. Similarly, data sets with information that is known to be erroneous should be considered incomplete. Data may contain incorrect information that goes unrecognized and is not labeled as an error; however, an error must be recognized in order to impact certainty. Errors are different than approximations. In approximation, the creator has done their best to get close to the right answer but recognizes their measurement may be inaccurate.

Credibility

In the last section we described how incomplete data lead to uncertainty. Erroneous data can also lead to concerns about the correctness of the rest of the data in a data set. We cast this concern as an issue of credibility. Credibility, as a type of uncertainty, can also arise because of the source of the data. Data that come from an unreliable source

Table 1: List of participants and their areas of work

Participant ID	*Area of work*
1	Databases
2	Computer Engineering, Project Management
3	Robotics, Probabilistic Modeling
4	Cryptography
5	Computational Biology
6	Computer Science (data analysis and study design)
7	Physiological Psychology & Software Engineering
8	Sales, Project Management
9	Human–Computer Interaction
10	Perceptual Intelligence, Perceptual Computing, Machine Learning
11	Computer Supported Cooperative Work
12	Social Psychology
13	Computer Vision, Computer Graphics
14	Intelligence and Law Enforcement, Computer Architecture
15	Journalism
16	Bioengineering, Clinical Trials, Medical Decision Support
17	Radiology
18	Zoology, Ecology

may be considered less certain than data from a trusted source. Reasons a source is not considered credible could be related to the methods used to get the data or concerns surrounding the biases or conflicts of interest with the creators of the data.

Study of Uncertainty

The goal of this study was to evaluate our classification of uncertainty and inform its refinement into a tool that can be used within information visualization to systematically develop domain-spanning uncertainty visualization techniques. To get a deeper understanding of uncertainty across domains, we conducted a formative interview-based study. We were particularly interested in gathering examples of uncertainty and learning about how people currently represent and handle uncertainty. We then used the data to improve our classification as well as gaining other insights to inform the design of ways to visualize uncertain information.

Participants

We recruited 18 knowledge workers, scientists and researchers (in the Greater Puget Sound area) from a variety of domains through emails targeted to people in specific domains and also to research email lists. Our goal was to recruit from a broad range of fields. The recruitment email explained that we wanted to interview people about uncertainty in their work, so people who contacted us had self-identified as having aspects of uncertainty in their work. Our participants came from both academic and industry settings and ranged from interns and students to established researchers and practitioners. Several participants worked in computer science with specialties including robotics, machine learning, databases, visualization, perceptual computing and computer graphics. One participant was a former radiologist and other participants were from psychology, journalism, biology, bioinformatics, intelligence, bioengineering and ecology. Table 1 lists the participants' areas of work.

Interview methods

We conducted a 30–60 mins interview with each of the 18 participants individually and took extensive field notes as well as audio recording. Some participants also provided screenshots or pointers to examples of uncertainty in their work. Most interviews were done in participants' offices, with a few exceptions for participants who suggested other locations.

Interviews followed an interview guide, but were open-ended and exploratory. We began by asking participants to tell us about the uncertainty they encounter in their work. By starting with a broad question we were able to learn about the participant's view of uncertainty without introducing our own language and thinking into their responses. As they described uncertainty we asked for specific examples, which we later used to test and improve our classification. We also asked them how they dealt with uncertainty. Towards the end of the interview we asked each person if they encountered disagreement, credibility issues or incomplete data (if those issues had not previously been covered). We also asked participants to define uncertainty, asked them how they represented uncertainty or had seen it represented and asked if they had seen any visualizations of uncertainty.

Figure 1: Our affinity diagramming in process. Yellow pages have titles or descriptions of categories and colored highlighters were used to mark other aspects of the data. Piles were refined and combined as analysis progressed.

Analysis methods

Within our team we used affinity diagramming to collaboratively analyze our data.[27] This process began with individual thoughts and examples from the interviews broken out onto pieces of paper. The aim of this study was to evaluate and improve on our classification, so we started with the categories in our original classification. As we went through the affinity diagramming pages we tried to classify the thoughts and examples into the existing categories. When we discovered examples that did not fit our scheme we placed them in a new stack or adjacent to the stack with the closest fit. When we had multiple examples that were clustered together, but did not fit into one of our existing classifications, we attempted to redefine and iterate on our classification to accommodate the new type of uncertainty. About two thirds of the way through the data we had a classification that accommodated all the examples in our data set. From that point on we were able to classify the remaining examples into the new classification without making further changes. Figure 1 shows the output of our affinity diagramming. After completing our classification using participant data we returned to the previous work to make sure that most types of uncertainty described in previous classifications for specific domains map to some part of our classification.

Results and Discussion

We present our results in the form of an improved classification with descriptions of how the interview data guided the classification. We begin with participants' definitions of uncertainty. Then we will discuss the types of uncertainty in more detail, how participants currently represent uncertainty and what they do with uncertainty. As we classified examples of uncertainty into different kinds of uncertainty, we began to see a pattern in the way uncertainty compounds or stacks in data sets. Participants were not describing just one type of uncertainty, but instead were discussing uncertainty about multiple aspects of their work and occasionally used the word 'level' to describe a higher or lower level form of uncertainty. We will describe the different levels of uncertainty using a project about robot dogs playing soccer. We believe our refined classification will help information visualization researchers in developing appropriate visualizations for each category of uncertainty. Our classification can also help them generate appropriate guidelines for each type of uncertainty.

Definition of uncertainty

Most participants had some difficulty providing a definition of uncertainty when asked. There seemed to be agreement that uncertainty often happens in situations without complete knowledge. Participants used phrases like 'imperfect knowledge,' 'inadequate information' and 'lack of absolute knowledge' to describe uncertainty. Some participants saw uncertainty as a time when the probability of something is not 1.0. When more than one event could happen, this was uncertainty. One participant articulated this as a 'partial belief' in something. Another participant phrased uncertainty in terms of hypothesis testing and model fitting. This participant claimed that we are constantly trying to fit what we see in the world into models to explain what we are seeing or predict what will happen next. When the fit is not perfect we have uncertainty. Three participants also pointed out that in research their goal was to identify and remove uncertainty; one participant described this as an attempt to 'reduce global uncertainty.'

Classification of uncertainty

As we discussed uncertainty with participants they revealed several different types. They also made distinctions between these different types as they spoke. For example, one participant (Participant 5) described uncertainty about the measurements he got from scientists and then said that on top of that there was also 'inference uncertainty.' Another distinction a few participants made explicitly was among different 'levels' of uncertainty. Participant 13 worked on computational photography and described the type of inference he used to try to remove blurring from images. He then distinguished uncertainty in the probabilistic inference from 'another level of uncertainty' caused by noise in the sensor and lens variables. The sense of multiple kinds of certainty and different levels of uncertainty in a data set or process are captured in our classification below (Figure 2).

Within the description of the categories, we will use Participant 5's uncertainty as an example and walk

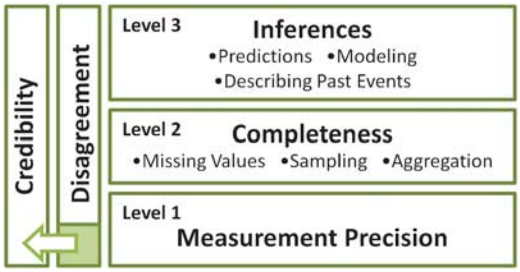

Figure 2: Improved classification. Credibility and Disagreement uncertainty span levels and sometimes disagreement results in credibility uncertainty.

through the types of uncertainty as they apply to his data. This participant worked with viral genetic data provided by scientists to create inferred phylogenies, or evolutionary trees, representing the historical evolution of the virus.

Measurement precision – Lowest level

Uncertainty due to imprecise measurements was common in our interview data and spanned domains. This category of uncertainty covers any variation, imperfection or theoretical precision limitations in measurement techniques that produce quantitative data. Sometimes this imprecision is represented explicitly by a range that the true value is probably in, for example in a confidence interval. Participant 3 used a 'bound on error,' which was a range of answers that the value was certainly in. Another example of a range representing measurement precision is with GPS location data. Participants 1 and 18 both used location data with precision represented as a distance range. However, measurement precision uncertainty is often simply a value that is known to be potentially flawed. In the example Participant 13 discussed above, there was measurement precision uncertainty from camera lens variability that was not constant enough to be modeled and adjusted for. He did not have a representation of certainty; instead, he had values known to be somewhat uncertain. In the case of Participant 5, he had uncertainty surrounding the genetic sequences he used to build the phylogenies because there is sometimes ambiguity in laboratory analysis about which nucleotide base appears at a specific point in a sequence. This uncertainty could be due to measurement error or to multiple bases appearing at that point in the sample.

Completeness – Middle level

Completeness was an issue across domains as well. Some participants described sampling as a strategy for representing the values of some population. Concerns about sampling methods and generalizing to the population from the sample lead to uncertainty in this category. Sampling could be sampling participants from a population or could be sampling rates for sensor or image data. Participant 1 worked with location data (which had measurement precision uncertainty from imprecise GPS data) and had to decide how often to sample a device's location. Location is a continuous variable; so, any sampling scheme is necessarily incomplete. Clustering and dimension reduction are two types of important data transformation to handle large data sets. Since users will lose some data details after these transformations, they also cause an uncertainty issue similar to sampling.

Missing values also represent incompleteness uncertainty but should be distinguished from sampling; missing values are intended to be included but are not present, whereas sampling implies deliberate extrapolation from a few specific measurements to cover a larger set of possible measurements. Participant 6, for example, did studies where she sampled the population but then had to throw out participants who did not finish her study. Thus, she had missing values within her sample. Both the sampling and the missing values lead to uncertainty in her results, but contributed in different ways. Participant 5 experienced uncertainty due to issues of completeness because ideally he would have a sequence for a representative sample of viruses around the world, but instead he had sequences that other scientists had collected and given him. He did not have confidence that his sample was representative of the entire population of viruses. The lack of completeness leads to uncertainty about later inferences done with the incomplete sample.

Aggregating or summarizing data in an irreversible way can also be a cause of uncertainty because once data have been summarized, information is lost and the data are no longer complete. Another example of uncertainty in our data was concern about choosing the right parameters or variables to measure or use for inference. This is related to inference uncertainty (discussed later), but according to our classification is actually an issue of completeness. The inference itself could be absolutely certain, but could produce uncertain results if an incomplete set of data is used to do the inference.

A related concept within completeness that spanned domains is the concept of unidentified unknowns. Participant 7, for example, when planning for a project, described the sense that he was probably only aware of a small fraction of the information (estimated at 10 per cent) out of all the information that might be important to consider. Participant 18 further distinguished the information you know (known knowns) from the information you know exists, but don't have (known unknowns) from the information you don't even know you are missing (unidentified unknowns). The participants who discussed this distinction agreed that the unidentified unknowns are the worst kind of missing information. These are information needs that you are unaware you have, which is worse than an information need you are aware of, but is unfulfilled. When you do not know you are missing important information you are more certain than you should be (see Figure 3). As unidentified unknowns are not identifiable, it could not be handled in our classification.

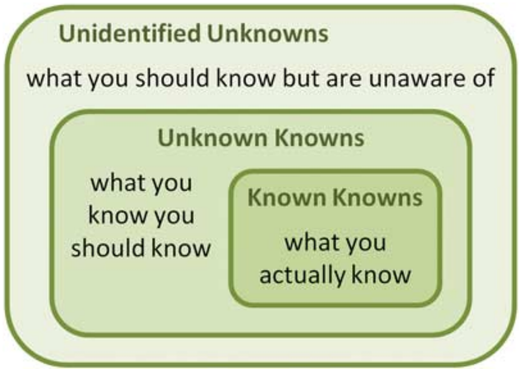

Figure 3: Level of uncertainty awareness: (1) Known Knowns: what you know; (2) Unknown Knowns: what you need to find out; and (3) Unidentified Unknowns: what you don't even know you need to find out. Participants described the unidentified unknowns, or unrecognized information needs, as the worst case.

Inference – Highest level

Inference is a fairly broad category, spanning all types of modeling, prediction and extrapolation. Inference has a tight relationship with decision-making: it is how data are infused with meaning and transformed into decisions.

Modeling of any kind, ranging from probabilistic modeling to hypothesis-testing, to diagnosis, falls in this category. For example, Participant 16 described the need to take a set of medical symptoms, either as a care provider or health consumer, and fit them into a model of illness. If someone is experiencing pain in their shoulder, they may pair that with information about moving furniture and fit that pain into their model of a tired muscle, instead of into their model of a heart attack. The degree to which the model is an appropriate causal abstraction of the data's properties determines the extent to which uncertainty is introduced by the model. For example, a health consumer may inappropriately rule out a heart attack because their model of the kind of person who has a heart attack is not realistic.

Prediction involves inferring future events by creating an abstraction of the causal relationship between current or past data and future occurrences – again, the correctness or incorrectness of the abstraction is a source of uncertainty. Participants described using past experiences to predict timelines and outcomes. They also described formal modeling of data to make predictions. For example, Participant 9 described watching the trend of exchange rates to predict their future values.

Extrapolation into the past, a complement to prediction, involves using data to try to recreate or make inferences about past events. For example, Participant 1 was interested in locations and paths of devices and people. He could use path data (inferred from location data) to try to identify where someone was in the past. Participant 10 described a project where he was interested in identifying patterns in past network traffic from malicious programs and using those models to predict and detect future malicious activities. Participant 5 extrapolated to past events by using known virus genetic sequences to try to reconstruct how a virus has changed over the years. This inference is uncertain because no one knows for sure how viruses change, which is illustrated by the existence of several different models of evolution.

Credibility – Spans levels

Credibility is a type of uncertainty that spans the three levels and is not easily measured. An information source that produces data that conflict with other data, has produced unreliable data in the past, or is otherwise suspect for some reason leads to uncertainty. Individuals may have different judgments about what constitutes a credible source. A human source may be considered untrustworthy based on past behavior or associations. Participant 17 described his experiences as a radiologist with uncertainty due to credibility. Much of his information came from other people, either patients or clinicians. When evaluating the information, he often considered the source and their level of expertise. Occasionally, he was concerned about intentional deceit from patients, but usually he was assessing whether the person was qualified to provide the information they had given him. For example, information from a specialist may lead to less uncertainty than information from a generalist because he trusts the specialist's knowledge and expertise. This participant described that over time he got to know other doctors and understood their strengths and weaknesses, which he then factored into his feelings of certainty or uncertainty about a case. Similarly, Participant 18, an ecologist, discussed building relationships with people and organizations over time and assigning different levels of credibility based on their level of expertise and on his experiences with them.

Non-human sources can also give rise to credibility uncertainty. Participant 18 ascribed levels of credibility or trustworthiness to different sources of information. In his work they tried to rank types of sources (for example, journal or book), but he described this as problematic and leading to uncertainty because different sources are more credible for different types of information. Machines or measurement tools can also be considered untrustworthy based on past behavior. In this case the credibility is superficially similar to measurement precision uncertainty. However, the distinction is that credibility is a judgment made by the human consumer of the information about the information source, rather than being a known precision limitation mathematically expressible by the information source itself.

Disagreement – Spans levels

Disagreement leads to uncertainty and spans the three levels. At the measurement precision level, disagreement happens when the same thing is measured multiple times

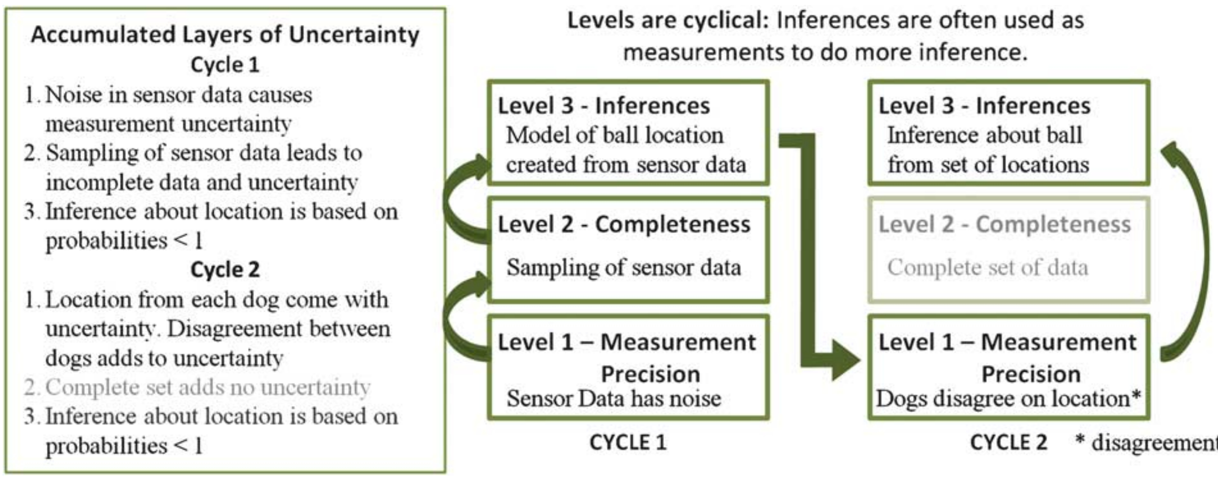

Figure 4: Example of Cyclic Uncertainty: when dogs disagree about the location of a soccer ball. By the time a ball location is predicted we have several layers of uncertainty below this one assertion.

or by different sources and the measurements are not the same. At the completeness level, disagreement comes from overlapping but not identical data sets. At the inference level, disagreement comes from two (or more) different conclusions being drawn from the same data. This could be two (or more) experts looking at a data set and coming to different conclusions, or it could be applying two different mathematical models to a data set to do inference. Participant 5 related an instance of disagreement at the inference level. Part of his work involved using multiple mathematical models of evolutions to predict the phylogeny of a virus. Each model produced a slightly different phylogeny and thus disagreement. The ecologist, Participant 18, also discussed substantial uncertainty due to disagreement. He gathers data from many sources to evaluate the number of species in different regions of the world. He uses databases, publications and personal contacts to fill in missing information and often encounters disagreement among the sources. One of his projects involved combining a map of watersheds with a map of dams in a region and the two did not match up. He described this as 'location uncertainty' and had to try to reconcile the two disagreeing sources of information. A slightly different example of disagreement in his work relates to counting the number of species in regions. He receives disagreeing information from different sources for this project too, but also experiences uncertainty due to disagreement with surrounding regions. If an area is reported to have 10 species, but regions around it are all reported to have 300 species, this disagreement leads him to be uncertain about the count of 10 species.

Disagreement and credibility are often associated because as soon as disagreement occurs, whether among people or among measurements, credibility is often called into question. In the case of the disagreement in number of species, the ecologist may check the sources of the numbers and decide if credibility is an issue.

Levels of uncertainty

As mentioned before, several participants described uncertainty about multiple aspects of their work and occasionally used the word 'level' to describe a higher or lower level form of uncertainty. After exploring this concept in the data, we assigned Measurement Precision to the lowest level type of uncertainty, Completeness to the middle level, and Inference to the highest level (Figure 2). Credibility and Disagreement are types of uncertainty that occurred along with, or on top of, each of the other types of uncertainty so they span the three levels. This does not mean that every data set or project will involve every level of uncertainty; but many projects involved more than one level of uncertainty. One reason levels of uncertainty are so crucial and problematic in our participants' experiences is that uncertainty within one level, even if well quantified at that level, rarely can be adequately transformed or accounted for at another level when the decision-making process requires a transition between levels.

We will use a project Participant 3 described about robot dogs playing soccer to illustrate the different levels of uncertainty and how they interact (Figure 4). When robot dogs play soccer, one of the problems is identifying the location of the soccer ball. Each dog has multiple sensors to figure out where the ball is. At the level of the individual dog (cycle 1 in Figure 4), there is some measurement uncertainty due to noise in the sensor data. On top of this measurement uncertainty, there is also uncertainty due to the incompleteness of the data because, although it is theoretically continuous, it is actually sampled. There is also uncertainty at the inference level when the incomplete sample of uncertain measurements is used to infer ball location because the model by which sensor data is translated to ball location is known to be imperfect. With just one dog there is no disagreement-based uncertainty (although there could be if, for example, duplicate sensors

Table 2: Examples of uncertainty from three different domains

Domain	*Ecology*	*Computational Biology*	*Medicine*
Goal	Identifying species diversity in different regions of the world	Modelling viral evolution	Diagnosis of disease
Data types	List (and numbers) of species and maps	Evolutionary tree and RNA sequences	Variety of formats such as written/verbal descriptions, numerical values and images
Inference	Predicting future populations based on past events or current states is imperfect	Models of evolution are imperfect so inference about past viral sequences based on evolutionary models is uncertain	Signs and Symptoms are used to infer diagnosis when often there is not a clear line between one set of signs and symptoms and one diagnosis
Completeness	Lack of data for some parts of the world leaves an incomplete data set	Do not have sequences for all strains of the virus, which would be required to have a complete data set	Decisions must often be made with incomplete medical records or without running all possible tests
Measurement precision	Identification of a species within a specific geographic region could be wrong because location data is imprecise	Errors can occur in lab processes for identifying sequences	Many diagnostic procedures (for example laboratory tests) are not perfectly precise
Credibility	Data gathered from some sources may be trusted over others based on source's qualifications	Viral sequences from one data source might be trusted over another	Some expert opinions (for example radiology) may be trusted over others
Disagreement	Different sources may disagree about the presence or absence of a species in a region	Different evolutionary models render different results	Two different tests might contradict one another or two doctors might disagree on a diagnosis

gave different readings). At the end of this cycle through the three levels we still do not know the location of the soccer ball; we just know where one dog thinks it is.

When we start compiling the opinions of multiple dogs we start the cycle all over, treating each dog's ball location as a measurement in level one and the set of all the dogs' ball locations as a complete set in level two. The only uncertainty in these two levels comes from disagreement in the measurements (previously considered individual inferences) from each dog. Once the set of locations from multiple dogs is compiled, inference is done to identify one probable location of the ball. This model of location uses probabilities and does not generate one definitive location, but instead generates a likely location. With just the assertion of where the ball is at any given time, we have several layers of uncertainty. Inferences are done on top of other inferences, which are done on an incomplete set of data, which is made up of imprecise data. While this example may seem riddled with uncertainty, this is fairly typical of our participants' data.

Another example where we can see multiple cycles through the levels of uncertainty comes from the radiologist, Participant 17. He receives a film, which may have measurement imprecision due to flaws in the imaging and it may or may not be a complete enough image for him to do a thorough analysis. He then attempts to fit what he sees in the imaging to models of disease through an admittedly incomplete inference process. Once he is finished, he delivers an interpretation, or a diagnosis, to the doctor who ordered the report. That doctor then goes though a second cycle where the radiology report, lab work and observed symptoms may all be considered measurements, from information sources with various degrees of credibility, which together make up a set of potentially incomplete information about a patient. The doctor then goes through his or her own diagnosis, or inference, process based on this set of information. Further examples of types of uncertainty from multiple domains are described in Table 2.

Dealing with uncertainty

The degree to which uncertainty from imprecise measurement from a soccer dog's sensor or a grainy x-ray impacts the eventual outcome is extremely hard to quantify. Participants described several strategies for dealing with uncertainty, but the predominant feeling seemed to be that the uncertainty was complex and difficult to describe, let alone deal with. Part of the problem may be that there is no way to clearly carry uncertainty through multiple steps in a project, to transform measurable uncertainty at one level into meaningful information at another level, or even to convey that uncertainty to others.

At a basic level, participants chose to do one of two things: live with the uncertainty or try to become more certain. Participants made this decision based on the potential impact of being wrong and based on how successful they felt they would be in improving their certainty. When participants decided to try to improve their certainty, they often gathered more data, either from the same source or a different type of data that informed the problem. Participant 10 described spending quite a bit of time selecting good parameters, or types of evidence, for identifying activities using machine learning. Participant 6 said that in his work, finding multiple sources of information indicating the same result improved his certainty. When participants chose to tolerate their uncertainty they often mentioned it was important to be aware of it and not use the data inappropriately. The ecologist, Participant 18, described avoiding making close calls with his data because he knows it is imprecise. Instead, he only makes distinctions between regions with vastly different numbers of species and avoids making distinctions about regions with relatively similar species counts. Some participants were concerned with conveying their uncertainty to others or described a shared (but often implicit) understanding of uncertainty within a group or within a field.

Representations of uncertainty

There has been a substantial amount of research on the difficulty of mapping qualitative descriptions of uncertainty to numerical probabilities.[7] At the end of each interview, we asked participants how they convey uncertainty and how they represent uncertainty. As previous work has indicated, one of the challenges for visualizing uncertainty is that it was often not expressed in a standard quantification.

Formats of uncertainty

Some participants had quantifications of uncertainty they routinely used. In computer science, participants tended to define uncertainty in terms of probabilities. They saw a probability as representing a belief that something is true. The other quantification of uncertainty we saw was some form of range. Sometimes this was a value plus or minus some amount, as in the case of a confidence interval. Other times this was simply a range where the true value was believed to fall. These quantitative expressions of uncertainty often occurred with measurement precision uncertainty or inference uncertainty.

Many participants had uncertainty that they did not quantify. Instead they used looser, more qualitative labels. Participant 8 described it in terms of t-shirt size: 'small, medium, large, and extra large.' These were not standardized definitions, but were constructs created and used within teams or groups. Participants also used words such as 'likely' and 'probably' to convey their own belief in an assertion or value. These are comparable to a 'subjective probability' or 'personal probability' judgment commonly used in intelligence analysis.[28] This often happened in discussions, in email, or in other writing, but was rarely stored with the data.

Visualization of uncertainty

By far the most commonly mentioned visualization was error bars. Participants explained that they used these to show ranges of all kinds and usually put them on bar graphs. Some participants expanded the idea of an error bar to apply to location as well, describing a point with a circle around it. One participant even described a sphere visualization with a buffer zone (or error bar) encompassing the sphere. Other visualizations of uncertainty included showing distributions with box plots and using data plots with quartiles. Participant 5, who dealt with evolutionary trees, mentioned tree alignment, described color coding branches and adding icons (often asterisks) to branches to indicate certainty. A participant also mentioned that humans often visualize uncertainty with body language and facial expressions. Several participants expressed frustration with the difficulty of communicating certainty and were intrigued by our research in visualizing it.

Mapping Previous Work to Our Classification

Our classification is based on data we gathered from a limited number of participants. Our sample and methods do not provide an exhaustive data set and so we cannot claim absolute generalizability. Our classification could be further improved by trying to fit it to data in domains we were not able to include and to more examples of uncertainty in data. In this section, we further the evaluation of our classification by identifying any aspects of other classifications not encompassed in our own and by examining how existing visualizations of uncertainty can be described by our classification.

Most types of uncertainty described in classifications for specific domains map to some part of our classification. Our category of measurement precision has similarities to Thomson *et al*'s 'Accuracy/errors' and to 'Precision,'[23] although we view known errors to be missing data. Gershon[26] also has a category for 'corrupt data' and this maps to our measurement precision category in some conditions. If the data are suspected to be corrupt we consider it a measurement precision issue with a possible credibility complication, but if it is known to be corrupt we could describe that as invalid, and therefore missing, data. Walker *et al*[24] describe a level of certainty called 'Statistical Uncertainty' and this also maps to our concept of measurement precision. There is a more straightforward mapping between Thomson *et al*'s 'Completeness'[23] and Gershon's 'Incomplete'[26] and our category for completeness.

The focus of our empirical investigation has been describing uncertainty the way researchers and scientists

Table 3: Mappings of existing visualizations to our classification

Visualization	*Classification*
Vis by Pang and Freeman	Disagreement uncertainty (inference level)
CandidTree	Disagreement uncertainty (inference level)
LISTEN	Disagreement uncertainty (inference level)
DaVis	Disagreement uncertainty (measurement precision level) and Completeness uncertainty
MANET	Completeness uncertainty
Restorer	Completeness uncertainty
Vis by Djurcilov and Pang	Completeness uncertainty
Vis by Collins and Penn	Inference uncertainty
Error bars by Olston and Mackinlay	Inference uncertainty
Ambiguation by Olston and Mackinlay	Measurement precision uncertainty

encounter it in their data sets. Pang *et al* take a process-oriented view of where uncertainty can be introduced in the information visualizations pipeline.[25] At the 'Acquisition' phase measurement precision becomes an issue. The step they call 'Transformation' has some mappings to our concept of completeness, since sampling and interpolation are included in their description of the transformation process. Our investigation was of how researchers and scientists describe the uncertainty in their data, so we have not focused on the uncertainty that can be introduced at the final 'Visualization' step of the pipeline. Pang *et al* describe how rendering, animation and other forms of manipulation for visualization can introduce further uncertainty. There is not a directly equivalent category for inference in the existing classifications, but Walker *et al*'s description of 'recognized ignorance' [24] is somewhat related. Interestingly, their discussion of 'recognized ignorance' versus 'total ignorance' came up with our participants in the context of recognized information needs (known unknowns) versus unrecognized information needs (unidentified unknowns), whereas Walker *et al* describe these two states in terms of inference limitations. The problematic nature of unrecognized information needs has also been described in other fields.[24]

Our concept of credibility also appears as a form of disagreement in Thomson *et al*,[23] where it is called 'Consistency,' and includes inconsistent or conflicting information from a single source or from multiple sources. Gershon also includes 'inconsistency' as a source of imperfect knowledge.[26] Thomson *et al* have a category for 'credibility' and also have categories for 'lineage' and 'currency.' We include information like lineage, and potentially currency of information, within the credibility category because they lead to uncertainty about the credibility of the data. Walker *et al* also discuss a similar concept and call it 'pedigree'.[24] In our classification some of the lineage is also captured in our description of layers of uncertainty. The purpose of the layers is to capture the accumulation of uncertainty and thereby provide information for judging the certainty of the end product.

Table 3 shows how uncertainties handled by the existing visualizations described in the Previous Work section above can be categorized into our classification. Although Table 3 may not cover all the currently available uncertainty visualizations, we notice that visualizations for disagreement uncertainty mainly from the inferences level and completeness uncertainty mainly for missing values have been investigated the most. Furthermore, visualizations for credibility uncertainty have not been investigated yet. We suspect it is because, compared to the disagreement uncertainty and completeness uncertainty, credibility uncertainty is much more difficult to quantify. For example, disagreement uncertainty is often computed by the differences between two results, which can be represented as numbers (for example, percentage). Completeness uncertainty can also be easily represented as boolean values or numbers. We may need to develop ways to express credibility uncertainty as categories or numbers that are meaningful to end users before we try to design new visualizations for credibility uncertainty. The 'subjective probability' or 'personal probability' judgment commonly used in intelligence analysis[28] can be used as a starting point. We then can either apply existing visualizations or develop new visualizations to reveal uncertainty about the credibility of data.

Future Work

Our motivation for categorizing uncertainty across domains was to create a way to discuss and understand uncertainty so that we can find good ways to display it to a wide range of users. The next step is to identify ways to visualize different types of uncertainty. One of the most promising aspects of our classification for uncertainty is the concept of 'layers' of uncertainty, which were discussed by participants in multiple domains. Layers add complexity to data that is not simple to conceptualize or convey with current techniques. This creates a wide-open opportunity for visualization.

Visualizing uncertainty is a relatively young research area and most of the proposed visualizations have not been empirically evaluated. Part of the reason may be that the appropriate approach to evaluation of a visualization of uncertainty is not clear. Zuk and Carpendale[29] did a heuristic evaluation of eight visualizations of uncertainty using principles or heuristics identified for information visualization in general. After principled evaluations, we need to progress to studies of actual use. Both the importance and difficulty of studying actual use arises in other areas of human–computer interaction and information visualization, and has been a popular topic of discussion.[30] The goal of this type of visualization is to provide users with a better understanding of the data by conveying information about the certainty of data. A lab study may be able to evaluate how well a sense of certainty is conveyed, but more relevant and rich data would probably be discovered by a study of real-world situations when people use data that matters to them. This leads us to propose *in situ* longitudinal evaluations of a visualization of uncertainty.

Conclusion

In this paper, we proposed a classification of uncertainty that spans domains. We first created a classification by reviewing literature on uncertainty from several domains on uncertainty, and then empirically evaluated and improved upon our classification by conducting interviews with 18 people working with uncertainty. Participants in our study were clearly aware of uncertainty at many levels in their data and expressed discomfort at their inability to be transparent about showing their uncertainty. Our classification better describes the broad range of uncertainty across domains, provides a structure for visualizing uncertainty and will help us in developing visualizations to make uncertainty more readily understood.

Acknowledgements

The authors thank our participants, who made this research possible. We also thank John Stasko for his thoughtful comments and the reviewers for the highly constructive and helpful comments.

References

1 Thomas, J.J. and Cook, K.A. (2005) *Illuminating the Path*. Los Alamitos, CA: IEEE Computer Society.
2 Johnson, C.R. and Sanderson, A.R. (2003) Visualization viewpoints: A next step: Visualizing errors and uncertainty. *IEEE Computer Graphics and Applications* 23(5): 6–10.
3 Klir, G.J. (2005) *Uncertainty and Information: Foundations of Generalized Information Theory*. New York: Wiley-IEEE Press.
4 Bernardo, J.M. and Smith, A.F.M. (2000) *Bayesian Theory*. New York: Wiley.
5 Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Los Altos, CA: Morgan Kaufmann.
6 Klir, G.J., Clair, U. St. and Yuan, B. (1997) *Fuzzy Set Theory: Foundations and Applications*. Englewood Cliffs, NJ: Prentice-Hall.
7 Renooij, S. and Witteman, C. (1999) Talking probabilities: Communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning* 22: 169–194.
8 MacEachren, A.M. *et al* (2005) Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science* 32(2): 139–160.
9 Norton, J.P., Brown, J.D. and Mysiak, J. (2006) To what extent, and how, might uncertainty be defined? Comments engendered by 'Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support' (Walker, 2003) *Integrated Assessment* 6(1): 83–88.
10 Lodha, S.K., Pang, A., Sheehan, R.E. and Wittenbrink, C.M. (1996) *UFLOW: Visualizing Uncertainty in Fluid Flow*. IEEE Symposium on Visualization 1996. Chicago, IL: IEEE Computer Society Press, pp. 249–255.
11 Wittenbrink, C.M, Pang, A.T. and Lodha, S.K. (1996) Glyphs for visualizing uncertainty in vector fields. *IEEE Transactions on Visualization and Computer Graphics* 2(3): 266–279.
12 Grigoryan, G. and Rheingans, P. (2004) Point-based probabilistic surfaces to show surface uncertainty. *IEEE Transactions on Visualization and Computer Graphics* 10(5): 564–573.
13 Gershon, N.D. (1992) Visualization of fuzzy data using generalized animation. IEEE Symposium on Visualization 1992. Chicago: IEEE Computer Society Press, pp. 268–273.
14 Lodha, S.K., Wilson, C.M. and Sheehan, R.E. (1996) LISTEN: Sounding uncertainty visualization. Proceedings of the Visualization 1996. Los Alamitos, CA: IEEE Computer Society, pp. 189–195.
15 Pang, A. and Freeman, A. (1996) Methods for comparing 3D surface attributes. Proceedings of SPIE-VDA 1996. SPIE/IS&T, San Jose, CA. pp. 58–64.
16 Lee, B., Robertson, G.G., Czerwinski, M. and Parr, C.S. (2007) CandidTree: Visualizing structural uncertainty in similar hierarchies. *Information Visualization* 6: 233–246.
17 Sulo, R., Eick, S. and Grossman, R. (2005) DaVis: A tool for visualizing data quality. Posters Compendium of InfoVis 2005. Los Alamitos, CA: IEEE Computer Society, pp. 45–46.
18 Unwin, A., Hawkins, G., Hofmann, H. and Siegl, B. (1996) Interactive graphics for data sets with missing values – MANET. *Journal of Computational and Graphical Statistics* 5(2): 113–122.
19 Twiddy, R., Cavallo, J. and Shiri, S.M. (1994) Restorer: A visualization technique for handling missing data. Proceedings of the Visualization 1994. Los Alamitos, CA: IEEE Compute Society, pp. 212–216.
20 Djurcilov, S. and Pang, A. (1999) Visualizing gridded datasets with large number of missing values. Proceedings of the Visualization 1999. Los Alamitos, CA: IEEE Computer Society, pp. 405–408.
21 Collins, C. and Penn, G. (2006) Leveraging uncertainty visualization to enhance multilingual chat. Proceedings of the CSCW 2006, Demonstrations. New York: ACM Press.
22 Olston, C. and Mackinlay, J.D. (2002) Visualizing data with bounded uncertainty. Proceedings of InfoVis 2002. Los Alamitos, CA: IEEE Computer Society, pp. 37–40.
23 Thomson, J., Hetzler, B., MacEachren, A., Gahegan, M. and Pavel, M. (2005) *A Typology for Visualizing Uncertainty*. Proceedings of the SPIE-VDA 2005: SPIE/IS&T, San Jose, CA.
24 Walker, W.E. *et al* (2003) Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment* 4(1): 5–17.
25 Pang, A.T., Wittenbrink, C.M. and Lodha, S.K. (1997) Approaches to uncertainty visualization. *The Visual Computer* 13: 370–390.
26 Gershon, N.D. (1998) Visualization of an imperfect world. *IEEE Computer Graphics and Applications* 18(4): 43–45.
27 Beyer, H. and Holtzblatt, K. (1998) *Contextual Design: Defining Customer-Centered Systems*. San Francisco, CA: Morgan Kaufmann Publishers.
28 Heuer, Jr. R.J., (1999) *Psychology of Intelligence Analysis*. Pittsburg, PA: Government Printing Office.
29 Zuk, T. and Carpendale, S. (2006) Theoretical analysis of uncertainty visualizations. Proceedings of the SPIE-VDA 2006. SPIE/IS&T, San Jose, CA.
30 Shneiderman, B. and Plaisant, C. (2006) Strategies for Evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. Proceedings of the BELIV 2006. ACM Press, pp. 1–7.