THE UNIVERSITY OF CALGARY

# Visualizing Uncertainty

by

Torre Dana Zuk

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

CALGARY, ALBERTA

April, 2008

# THE UNIVERSITY OF CALGARY

# FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a dissertation entitled "Visualizing Uncertainty" submitted by Torre Dana Zuk in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY.

Supervisor, Dr. Sheelagh Carpendale
Department of Computer Science

Dr. William Ghali
Department of Medicine

Dr. Brian Wyvill
Department of Computer Science, University of Victoria

Dr. R. Daniel Jacobson
Department of Geography

External Examiner, Dr. Alan M. MacEachren
Department of Geography, Pennsylvania State University

April 29, 2008

Date

# Abstract

Uncertainty is a normal part of everyday life. It appears in the environment around us from the weather to the stock market, internally to some degree in almost every plan or decision we make, and is inherent in our daily communication, both verbal and visual. The form this uncertainty takes is often qualitative or unquantified and so fits poorly with the initial issues of representation, computability, and efficiency often the driving forces in initial visualizations of information. Understanding what may assist in visualizing uncertainty is the subject of this research.

Initially I provide a literature review of existing work in uncertainty visualization. This review continues with an exploration of heuristic evaluation specifically on uncertainty visualization but then looks deeper at the process of heuristic evaluation itself. Moving toward user constraints and cognitive tasks I coalesce existing work relating to reasoning under uncertainty. From this I propose further linking and integrating the uncertainty visualizations into the process of reasoning which encompasses all visualization tasks.

The second half of the dissertation turns to investigate uncertainty visualization in specific domains. In the first domain, results of research into visualizing temporal uncertainty in archaeological reconstructions are provided. This is followed by visualizations developed for uncertainty in rock property modelling in the seismic domain. The final domain of evidence-based medical diagnosis is explored with an observational study, participatory design of new visual support, and a final evaluation.

Finally I present a framework for assisting with the development of visualizations dealing with uncertainty by breaking out several important factors and cognitive tasks to consider based on generalizing and applying the practical and theoretical developments. In summary my contributions include specific visualizations for particular application domains along with more general aspects relating to evaluation, applicability of cognitive theory, and a framework to aid uncertainty visualization.

# Acknowledgements

This dissertation is the result of the work and support of many people. First and foremost I would like to thank all of my family for all their patience and support. I deeply thank my supervisor Sheelagh Carpendale for all the direction, time, and effort; all my domain collaborators for their contributions – Bill Glanzman for the archaeological discussions over beer, Bill Ghali and Barry Baylis and the the rest of the Ward of the $21^{st}$ Century research group for their insights and ideas, and Jon Downton and Dave Gray for their anisotropic discussions; and all my other co-authors for their contribution: Petra Isenberg, Lothar Schlesier, Mark Hancock, and JD Liang. I would also like to thank the last of my supervisory committee not already mentioned, Brian Wyvill, the others in my defense committee, Alan MacEachren and Dan Jacobson, and the rest of my candidacy committee, Mario Costa Sousa, for all their time and constructive feedback. Countless additional thanks to all of the ILab members and administrators past and present, and my other friends at the UofC, for their creative inspiration and sense of community. I would also like to gratefully thank the department administration and support teams for all their assistance, as well as NSERC, CGGVeritas, and the Department of Computer Science for funding this work.

*To Gretchen, Xander, and Ronan.*

# Contents

# List of Tables

# List of Figures

As far as the laws of mathematics refer to reality, they are not certain;
as far as they are certain, they do not refer to reality.

– Albert Einstein. (1879 – 1955)

# Chapter 1

# Introduction

*Solum certum nihil esse certi ...*
– Pliny the Elder. *Historia Naturalis* (23 CE – 79 CE)

Only in theory can one be certain.
– Translated by T. Zuk. *Ph.D. Dissertation* (2008)

In this chapter the problem domain is introduced in very practical terms and I point out some of the great variety it encompasses. I will then define the restricted scope of my research and present my goals followed by a high-level summary of the methodologies used to pursue them. The chapter ends with a brief overview of the entire structure of the dissertation.

## 1.1 Problem: Uncertainty Visualization

Uncertainty is not isolated to statistical numerical processes but is a normal part of everyday life. It appears in the environment around us from the weather to the stock market, cognitively to some degree in almost every plan or decision we make, and is inherent in our daily communication, both verbal and visual. The form this uncertainty takes is often qualitative or unquantified and so fits poorly with the initial issues of representation, computability, and efficiency, which are often the first driving forces in visualizations of information. This may be the reason why it has not received much direct attention until relatively recently.

To frame this research I will initially define uncertainty in reference to this dissertation. Then it will be described in terms of data, which has been the standard basis for investigation in the field of information visualization. Expanding into a more general and larger scope uncertainty will be related to communication in general, and then what is often the result of communication: decisions and actions. To end on a positive note, some beneficial aspects of uncertainty will be described.

### 1.1.1 Defining the Undefinable

Based on the goal of including most types and sources of uncertainty in data, Pang et al. [1997] defined uncertainty to include statistical variations or spread, errors and differences, minimum-maximum range values, and noisy or missing data. This parallels their use of the National Institute of Standards and Technology (NIST) report's four ways of expressing uncertainty: statistical, error, range, and scientific judgment [Taylor and Kuyatt, 1994]. The method of scientific judgment is not considered in their discussion of visualization but is the only one that directly includes user considerations. This should be an important type to directly consider, as the end result of interpreting a visualization will often be judgments. Therefore, I define uncertainty more broadly and include cognitive uncertainty of the user. As can be seen from the quotation of Pliny the Elder, with this broader definition we run the risk of including everything but it is important to consider the power of uncertainty in its ubiquity and the increased freedom from constraints.

### 1.1.2 Uncertain Data

Data or information with additional uncertainty attributes, may counter-intuitively be considered superior in quality to raw data. Almost all data outside of the theoretical realm has some associated uncertainty, and so presenting data without this uncertainty usually means something is hidden. The viewer is then left only with the option to hope that it is insignificant uncertainty.

Without uncertainty information, data is missing some characteristic properties which capture aspects of how it was acquired, processed, or encoded. These aspects may be essential for judging the validity of data before accepting it or even incorporating it into ones knowledge. Visualizations have been created to reveal both the data and its uncertainty for many data types and tasks (e.g. [Botchen et al., 2005, Grigoryan and Rheingans, 2004]). These uncertainty visualizations are designed with the goal of generating the appropriate confidence in the data, and user confidence itself may be considered uncertainty data to be visualized.

### 1.1.3 Lost in Communication

During one collaborative discussion with an archaeologist, conversation lead to my statement that I had read the book The Nibelungenlied [anonymous, translation of c. $13^{th}$ Cent. CE text]. His comment was "... in the original Old German"? To which I replied, no, that it was a translation. "Then you haven't read the Nibelungenlied", he stated mostly in jest. Uncertainty usually exists in what is lost both in translation and communication[1].

Translation is an excellent exemplar of all the issues of uncertainty in communication confounded by uncertainty in re-representation. If we consider my translation in the epigraph I provide what could be considered a valid translation of Pliny the Elder's statement, albeit my translation is more ambiguous and may be misinterpreted as a statement of the supremacy of theory over practice.

Visualizing the particular aspect of the uncertainty in statistical lattices used in automated translation has been presented by Collins et al. [2007]. In one of their visualizations, shown in Figure 1.1, multiple variations in the possible translation are revealed with encodings revealing their statistical weight. However the general process of literary translation is complex and encompasses issues of potentially preserving the visually evoked images, allusion, rhyme, rhythm, pun, alliteration, onomatopoeia, and of course the meaning (as it is understood by the translator). Paul Wilson noted in translating the passage taken from a discussion among Czechoslovakian factory workers, "... the moon is really no bigger than

[1] For example the allusion to the movie "Lost in Translation" in this section title would itself likely be lost in translation, just as it will become lost with time

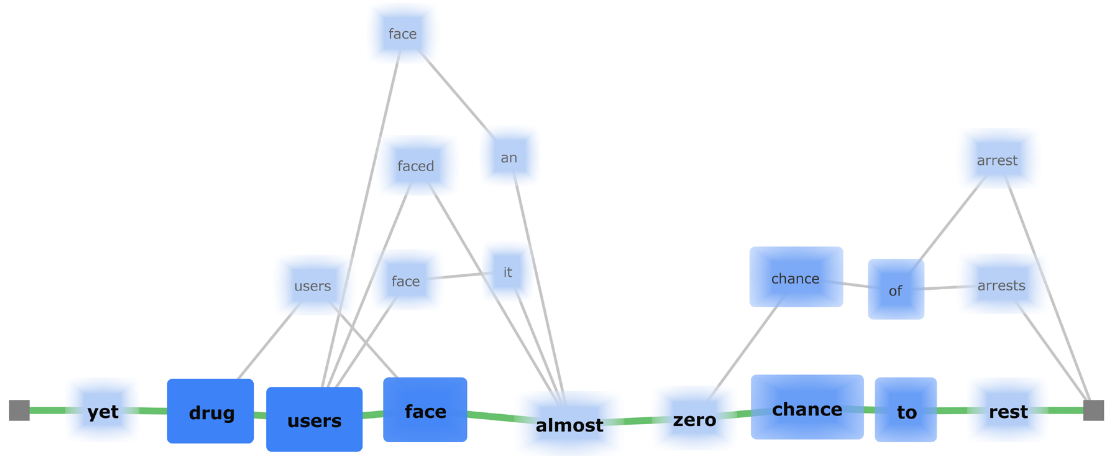



Figure 1.1: Visualization of uncertainty in statistical lattices used for translation [Collins et al., 2007]. The most probable translation determined by the algorithm is found along bottom (green linked) path. [©2007 C. Collins]

a loaf of bread," it is essential that the cultural context that bread in the area was shaped like a ball be known, and so required the insertion into the translated text of what is termed an invisible footnote [CBC Radio, 2007].

Visual communication is a fundamental part of this work and thus visualizations may be confounded by uncertainty in both the sender and receiver. This uncertainty enters in the general processes of encoding and decoding and is in addition to the data uncertainty the visualizations wish to portray [Saussure, 1965, Shannon and Weaver, 1949]. In the end, visualizations are simply communication between people. Thus, when creating uncertainty visualizations, it may be a useful abstraction to think about just two people trying to communicate with each other.

### 1.1.4 Uncertainty in Action

Traditionally uncertainty has had closer ties to inaction than action, as it is normal that one tries to reduce uncertainty so one may get deterministic results. An important example of uncertainty leading to inaction is climate change. Data and models, with the earliest dating all the way back to 1896, have shown potential warming consequences for failing to address a build-up of carbon dioxide in the atmosphere [Weaver, 2003], but up until recently policy makers and the public have been able to discount the value of action based upon perceived uncertainty[2]. Whether the primary uncertainty and thus inaction has had more to do with presentation than with the data itself is worth considering [Gore, 2006], as presentation takes on a stronger role with less understood processes. This relates at a semiotic level to the strength of codes, which are independent of scientific strength [Eco, 1986].

In contrast, a recent advisory from the University of Calgary Department of Risk Management is an example of action motivated by uncertainty:

> Please note that Iran has been raised to an Extreme Rating with respect to international travel on University of Calgary business. This is due to recent ambiguous comments by the Iranian President regarding academic faculty members in that country. As such, all travel on University business to Iran has been suspended until the situation has been appropriately clarified and/or resolved.

This advisory is itself filled with uncertainty. Neither the cause "ambiguous comments" nor the conditions "appropriately clarified and/or resolved" are clearly defined. The use of "Extreme Rating" is also a vague category, and without further clarification it lacks grounding in specific risks.

Similar uncertainty in communication exists with the U.S. Homeland Security Advisory System's Color-coded Threat Level System which has five levels: severe, high, elevated, guarded, and low risk of terrorist attack. The levels are not clearly mapped to expected public responses, as the current level of elevated has the guidance "All Amer-

[2]Many policy makers including those in Canada, United States, and Australia continue to discount the value of action based on uncertainty in current economic costs versus those of the future (which they do not have to face).

icans should continue to be vigilant, take notice of their surroundings, and report suspicious items or activities to local authorities immediately" which one would expect to be applicable at all threat levels. This mapping of uncertainty to decisions and actions is an interesting area that requires further examination, as a visualization should consider in its design how any resulting responses or actions relate to a user's tasks.

### 1.1.5 Benefits of Uncertainty

With uncertainty negative connotations such as stress resulting from fear of the unknown are often thought of first, but it also has its positive side. Intentionally added ambiguity in video communication (e.g. Gaussian blur filtering for telecommuting as shown in Figure 1.2) may provide benefits in the form of privacy [Boyle and Greenberg, 2005]. Uncertainty is also often part of a normal encoding process, as we reduce or compress data down to work more efficiently with it. If something only requires a yes or no encoding, there may be some uncertainty as to the level of agreement, but this is much easier to work with and communicate than something such as a rating out of 100.

Figure 1.2: Privacy via uncertainty for telecommuters. Right image shows a trade-off between awareness and privacy via Gaussian blur filtering [Boyle and Greenberg, 2005].

Thus there may be a trade-off between uncertainty (often in the form of precision) and efficiency for both communication and calculation. At the sub-atomic level Heisenberg's uncertainty principle tells us that beyond a specific point we can not gain certainty in position without gaining uncertainty in momentum, and vice versa. It is therefore something fundamental that one must trade in uncertainties. Mathematically spatial location certainty must be traded for spatial-frequency certainty (an exact spatial frequency must have infinite domain). Daugman [1985] has even shown this trade-off exists for the human perception of position, orientation, and size.

Uncertainty may also be beneficial to collaborative design and creative processes. It has been found that rough sketches of architectural designs (via non-photorealistic rendering) may promote more discussions and active participation than shaded rendering or traditional computer-aided architectural design (CAAD) plots [Strothotte and Schlechtweg, 2002]. Apparently showing ambiguity or uncertainty in details opens the door to alternative interpretations or different ideas. Uncertainty is extolled for this reason by the philosopher Eric Fromm in his statement that "creativity requires the courage to let go of certainties".

## 1.2 Motivation and Goals

As highlighted in the previous section uncertainty and its visualization covers a vast territory for potential research. Looking at specific challenges noted for the area of uncertainty visualization and my goals will carve out a more manageable area for investigation.

### 1.2.1 Challenges in Uncertainty Visualization

There still exist many challenges for uncertainty visualization, for which some major ones have been summarized by MacEachren et al. [2005] as:

1. understanding the components of uncertainty and their relationships to domains, users, and information needs,

2. understanding how knowledge of information uncertainty influences information analysis, decision making, and decision outcomes,
3. understanding how (or whether) uncertainty visualization aids exploratory analysis,
4. developing methods for capturing and encoding analysts' or decision makers' uncertainty,
5. developing representation methods for depicting multiple kinds of uncertainty,
6. developing methods and tools for interacting with uncertainty depictions, and
7. assessing the usability and utility of uncertainty capture, representation, and interaction methods and tools.

The first four challenges relate to high-level user issues, such as decision processes which have been to now under explored, but will be examined from the starting point of uncertainty in cognitive processes in Chapter 4. Challenges 5 and 6 have been closer to the recent research in this area and are briefly summarized in Chapter 2. Lastly, Challenge 7 is also often neglected in research presentations and is the focus of Chapter 3. All these challenges should be kept in mind as you proceed through the chapters, and we will return to them directly in Chapter 9.

### 1.2.2 Goals of this Research

Finding commonality in all the types of uncertainty visualization is the concern of this dissertation. To reach this general goal, smaller sub-goals are to develop uncertainty visualizations to aid the understanding of domain specific uncertainty. Johnson and Sanderson [2003] state a primary goal of effective visualization is to provide a complete and accurate visual representation and this is a goal of any specific visualizations. They also note an important criterion is the user's psycho-physical ability to effectively understand the visualization. Carrying this further into the cognitive aspects, a larger goal is the pursuit of understanding of how uncertainty fits into a complete and accurate interpretation and decision model.

## 1.3 Methodologies: Micro and Macro

The research strategy to work toward my goals has been to use both a micro or bottom-up style for grounding in specific domains, as well as a macro or top-down approach using more theoretical knowledge to provide initial partitioning of the problem space. The top-down methodology involved looking at the problem from different perspectives extracted from a literature review, and focusing separately on the issues of:

- visual representations of uncertainty,
- analysis and evaluation of uncertainty visualizations,
- cognitive constraints when thinking about uncertainties, and the
- requirements of the user's task.

These issues, however, are not clearly distinguished as in any uncertainty visualization the role of each is interdependent.

The bottom-up approach involved delving, at varying depths, into uncertainty visualizations to support tasks pertaining to archaeological site data and reconstructions, rock property modelling in the seismic industry, and medical diagnostic support. The choice of three distinctly different domains is important for the purpose of making true generalizations. Thereby, any concepts that are found to apply to all three areas, will have more chance of applying in general.

In each domain, initial work was only to get an accurate understanding of the issues involved. For this bottom-up approach the methods varied across the domains but the most formal methodology was used for the problem of medical decision making. This included an observational study, contextual interviews, participatory design with the domain experts, and final evaluation using a form of pluralistic walkthrough (for a description of pluralistic walkthroughs see Bias [1994]). From qualitative research methodology the methods I utilized were mainly based on the phenomenological and grounded theory traditions [Creswell, 1998]. These methods have the potential of allowing the essence of experiencing uncertainty to be examined, as well as providing for the development of theory about it based on abstracting from the data. A qualitative methodology was chosen as the goals were to look at the big-picture of working toward a rich understanding of the key components and issues rather than specific details of any one implementation.

The fact that domain knowledge may be fundamental to interpreting observations (the epistemological assumption [Creswell, 1998]), implies that collaboration with the domain specialists and being immersed in the domain is important. For visualization support of medical decision making I have collaborated with physicians using a participatory design methodology, as well as being a part of the Ward of the $21^{st}$ Century research initiative at Foothills Hospital. Similarly for the archaeology domain I have taken a course in the area as well as participated in archaeological digs. For investigations in the seismic industry I have been working within the industry at CGGVeritas for almost six years and been a part of various collaborative research and development. This attempt to reduce the interdisciplinary separation as well as the collaboration with domain experts was very important to provide a grounded check on the understanding of the problem as well as the validity of any results.

Generalizing from both the macro and micro strategies was based around a form of thematic analysis [Boyatzis, 1998]. Using this approach, general themes may be sensed and coded to allow the analysis of qualitative data. Using this process to look at the specific uncertainty visualizations and the higher-level issues, generated a set of directives which detail important factors to consider when designing or evaluating uncertainty visualizations.

## 1.4 Organizational Summary

I have organized the dissertation in a roughly chronological manner following the order of my literature review, analysis, and the domain investigations outlined in the previous section. The first half of the dissertation is more top-down driven consisting of Chapters Two to Four. The second half encompasses more of the bottom-up approach to my research, ending with the results of integrating the two approaches.

### 1.4.1 Top-down Approach: Chapters Two to Four

Chapter Two provides a literature review of existing work in uncertainty visualization. The following chapter, Chapter Three, separates out what can be considered one aspect of a formal review of existing work, which is the analysis of visualizations. Chapter Three begins with an exploration of heuristic evaluation specifically on uncertainty visualization but then moves on to more general analysis of all visualizations. Chapter Four presents a summary of existing work in the area of cognitive psychology relating to reasoning under uncertainty and proposes some approaches for linking and grounding the uncertainty visualizations into the more fundamental process of reasoning which encompasses all visualization tasks. This chapter is chronologically out of order, as it occurred after some field work, but is presented here to give a high-level or "big picture" reference to cognitive issues which will be considered in the domain investigation chapters that follow.

### 1.4.2 Bottom-up Approach: Chapters Five to Eight

These chapters form the predominant strategy of the entire dissertation in that specific domains are investigated followed by a distillation of key factors that may generalize to uncertainty visualization as a whole. How the problem domains have been framed however is set down in the first half of the dissertation.

Chapter five provides the results of research into the specific domain of temporal and spatial uncertainty in archaeological reconstructions. An example visualization showing a site reconstruction is shown in Figure 1.3.

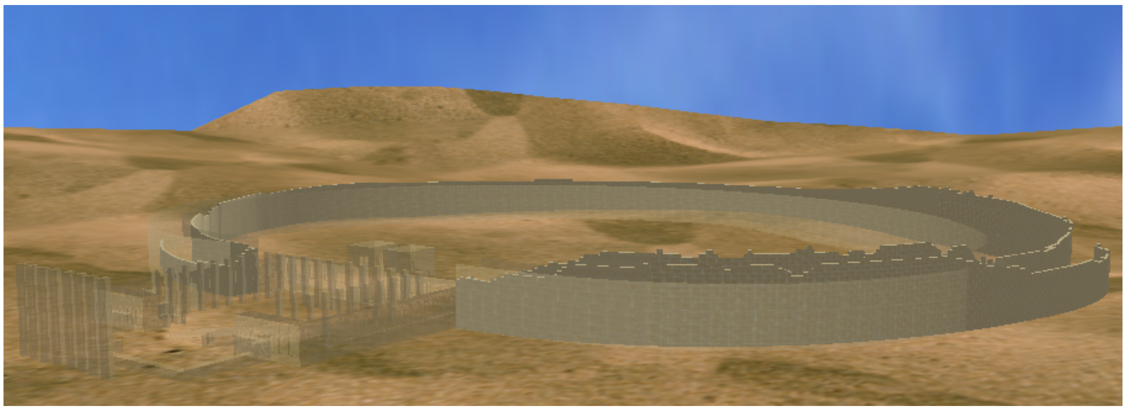


Figure 1.3: Example uncertainty visualization from archaeological site reconstruction. Theoretical early site is shown transparently in juxtaposition to recent survey data.

Chapter Six provides uncertainty visualizations developed for rock property modelling in the seismic domain. Two alternative representations were developed for the bi-directional vector field (with uncertainty in both orientation and magnitude) resulting from the modelling process: a static glyph[3] and an animated flow. Examples of the two representations are provided in Figure 1.4.

The final domain of medical diagnosis has been split into two chapters: Chapter Seven covering the study of the problem itself and analysis of the uncertainties involved, and Chapter Eight which covers the visualizations that were developed and their analysis. An illustration of the visualization system developed to assist in this task is shown in Figure 1.5.

[3]A glyph refers to an abstract encoding of multiple attributes forming a sign or other discrete graphical object.

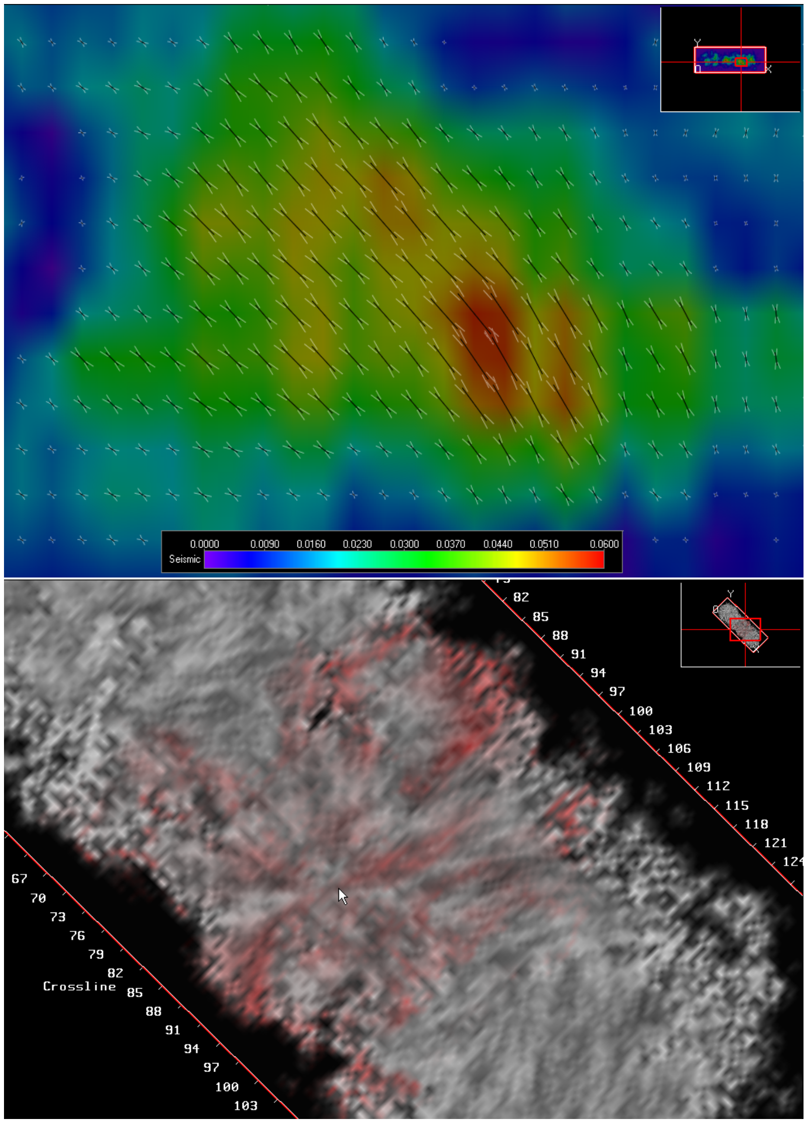



Figure 1.4: Uncertainty visualizations for use in seismic rock property modelling. Static glyphs, visible as black and white line segments, are shown in the top image and a flow based representation in the bottom image.

### 1.4.3 Integrating the Two Strategies

Chapter 9 returns to the top-down analysis to integrate it along with the bottom-up findings from the specific domains. It introduces a cognitive uncertainty categorization and by breaking out several important factors and tasks to consider (as directives), forms a

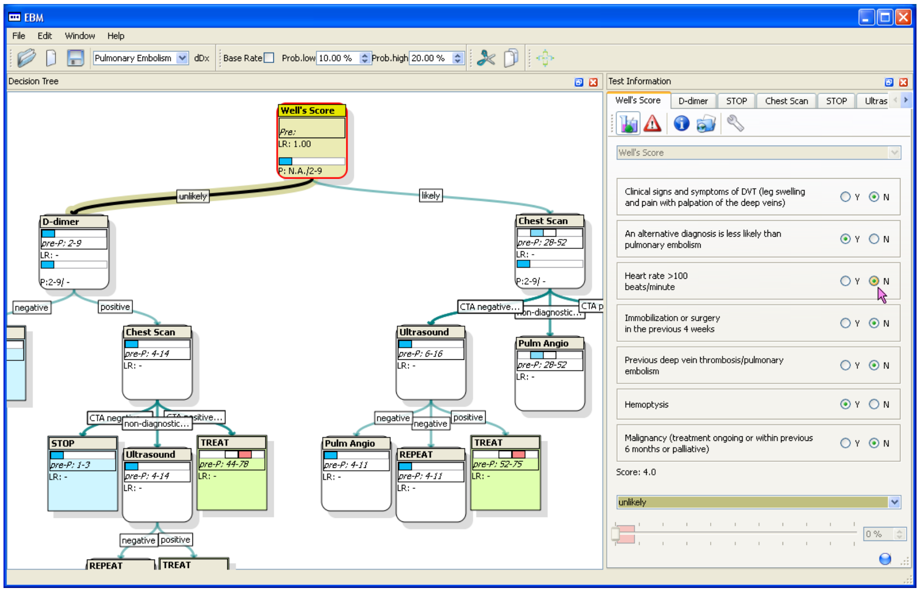



Figure 1.5: Uncertainty visualization for medical diagnostic reasoning support. View shows decision tree linked with actuarial scoring used for determining pretest probability.

framework for assisting the development of visualizations dealing with uncertainty. For the purposes of this dissertation I define a *framework* as any set of assumptions, concepts, or practices, that can be applied to structure a problem space or methodology. The framework directives are then applied in a post-hoc[4] evaluation of the visualizations that were developed. The final chapter reviews the results from all the chapters providing a summary of all contributions and recommendations for future work.

[4]The framework directives may have in fact been utilized in some primitive form during their development.

# Chapter 2

# Overview of Research in Uncertainty Visualization

*Quòd tertio loco à nobis fuit obferuatum, eft ipfiufmet LACTEI Circuli effentia, feu materies, quam Perfpicilli beneficio adeò ad fenfum licet intueri, vt & altercationes omnes, quæ per tot fæcula Philofophos excrucia runt ab oculata certitudine dirimantur, nosq́ue à verbofis difputationibus liberemur*.
– Galileo Galilei. *Sidereus Nuncius* (1610)

What was observed by us in the third place is the nature or matter of the Milky Way itself, which, with the aid of the spyglass, may be observed so well that all the disputes that for so many generations have vexed philosophers are destroyed by visible certainty, and we are liberated from wordy arguments.
– Translated by A. Van Helden *Sidereus Nuncius, or the Sidereal Messenger* (1989)

Visual certainty as stated by Galileo in the above quote, or the process of directly seeing evidence, is relevant as I review various existing visualizations aimed at increasing certainty by graphically exposing uncertainty. The power of visualization resounds in the old adage "seeing is believing", which, however, makes no mention of understanding. Therefore, we should be wary of the potential for a visualization to create certainty beyond what is appropriate.

This chapter as well as Chapters 3 and 4 comprise the literature review component of the dissertation. Initially I provide a meta-level review of existing work in uncertainty visualization. Following this Chapter 3 separates out what can be considered one aspect of a formal review of existing work, which is the analysis of uncertainty visualizations. This begins with an exploration of heuristic evaluation specifically on uncertainty visualization but then moves on to heuristic evaluation of visualizations in general. Moving toward user constraints and cognitive tasks in Chapter 4, I coalesce existing work in cognitive psychology relating to reasoning under uncertainty.

## 2.1 Review of Uncertainty Visualization

Uncertainty visualization has recently received more attention as the need for visualizing uncertainty along with data now has more general acceptance [National Academy of Sciences Workshop, 2005]. Visual representations for numerous specific data models with uncertainty have been proposed by various researchers [e.g. Cedilnik and Rheingans, 2000, Lodha et al., 2002b, Kao et al., 2001, Rheingans and Joshi, 1999]. The performances of some of these visualizations have also been evaluated with users for specific tasks such as by Grigoryan and Rheingans [2004] and Wittenbrink et al. [1996]. This is appropriate as it has been suggested that most visualization applications must be task-specific to be effective [Treinish, 1999]. To help guide research in this area, Johnson and Sanderson [2003] have called for more theoretical frameworks and visual representations for visualization tasks that involve uncertainty. In this section we will first review and critique one of the best surveys of the area, that which was provided by Pang et al. [1997].

Adding uncertainty into a visualization was described by Pang et al. [1997] as a parallel process to the visualization pipeline, which is shown in Figure 2.1. While this reveals their grounding in physical phenomenon rather than abstract data, nevertheless it is applicable to information visualization. To digress, information visualization refers to the visualization of data without an inherent spatial mapping[1], with focused theoretical aspects such as those of Ware [2004], Bertin [1983], and Tufte [2001], to be described in Chapter 3. The third component in the uncertainty visualization pipeline (see Figure 2.1) is important for design as it separates out the uncertainty introduced by a representation and the visualization itself, an issue often not carefully detailed in the presentation of a new visualization method. For user evaluation this also suggests the benefits of comparing multiple visualizations so that the uncertainty in this component may be roughly estimated.

[1]Information visualization may be considered more general, but it would be difficult to create a complete visualization without aspects that are abstract. The separation of visualization and information visualization is based more on historical reasons and I consider it vestigial.

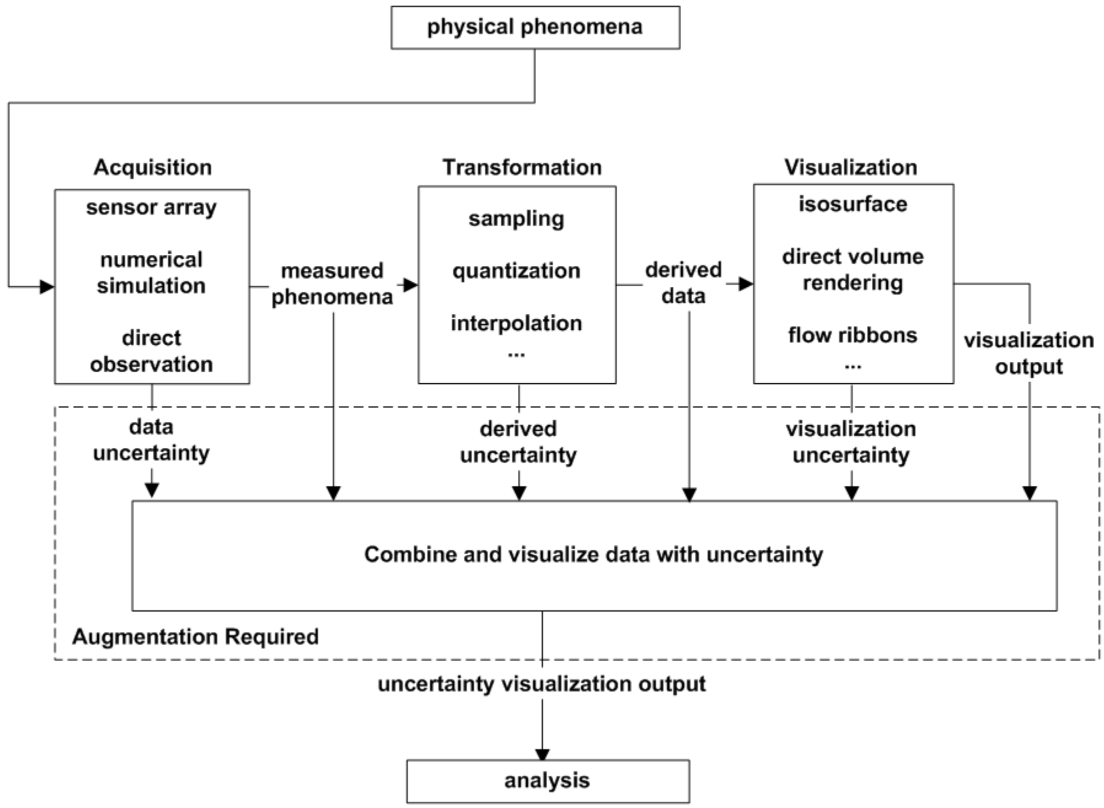



Figure 2.1: Pang et al.'s [1997] visualization pipeline showing three types of uncertainty potentially augmenting a data visualization.

Pang et al. [1997] reviewed uncertainty visualization in general and provided multiple classification schemes based on the uncertainty, data, and on the methods used to create visualizations. Their classification of the types of datum values[2] (scalar, vector, tensor, multivariate) is shown in Table 2.1 with a few example visualizations. They also used the location of the datum (0D, 1D, 2D, 3D, time, etc.) and its associated positional uncertainty for categorization. As this location of the datum category appears to combine competing factors and influences of space-time, I would propose that it be kept as three separate sub-criterion. One of these criterion would be the traditional visualization/information

[2]One could refer to the types of datum values in a more general sense as data types, but I will use their terminology in this section for consistency.

Table 2.1: Pang et al.'s [1997] typology of datum values applied to other researchers uncertainty visualizations.

| Value | Example Uncertainty Visualizations |
|---|---|
| scalar | line median, standard deviation [Tufte, 2001]<br>box plots, box-and-whisker plots [Chambers et al., 1983]<br>notched box plots [Chambers et al., 1983] |
| multivariate | scatter plots [Chambers et al., 1983]<br>probability map [Kao et al., 2001][van der Wel et al., 1994]<br>Bayesian model [Hanson et al., 1999] |
| vector | glyphs [Tufte, 2001]<br>fluid flow [Lodha et al., 1996b]<br>glyphs & environmental vector fields [Wittenbrink et al., 1996]<br>reaction-diffusion texture [Sanderson and Johnson, 2003] |
| tensor | glyph tensor probe [Pang et al., 1997]<br>principle axis ellipsoid [Leonhardt et al., 2006] |

visualization dichotomy based on there being a spatial location of the datum or not, for which Pang et al. also provide an explicit category relating to the visualization, rather than the data, called visualization axes mapping: experiential or abstract. With experiential being defined by them as visualizations for which the axes mappings replicate the viewers experience with the phenomenon, versus abstract for those that do not. The second criterion could be non-temporal dimensionality on its own, with the third criterion being if time is an additional dimension.

Pang et al.'s [1997] last datum based criterion is the extent of both location and value being either discrete or continuous. This characteristic combines two different things which I would prefer to more clearly break apart, the first, is the sampling in each dimension of a discrete or continuous phenomenon in that dimension[3]. The second criterion, is the valid range of datum values (individual samples) being either discrete or representative of a continuous function (i.e. real valued). The separate consideration of sampling and sample type may have been their intent but they only ever provided a single categorization

[3]Sampled data is itself always discrete and so the original phenomenon must be the reference point.

(continuous or discrete) for any one visualization.

Relating aspects of the data to those of the visualization, Pang et al. [1997] provide the two characteristics: visualization extent, and visualization axes mapping. The visualization extent is used to characterize whether the chosen representation indicates a discrete (e.g. points, glyphs) or continuous range of data (e.g. curves, surfaces). To parallel our separate consideration of datum value, adding a characterization of the value encoding being perceived as discrete (e.g. 8 colourmap entries) versus continuous (e.g. 256 levels of grayscale) should be considered. Pang et al.'s final characteristic of experiential or abstract visualization axes mapping was already described, and they also noted its parallel to the historical schism between visualization and information visualization. A relevant point is almost all the visualizations they described were classified by them as experiential (5 out of 6 pre-existing, and 19 out of 20 new ones) and so one might argue the abstract or information visualization categorization may not be very well analyzed with these criteria.

Pang et al. [1997] additionally organized their new uncertainty visualizations based on how the uncertainty information is encoded. For this they provided seven categories: add glyphs, add geometry, modify geometry, modify attributes, animation, sonification, and psychovisual approaches. With sonification they moved into non-visual input, and so haptics and olfactory might be additional top level categories. Non-visual representations may assist interpretation as redundant or additional encodings of uncertainty [Jacobson, 2002]. Pang et al.'s psychovisual approaches included stereo based blurring, and subliminal images. The majority of these visually based categories I would group into two styles: encoding using additional uncertainty representations (in addition to data representations), and uncertainty encoding through modification of data representations (including animation). These two styles may be more generally called the modification of existing graphic variables (termed verity visualizations by Wittenbrink et al. [1996]) versus the use of additional graphic variables. Modifications are inherently tied to the underlying data representation, while additional representational encodings need to be cognitively linked to allow the usual treatment of uncertainty as meta-data. The use of additional variables at the same location in the view plane is called overloading, as each variable can carry information. This usage of the term overloading implicitly refers to the primacy of the view plane graphic variable (i.e. spatial encoding) as it is one of the most flexible and rich variables for encoding information [Bertin, 1983].

In Griethe and Schumann's [2006] review of uncertainty visualization they discuss Pang et al.'s categorization, but based on the dominance of existing methods for scalar uncertainty visualizations propose two main categories: direct uncertainty visualization, and using uncertainty indirectly. Indirect use they term parameterization, which is the use of uncertainty in the filtering, mapping, or rendering of the raw data itself. They provide filtering out data based on an uncertainty threshold as an example of indirect use, but how it can be used later in either mapping or rendering and not be considered "direct" was not made clear. Thus I would place their indirect methods into what I grouped together as the modification of existing graphic variables. Direct uncertainty visualization Griethe and Schumann break down further into five categories based on the use of: unused graphical variables, additional graphical objects, animation, interaction, and other non-visual human senses.

The nature of the data will constrain what uncertainty representations are appropriate, and uncertainty representations for various data types have been proposed (for examples see Davis and Keller [1997], Djurcilov et al. [2002], Botchen et al. [2005], Lee et al. [2007]). Lodha et al. [2002b] present techniques for probabilistic points and their movement. Grigoryan and Rheingans [2004] represent surfaces with uncertainty using point clouds perturbed from the original 3D surfaces based on a probability distribution of the data. Numerous other uncertainty representations have been published, and many are listed in the survey by Pang et al. [1997]. Table 2.2 provides a listing of some un-

Table 2.2: Fundamental geometric data representations and uncertainty visualizations.

| Representation | Example Uncertainty Visualizations |
|---|---|
| point/<br>particle | flow [Lopes, 1999]<br>particle movement: galaxy, opacity, colour[Lodha et al., 2002b]<br>GIS position [Lodha et al., 2002a] |
| line/<br>contour | contouring dust-cloud [Lopes, 1999]<br>architecture lines [Masuch and Strothotte, 1998]<br>procedural line annotation [Cedilnik and Rheingans, 2000] |
| surface | hue and texture [Rhodes et al., 2003]<br>interpolants [Lodha et al., 1996a]<br>marching cubes [Lopes, 1999]<br>points on surfaces [Grigoryan and Rheingans, 2002]<br>isosurface colour[Rhodes et al., 2003] |

certainty visualizations created around the basic geometric primitives[4] for one, two, and three dimensions. The addition of uncertainty is often performed by extruding these basic representations in the space or time dimension (e.g. a point becomes a region). The review of visualizing errors and uncertainty by Johnson and Sanderson [2003] concludes with the need for more formal evaluations, new representations, and more widespread presentation of errors and uncertainty.

Thus enough graphical constructs and algorithms exist to provide a plethora of visualizations. In order to prune down our design space we can use theory from information visualization (e.g. Ware [2004]) and cognitive psychology (e.g. Shelton and McNamara [2001]) as a guide as to which construction styles may be the most comprehensible and effective. Considering a representation such as lines, they are a basic drawing primitive, yet there exist a large number of ways to render lines (strokes) to express different information. Strothotte and Schlechtweg [2002] discuss various line rendering techniques and how they can be used to provoke different interpretations. Strothotte et al. [1999b] and Cedilnik and Rheingans [2000] have shown how lines can be rendered in various ways to

[4]Other primitives may also be considered fundamental, but these share the property that they are all supported directly by standard graphics hardware.

Table 2.3: Example categorization of uncertainty visualizations with examples.

| **Category** | **Example Uncertainty Visualizations** |
|---|---|
| spatial/positional | medical scan segmentation [Grigoryan and Rheingans, 2004]<br>molecular [Rheingans and Joshi, 1999] |
| temporal | archaeological [Zuk et al., 2005] |
| spatiotemporal | archaeological [Strothotte et al., 1999a]<br>global positioning systems [Lodha et al., 2002a] |
| non-spatiotemporal | translation confidence [Collins et al., 2007] |
| interpolation<br>& extrapolation | IFS interpolation [Wittenbrink, 1995]<br>interpolation [Pang et al., 1994]<br>missing data [Twiddy et al., 1994, Wyvill and Wyvill, 2000] |

express uncertainty.

One domain where a considerable amount of uncertainty visualization research has been done is the field of geographical information systems (GIS) (for examples see Howard and MacEachren [1996], Plewe [2002], Lucieer and Kraak [2004], Lucieer et al. [2005]). In a book devoted to the subject of GIS uncertainty, Zhang and Goodchild [2002] group uncertainty based on the types of data the uncertainty relates to: continuous variables, categorical variables, and objects. GIS objects refer to higher level abstractions that are often region-based (e.g. road, building). Zhang and Goodchild focus on the models and processes related to spatial data. They describe spatial interpolation models based on Kriging [Krige, 1962] that are somewhat unique as they inherently create an uncertainty model as part of the interpolation process. More specifics of the types of visualizations investigated for GIS and some of the few studies which compare different types of visualizations will be covered in the next section.

The various existing uncertainty visualizations cover a wide-range of data, uncertainty types, and user tasks. The question one can ask is how do these categorizations help us understand how to create new visualizations of uncertainty or what are important factors. If looking to deepen one's understanding of uncertainty issues on space and time one might consider a spatiotemporal categorization as well as other highly related but general issues such as the process of filling in missing data. Table 2.3 provides a potential categorization based on a spatiotemporal focus, and a listing of some uncertainty visualizations categorized by it.

Taxonomies may help us avoid reinventing the wheel for specific problems, by understanding common traits and how they might apply to similar problems. Provided the large number and variety of uncertainty visualizations, taxonomies and categorizations can be used to analyze how these different methods compare and determine what are the important aspects. Information visualization theory may also provide a good basis for this comparison and analysis, and is explored further in the next section.

## 2.2 Design and Evaluation of Uncertainty Visualizations

Information visualization theory provides us with a source of knowledge about what and how visual representations might be used for efficient and accurate visual processing. This knowledge applies to both combining uncertainty information into a visualization (e.g. overloading with multiple visual variables), as well as creating a separate visualization of the uncertainty, where cognitive integration issues are relevant [Ware, 2004]. However many information visualization theories are founded on behaviour observed in isolation and so one must be careful in applying them to practical implementations, because in everyday situations the user may be multi-tasking. Therefore applicability will be strongly influenced by domain and task considerations, which are explored in Chapters 5 to 8.

There has also been research into the best representations for uncertainty. MacEachren [1992] discussed the visualization of uncertain information in GIS. He broke uncertainty down into visualizing accuracy, and visualizing precision, as separate tasks requiring different strategies. MacEachren proposed the use of colour saturation and blurring as being conducive to indicate uncertainty. This recommendation may relate to the potential for intuitive reading based on Pierce's three types of signs: icons, indexes, and symbols [O'Sullivan et al., 1994]. Pierce's icons have a direct perceptual resemblance to what they indicate, and it is the ambiguity created by the colour saturation and blurring that MacEachren [1992] suggests makes them logical to use. He related how Bertin's [1983] original graphic variables and these new variables could be applied to uncertainty. MacEachren [1992, p.13] proposed that the graphic variables

> ... size and value are the most appropriate for depicting uncertainty in numerical information, while color (hue), shape, and perhaps orientation can be used for uncertainty in nominal information. Texture although it has an order, might work best in a binary classification of "certain enough" and "not certain enough" that could be used for either nominal or numerical data.

MacEachren's use of the term texture is referring to Bertin's graphical variable grain of resolution. Texture is the word used in the English translation (a bad translation for Computer Graphics people), [5], and thus I will refer to this variable henceforth as grain. He also proposed evaluation based on tendency to Type I visualization errors (seeing patterns that do not exist) and Type II (failure to notice patterns and relationships) visualization errors (Table 2.4).

Table 2.4: Classification of visualization errors [MacEachren, 1992].

| **Category** | **Definition** | **Statistics** |
|---|---|---|
| Type I | seeing patterns that do not exist | false positives |
| Type II | failure to notice patterns and relationships | false negatives |

In the area of GIS a variety of user evaluations have also been performed to assess the value of the uncertainty visualizations [Evans, 1997, Leitner and Buttenfield, 2000, Leitner and Curtis, 2006]. In one example of using multiple evaluations, Slocum et al. [2003a] describe three successive evaluations, with intervening refinements, of a visualization for

[5] Appropriately a warning about misinterpretation is given by the translator W. J. Berg.

global water models and their uncertainty: the first evaluation was with domain experts, the second with usability experts using heuristic evaluation, and the final evaluation with decision makers. They suggest that it may have been worthwhile to get the decision makers involved at an earlier stage of the process. Determining if visualizations can be shared across user groups is likely a difficult task in itself.

Kardos et al. [2003] studied the effectiveness of various spatial data representations of uncertainty. In their qualitative user opinion survey they compared the use of fog, an adjacent map, texture overlay (grain), blur, blinking pixels, sound, colour saturation, pixel mix (hue count), and animated regions to demonstrate regions of uncertainty. In this study with 44 participants, only blinking pixels was consistently judged more useful than non useful, with adjacent maps and texture overlays (grain) judged marginally more useful than non useful. Rating categories were: non-useful, ineffective, limited, moderate, good, and excellent[6]. They introduced a hierarchical tessellation (quadtree) overlay with the level of subdivision based on uncertainty, but did not advance to the point of user testing. More recently Kardos et al. [2005] did a web-based survey to test the effectiveness of hierarchical hexagonal or rhombus (HoR) tessellations against hierarchical square tessellations, blinking areas, adjacent map, texture overlay, fog, blur, and animation. In this survey they found the HoR tessellations had similar expressive power to blinking, adjacent maps, and texture overlay. These techniques were considered to be better than square tessellations, fog, blur, or animation. These results one would almost expect from the earlier study, as the tessellations form a type of texture overlay. The surveys were not described in detail and most participants were experienced in GIS, so it is difficult to say if these results can be generalized.

With spatial uncertainty, visual representations were found better than verbal representations in one decision problem looked at by Kirschenbaum and Arruda [1994]. Finger

[6] If not non-useful the participant could also choose to enter their own description. So useful indicates any one of the last five categories or user defined.

and Bisantz [2002] compared icons with levels of blur, and with and without text, and found the addition of text provided no statistical advantage. Therefore for some tasks uncertainty representations with larger granularity (capable of encoding fewer bits) may suffice for expressing the uncertainty necessary for the decision process at hand. Many other representations have also been evaluated but further evaluations and research is still needed [MacEachren et al., 2005].

Information visualization theory integrates the perceptual and cognitive theory that may help understand why certain visualizations work well and some do not. This understanding can provide the design patterns for visualization that help us avoid pitfalls. It also helps us with the more nebulous problem of how these visualizations are to be used for specific tasks and how to judge their relative performance. This also falls into the area of HCI and human factors.

Human factors and how they apply to visualizations has been surveyed by Tory and Möller [2004] and they found a somewhat limited utilization of the theory in visualization research. They summarize how the user-centered (participatory), task-based, and perception and cognition-based design, can focus on satisfying the users' goals by understanding their strengths and constraints. User studies are another approach used in human factors research. Kosara et al. [2003] discuss when user studies should be done and review some common problems and limitations of these studies. One example of the difficulty to generalize results comes from a recent user study with uncertainty representations for airline traffic flow. Masalonis et al. [2004] found for one task that the probability density function graphs that provide the most uncertainty information were given lower subjective ratings than best guess and range displays. It was unclear why the participants did not want to utilize the more detailed uncertainty information, and thus generalization is difficult. Another major problem with user studies is that they are naturally biased against new techniques which require a long period of training. This is because the costs to perform the training for such a study are prohibitive, and as a result the studies are run with under-trained par-

Table 2.5: Typology for visualizing uncertainty [Thomson et al., 2005].

| Category | Definition |
|---|---|
| Accuracy/error | difference between observation & reality |
| Precision | exactness of measurement |
| Completeness | extent to which info is comprehensive |
| Consistency | extent to which info components agree |
| Lineage | conduit through which info passed |
| Currency/timing | temporal gaps from info collection |
| Credibility | assessment of info source |
| Subjectivity | amount of judgment included |
| Interrelatedness | source independence |

ticipants. Avoiding this evaluative shortcoming is the motivation of long-term case studies as described by Shneiderman and Plaisant [2006].

Beard and Buttenfield [1999] created a GIS based framework for error and uncertainty that involves an initial phase of mapping data to error analysis and a second phase of mapping to graphical display. There has been a call, however, for more theory for visualizing uncertainty [Johnson and Sanderson, 2003, MacEachren et al., 2005], and a few have been put forward such as Thomson et al.'s typology of uncertainty [Thomson et al., 2005]. They consider contributions from Pang et al.'s [1997] classification and Gershon's [1998] high-level taxonomy of uncertainty. This typology was developed for geospatially referenced data and for intelligence analysts to use in analytic design (shown in Table 2.5). However, the typology was found to be general enough to be useful when applied to reasoning uncertainty [Zuk and Carpendale, 2007], which will be described in more detail in Chapter 4.

Looking at the relationship of analytic tasks to representations, Amar and Stasko [2004] present a set of knowledge precepts for design and evaluation of information visualizations. They describe a rationale gap, as being the separation between seeing a relationship and confidently understanding it in terms of making a decision. They proposed three rationale precepts to reduce this separation:

1. Expose Uncertainty

   ... a system can help bridge the Rationale Gap by exposing uncertainty in data measures and aggregations, and showing possible effect of this uncertainty on outcomes

2. Concretize Relationships

   ... a system can help bridge the Rationale Gap by clearly presenting what comprises the representation of a relationship, and present concrete outcomes where appropriate.

3. Expose Cause And Effect

   ... a system can help bridge the Rationale Gap by clarifying possible sources of causation.

While the first precept explicitly mentions uncertainty, the second two implicitly also deal with uncertainty. Both the concretization of relationships and formulation of cause and effect can be considered tasks directly aimed at reducing the uncertainty in knowledge.

Generalizing from Amar and Stasko [2004], design rules and recommendations can often be utilized for evaluation, and this dual role will be utilized throughout this dissertation. Chapter 3 will return to this topic and looks deeper into the evaluation of uncertainty visualizations. Specific evaluations will also be described in Chapters 5, 6, 8, and 9.

## 2.3 Cognitive Aspects of Uncertainty Visualization

In this section I move to some higher level theory relating to uncertainty from a human factors point of view. Gershon [1998] framed uncertainty as a part of imperfect knowledge and presented six types of causes for it:

1. incomplete information,
2. inconsistency,

3. information too complicated,
4. uncertainty,
5. imperfect presentation, and
6. corrupt data/information (imperfection).

Singling out imperfect presentation he broke it down further into: information overload, inappropriate presentation, and inappropriate device. These are challenges to the visualization process itself to avoid the creation of imperfect knowledge (if we abstract away the data/information). He also called for the development of principles for imperfection/uncertainty management, and noted that potential user variation should be considered.

From a more general cognitive perspective one taxonomy of reasoning uncertainty was presented by Kahneman and Tversky [1982] as the variants of uncertainty:

1. External (Dispositions)
   (a) Distributional (Frequencies)
   (b) Singular (Propensities)
2. Internal (Ignorance)
   (a) Reasoned (Arguments)
   (b) Introspective (Confidence)

The aspects of knowledge noted by Gershon relate to the reasoned aspects of information, with information usually relating to the external. Therefore the introspective (confidence) category may be understated in looking at imperfect knowledge as the end game. Similarly most of the work described in Section 2.1 related to external uncertainty. However, I will attempt to consider and place more focus on these internal aspects as I expect they can provide guidance for developing more insights into uncertainty visualization. This consideration of reasoning uncertainty itself and the related work will be investigated further in Chapter 4.

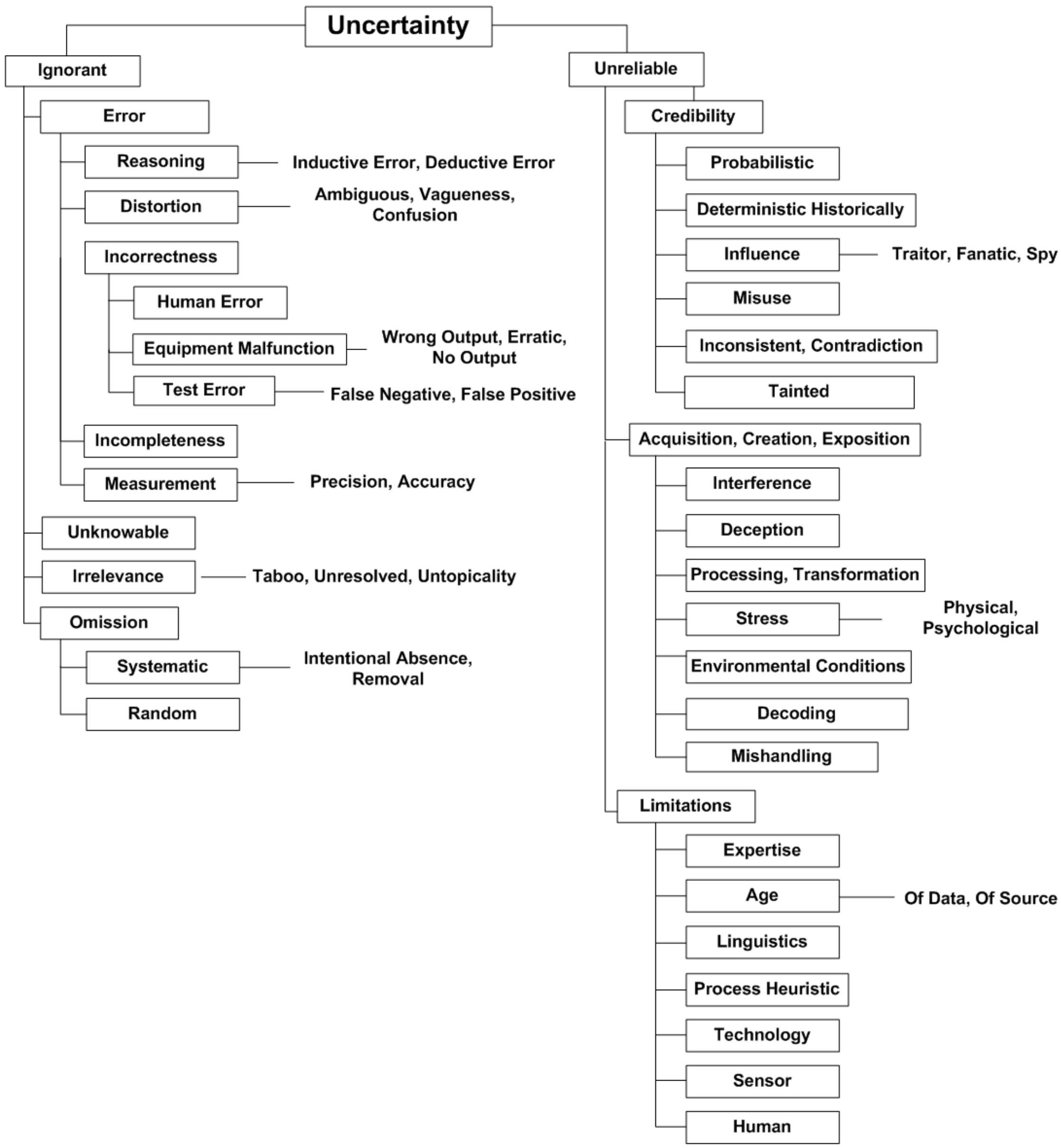



Figure 2.2: Watkins [2000] typology of uncertainty (components of physical and psychological stress have been left out of the figure).

For the purpose of visualization to aid decision support, Watkins [2000] examined cognitive aspects of uncertainty and also developed a taxonomy of "how and why" things are uncertain, which is shown in Figure 2.2. Other taxonomies of cognitive uncertainty exist (e.g. Howell and Burnett [1978]), and one extensive one including management strategies for dealing with uncertainty has been provided by Hutton [2004]. It is an open question as to whether all these types of uncertainty, or causes of uncertainty, have specific visualization needs. For many problem areas the concise typology provide by Thomson et al. [2005] in Table 2.5 may be a sufficient starting point for differentiating uncertainty.

If one needs further motivation on why this internal uncertainty (reasoned and introspective) is worth singling out we need only to look at the fact that we deal with uncertainty on a daily basis but have internalized its management to the point where we may hardly be aware of it. MacEachren et al.'s [2005] challenges listed in Chapter 1 mostly call for guidance and higher levels of understanding. Bridging the gap between Amar and Stasko [2004] general precepts and how to design uncertainty visualizations that can best assist interpretation will require further research. This is the area to which the rest of this dissertation will attempt to contribute.

## 2.4 Conclusions

In creating a new visualization one does not find the answers to design options simply by seeing what visualizations have been created for a particular type of data. Hence the visualizations listed in Tables 2.2 and 2.3 were by no means meant to be an exhaustive list. Furthermore, even when a given representation is chosen, to determine the best interaction methods one can not just choose one from an existing list of methods and benefits. The reviewed works provided various structures for understanding the problem in general, but how these "tools" relate to the users problems may be a challenge to determine. As was suggested in the preceding section the design should be equally driven by the user and task considerations. Therefore evaluations for visualizations in particular tasks will be important for estimating transferability and reuse of visualization techniques. Additional related work will be discussed in each of the following chapters pertaining to the corresponding topic or domain.

Information visualization theory integrates the perceptual and cognitive theory that may help us understand why certain visualizations work well and others do not. This understanding may provide the design patterns for visualization that help us avoid common problems, but it does not bridge all the gaps in determining the best designs. Continuing with the goal of generalizing knowledge from existing uncertainty visualizations, the next chapter includes a more formal analysis and evaluation of some particular uncertainty visualizations, and extends into information visualization evaluation in general.

# Chapter 3

# Analysis and Evaluation of Visualizations

Everything is vague to a degree you do not realize till you have tried to make it precise.
– Bertrand Russell. *The Philosophy of Logical Atomism* (1918)

Analysis of visualizations in general requires a deeper understanding of their composition. Although a number of theories and principles have been developed to guide the creation of visualizations, it is not always apparent how to apply the knowledge in these principles. We describe the application of perceptual and cognitive theories for the analysis of uncertainty visualizations. General theory from Bertin, Tufte, and Ware are outlined and then applied to the analysis of eight different uncertainty visualizations. The theories provided a useful framework for analysis of the methods, and provided insights into the strengths and weaknesses of various aspects of the visualizations[†].

## 3.1 Introduction

The need for visualizing uncertainty along with data now has widespread acceptance. However the task of including the additional uncertainty information into an existing or new visualization while maintaining ease of comprehension for both the data and the uncertainty is not easy. As a result, the visualization of uncertainty is still not standard practice. Various researchers have proposed visualization methods to present uncertainty [National Academy of Sciences Workshop, 2005] and some have used HCI methodology to analyze and evaluate the visualizations.

[†]Portions of this chapter have been previously published in Zuk and Carpendale [2006], Zuk et al. [2006]. Thus any use of "we" may refer to Torre Zuk and Sheelagh Carpendale or Torre Zuk, Lothar Schlesier, Petra Neumann, Mark S. Hancock, and Sheelagh Carpendale

Recently Johnson and Sanderson [2003] called for the development of formal theoretical frameworks for visualizing error and uncertainty. Before developing new frameworks it is worth examining existing perceptual and cognitive frameworks to better understand them with their strengths and their short comings, and to ensure we are utilizing those frameworks that already exist. With this goal of more fully understanding research in this area, we chose the three commonly cited theoretical approaches and use their principles to analyze eight representative uncertainty visualizations across a wide variety of domains.

A variety of theories and frameworks for analysis of uncertainty and uncertainty visualization are available. Some of these were mentioned in Chapter 2 such as Pang et al. [1997], van der Wel et al. [1994] and Beard and Buttenfield [1999]. While these and other frameworks could be applied and would be useful, at this time we are focusing on the perceptual basics and will only consider the general perceptual and cognitive theories described in the next section.

## 3.2 Perceptual and Cognitive Theory

From the large number of contributors to perceptual design theory we have chosen the subset of Bertin [1983], Tufte [2001], and Ware [2004] as perspectives for our analysis. Other perspectives could have been chosen and may be equally valid, Chambers et al. [1983] or Slocum et al. [2003b], for example. However, since the theories from Bertin, Tufte, and Ware are widely cited, they were deemed to be a good starting point. While we are simply citing Ware's text, we recognize that Ware's collection of theories includes explanations from many cognitive scientists.

Each of this trio of researchers (Bertin, Tufte, Ware) has an extensive set of principles. Therefore to limit the scope we will consider a selection of the trio's perceptual and comprehension driven principles. This will include Bertin's [1983] framework of the plane and retinal variables, Tufte's [2001] theory of data graphics, and excerpts from Ware's [2004] textbook on information visualization.

The following overview is included to provide the cognitive context for the specific details that follow. It summarizes the main components from the theory of the Bertin, Tufte, and Ware as used in this chapter. Many of these theories and guidelines are represented in all three sources, but from slightly different perspectives.

### 3.2.1 Bertin

In Bertin's [1983] framework called the Properties of the Graphic System, he presented eight visual variables[1]. The planar dimensions $(x, y)$ are two of Bertin's visual variables, and for any location on the plane a visible mark can utilize any or all of the six retinal variables: size, value, grain (a retranslation of the original variable name), colour, orientation, and shape. While developed for the printed page, Bertin's framework is still generally applicable to digital displays as it has been shown useful by many researchers (e.g. MacEachren [1992], Beard and Buttenfield [1999]). There are some adjustments, however, that should be made when applying it to current display technology, which will be briefly discussed after reviewing the visual variables. MacEachren [1995] and Ware [2004] have proposed some modifications to these variables and describe additional variables, but for our purposes here we will limit ourselves to Bertin's original variables.

Each of the eight variables is categorized based on its potential for immediate perceptual group selection, perceptual grouping characteristics, natural perceptual ordering (not learned), ability for quantitative comparisons, and length (the number of discernible elements that can be represented in the set, i.e. cardinality). In terms of perceptual processing speed, a variable is called selective if it can be perceived immediately over the entire plane without considering individual marks sequentially. The performance of this parallelized perceptual task has been labeled preattentive processing [Ware, 2004], in which the

[1] This framework is also described in Bertin [1981], which was translated first, but is subsequent work to Bertin [1983].

number of distractors does not impact performance. Selective classifications may break down when encodings use multiple variables, or shape, which has components of the two other variables: size and orientation. Thus Bertin classifies shape as not being selective while Ware does call it preattentive. This is likely due to the fact that shape is a complex variable (has infinite length) and the number of different types of distractors does have an impact. To be used for selective processing the usable length of any variable must be greatly reduced. If variation in a variable could be ignored so as to consider variation only in other variables, Bertin called the variable associative. This notion of associativity is closely related to the characterization of separable and integral variables Ware [2004], which is relevant to uncertainty visualization if the user needs to consider the data and its' uncertainty independently (separable), or it is more important to see them as a whole (integral). MacEachren et al. [1998b] found in one study that integral encoding of data and uncertainty negatively impacted the performance of detecting clusters in the data. Bertin's classifications of variables (e.g. which have natural perceptual ordering) may be contested in specific scenarios, or over subsets of a variable's length, but in practical terms we consider them to be useful.

### Visual Variables

Bertin describes *the two planar variables* (x and y) as the richest of the variables in that they are selective, associative, ordered, and quantitative. To make use of the retinal variables to change the appearance and thus encoding, of a mark; the mark must first be implanted at some location (x, y) on the plane. Bertin categorized this implantation as being point, line, or area based. The type of implantation affects the length of the retinal variables. Area implantation raises two issues: the orientation variable can no longer be processed selectively, and the meaning of any variable is read over the entire region of implantation (i.e. quantities must be normalized per unit area or they may be incorrectly read).

*Size* is the only retinal variable that can be quantitative (allowing ratios of data to be directly perceived). It is selective but not associative, and it is ordered. While any variable can be implanted as an area, size is naturally implanted (encoded) as an area. Therefore given a fixed area, the area itself cannot change size but its constituent points or lines can, and is then classified as the separate variable *grain*.

*Grain* is the variation in scale of the constituent parts of a pattern[2]. As grain can be considered a composite of size, it can be ordered on that basis. It is both selective and associative. The length of this variable is affected by the size of implantation. Thus making a larger mark allows more steps that can be distinguished.

*Value* is the ratio of perceived black to perceived white on a given mark. It is ordered. Bertin's usage is similar to the value in the HSV colour model [Ware, 2004]. Contrasting this, Ware makes a clear distinction in the definitions of luminance, brightness, and lightness from a perceptual context.

*Colour* is the chromatic variation of two marks with the same value. It is more closely associated with hue variation than saturation. As the pure or monochromatic colours associated with full saturation (not in the HSV sense) do not have equal value, Bertin did not create separate variables for hue and saturation. The colour variable has no implicit order[3], but is selective and associative.

*Orientation* is the variation in the angle between marks. This variable is associative, but only selective for point and line implantations. It is not ordered[4]. Numerous ways exist to split the 360 degrees of orientation into steps (theoretically infinite). Bertin states, however, that using only four steps provides for maximal selectivity. To enable the utilization of perceptual sensitivity to parallelism is a main reason for restricting the length to four.

[2]Pattern is texture in French. Thus texture was the translation of this variable name in Bertin [1983].

[3]MacEachren [1995] suggests colour saturation is ordered, and even some subsets of hue are ordered, based on a HSV decomposition.

[4]It might be considered ordered given symbolic associations, but not at the perceptual reading level to which Bertin refers.

*Shape* is the most ambiguous variable as it incorporates aspects of size and orientation. It has no perceptual ordering but has symbolic ordering (e.g. triangle, square, pentagon, ...). It has infinite length, but is only associative. Its flexible nature allows complex symbolism but this must be learned and therefore is never universally understood [Bertin, 1983].

#### Application of the Framework to Digital Displays

With digital displays Bertin's original visual variables will each exist in a slightly different representational domain and so considerations should be made. We will refer to Bertin's original domain as the page, and the digital display as the screen. The reduction of the length of a variable is one major difference that must be accounted for (it effectively reduces the amount of information/bits that can be carried in a variable). The two planar dimensions of the screen are currently lower resolution than is possible with the page, and so to pack the same amount of information into a small space is not possible. This effectively reduces the length of that variable as human visual acuity is beyond current screen technology. This means that to perceive the same amount of information with a screen the eye must make saccadic movements to cover more area, and this has subtle repercussions for perception. A large printed map contains more information than most large displays are capable of showing at once, and so another significant change is the user interaction required for scrolling.

Bertin's first retinal variable, size, is similarly affected as the plane was; its length is effectively reduced. Similar arguments can be made for affecting the length of texture (grain), orientation, and shape as they are all implanted on the plane as marks. Value is affected in that the dynamic range and resolution of luminance from the page can not yet be equalled on the screen. Leitner and Buttenfield [2000] found that the prominence of dark over light (value) was reversed from paper to the CRT (reflective versus emissive). Colour length reduction will not always be the case, as the gamut of printing technologies varies significantly and so in some cases it may be that an increase in length of the colour variable can be achieved. The nature of association, selectivity, order, and quantity are quite transferable to the digital medium. Quantity is tied to the length of a variable and so will be reduced for the plane and size variables on the screen.

The greatest affect of the digital medium is the introduction of (or just simplification of adding) more visual variables. Visual variables in the digital domain are analyzed using Bertin's perspective by MacEachren [1995] and Ware [2004]. Strong variables such as motion, disparity (stereo-displays), and blinking are not possible on the page. Other variables such as blur, concavity, and shape from shading, while not described by Bertin are also transferable back to the printed page. Bertin's variables can be thought of as one possible set of basis vectors that span a sub-space of 2D visualizations. Expansion beyond the limits of the printed page adds additional visual variables and added dimensions of depth and time. However, 3D visualizations after being projected to 2D, can be treated as implantations on the plane, and thus can be analyzed using Bertin's framework.

### 3.2.2 Tufte

Tufte has written a series of books on the graphical presentation of information [Tufte, 2001, 1990, 1997, 2006]. Here we will primarily utilize Tufte's principles for graphic excellence and integrity [Tufte, 2001]. These are general principles that usually lead to good visualizations. Other aspects of his theory provide optimization rules and design patterns. Tufte has summarized most of his concepts in one complex principle: *graphical excellence*, which he defines as that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space [Tufte, 2001].

In order to create *graphical excellence* Tufte [2001] has specified guidelines such as: avoid distorting what the data shows; encourage the eye to compare the data; present a large amount of data in a small space; reveal multiple levels of detail in the data; and closely integrate statistical and text descriptions with the data. These encourage graphical clarity, precision, and efficiency [Tufte, 2001]. Tufte provides numerous examples of graphical excellence most of which are multivariate.

To promote *graphical integrity* Tufte [2001, p.77] provides six principles to be followed:

1. graphic representations relating to numbers should be directly proportional to the quantities represented,
2. clear and detailed text should be used wherever needed to avoid ambiguity,
3. show data variation and not design variation,
4. money in time series must be adjusted for inflation[5],
5. the number of dimensions used for reading data should not exceed the number of data dimensions being represented (e.g. don't make scalars an area), and
6. do not show data out of context.

As the name integrity suggests, following these principles avoids deception and misinterpretation.

*Data-ink maximization* is a principle that pushes the graphic designer to present the largest amount of data with the least amount of ink. Extra ink can be a distraction and take the eye away from seeing the data or making comparisons. This may have its limits in that one should not keep trying to save ink to the point of breaking of Gestalt Laws [Koffka, 1935] (to be covered in the next section).

*Data density* refers to the amount of data elements divided by the area of the graphic. If this is too low the graphic may be reduced in size, or a table may even be more appropriate. Tufte's *small multiples* is a design pattern for comparing data that creates an animation through a series of stills. It states that for a series of graphics the design must remain constant so only the data varies. This should be intuitive, as with scientific experimentation we often hold all variables constant except for the one we are trying to investigate.

[5]While not directly relevant to our purposes this principle is included for completeness of the principles. A generalized rephrasing could be, when appropriate "normalize" data to remove misleading variation.

### 3.2.3 Ware

Ware has created a textbook on information visualization that draws on numerous researchers' theories (including his own) on visual perception and comprehension [Ware, 2004]. In general it is grounded in physiological, perceptual, and cognitive psychology research rather than the more experientially grounded theories of Bertin and Tufte. This research is usually compatible with the previous two's theories and often supports their principles with experimental data from user and electrophysiological studies.

*Preattentive Processing* — Additional visual variables have been shown to be preattentively processed (Bertin's selective category), some examples are: curvature, spatial grouping, blur, added marks, numerosity, flicker, direction of motion, and stereoscopic depth.

*Gestalt Laws* — The German Gestalt school of psychology created a set of fundamental laws of pattern perception (Bertin also refers to Gestalt theory) [Koffka, 1935]. Some of these laws describe how properties such as proximity, similarity, continuity, symmetry, closure, connectedness [Palmer and Rock, 1994], and relative size have major influence on the perception of patterns. They can be used as design principles in creating visualizations [Ware, 2004].

*Words and Images* — Text may often be superior to images for presenting abstract ideas, logic, and conditional information [Ware, 2004]. Consistent with Tufte's Graphical Excellence principle of integrating text descriptions with a graphic, Ware states that the Gestalt Laws (e.g. proximity, or connectedness) apply when adding text.

*Thinking with Visualization* — Ware groups and reviews related research dealing with the problem solving aspects of visualization. Memory categories such as iconic memory, long-term memory, and visual working memory are discussed. Theories on eye movement patterns and cognitive data structures are also presented. The implications of these cognitive constraints on problem solving strategies are also reviewed.

## 3.3 Analysis of Uncertainty Visualizations

In this section we applied the perceptual and cognitive theory outlined in the previous section to eight visualizations which have incorporated uncertainty: vector fields, molecular structure, archaeological reconstructions, 2D stochastic simulation, grid-based annotation lines, particle movement, air traffic control decision support, and surfaces. These eight uncertainty visualizations were chosen to cover a wide variety of domains. The visualizations also vary from highly data specific to more generally applicable. They will be covered in roughly chronological order. Each in turn will be briefly analyzed using the perceptual theories presented by Bertin [1983], Tufte [2001], and Ware [2004]. Our methodology borrows from the ideas of heuristic evaluation [Nielsen and Mack, 1994] as conducted in HCI in which each aspect of a set of heuristics is applied to the interfaces to be analyzed. In this context, a heuristic can be defined as a rule that will in general lead to an improved design.

### 3.3.1 Vector Fields

We will discuss vector field uncertainty glyphs that Wittenbrink et al. [1996] introduced in what they called verity visualizations. Uncertainty glyphs represent uncertainty integrated with the data without the use of additional visual variables (colour, value, ...). This was done with vector glyphs that holistically show uncertainty in magnitude and orientation. An example vector field using the uncertainty glyphs is shown in Figure 3.1.

The authors evaluated their methods using some measures including Tufte's data-ink ratio, as well as performing qualitative evaluation with a user study. Their quantitative analysis found that the mean error for decoding direction with and without uncertainty was not significantly different. This indicates that the addition of their uncertainty visualization was not detrimental to the simpler task which ignored uncertainty. Decoding the magnitude with the presence of the additional uncertainty encoding, was found to be prac-

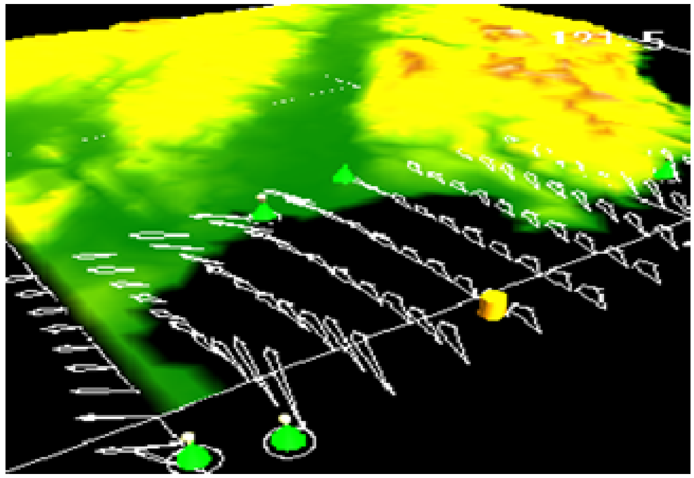


Figure 3.1: Vector field of glyphs showing uncertainty in orientation and area indicating magnitude [Wittenbrink et al., 1996]. [©1996 IEEE]

tically no different, but statistically the errors were not as small. The uncertainty decoding error was comparable to the respective magnitude and direction error.

**Bertin** — Wittenbrink et al. utilized a combination of Bertin's plane, shape, and size variables for their verity glyph for vector uncertainty. To allow users of their visualization to make quantitative estimates their use of the plane and size variables is appropriate, because they are the only variables Bertin claims may be read quantitatively. Showing multivariate data and avoiding using additional visual variables means that the plane and size variables must be overloaded. Decoding these overloaded variables may then be more difficult as the authors discuss.

Bertin's 2D framework may provide insight into potential interpretation problems with 3D viewing. For example, as a vertical line or surface rotates away from the viewer the visible length or area is reduced by the cosine of the angle between the surface normal and the view vector. This directly affects the reading of most variables (to a lesser extent the colour and value variables). Therefore accurate reading of the area glyph presented at varying angles will require more complex cognitive processes dealing with depth perception to compensate for rotation in 3D.

**Tufte** — Wittenbrink et al. [1996] discussed and utilized Tufte's principles. They utilize the data-ink maximization theory to design their uncertainty vector glyphs. In integrating the orientation uncertainty into the glyph they found they had to scale the area to the vector magnitude. This was needed as the orientation uncertainty made the glyph larger and area is perceived over length [Tufte, 2001] (length previously being vector magnitude). This treads on Tufte's integrity principle that the number of information carrying dimensions should not exceed the dimensions of the data (one for magnitude). This type of required trade-off in using these principles is to be expected, but even when not followed, the principles provide a warning to potential areas of misinterpretation.

**Ware** — Ware and Tufte's principle of close integration of text and graphics could be used to provide interactive queries of the glyphs exact values. As the authors were determining how well the new glyphs could be decoded, text was not appropriate, but it could be useful in a final visualization. Gestalt theory also provides a check on the glyph design: symmetry and closure exist with orientation only glyph, but when the magnitude uncertainty is added with an extra leading edge on the arrow head [Wittenbrink et al., 1996] (illustrated in Figure 3.2) it only is perceived as a unit on the basis of proximity. Therefore this perception could become ambiguous when very large magnitude uncertainties exist. A single line from the tip of the arrow to the extra leading edge could provide connectedness to avoid this problem as shown in Figure 3.2. The trade-off is that this reduces Tufte's data-ink ratio.

### 3.3.2 Molecular Structure

Methods for the visualization of molecular positional uncertainty were presented by Rheingans and Joshi [1999]. The uncertainty representations used transparency, volume rendering, and iso-surfaces. Traditional ball and stick models were rendered making dynamic portions of the molecule more transparent. In additional methods presented, Gaussian

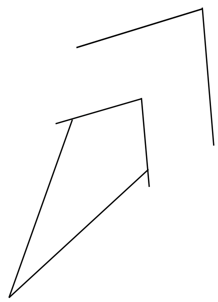


Figure 3.2: Large uncertainty in magnitude showing the weakened proximity gestalt in the glyph on left. Glyph redesign using connectedness to reinforce gestalt on right.

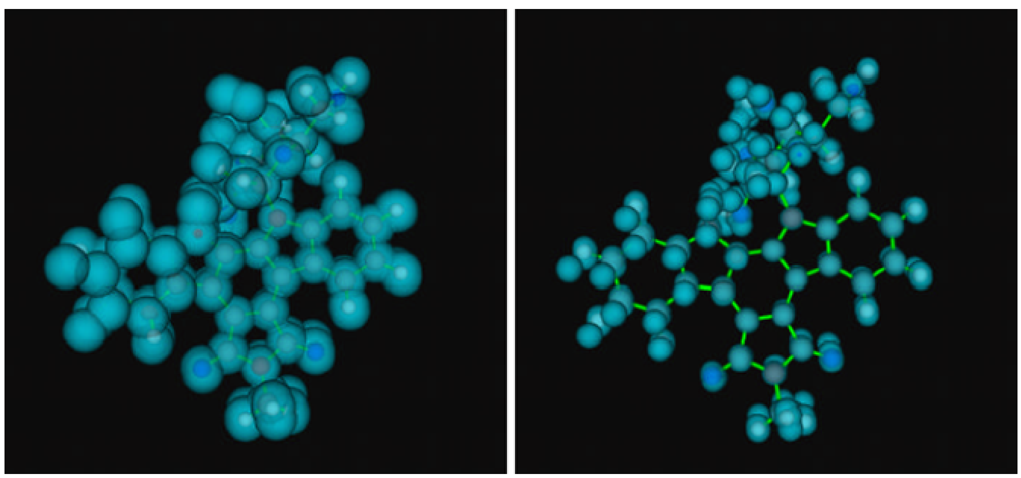


Figure 3.3: Likelihood iso-surfaces. Left and right images show the same data but with different iso-level values [Rheingans and Joshi, 1999]. [©1999 Springer]

distribution functions representing atoms were first aggregated. These were then either directly volume rendered or iso-surfaces were created based on confidence thresholds (iso-levels). Two example visualizations using the iso-surface method are shown in Figure 3.3.

The authors found that their ball and stick and iso-surface visualizations were flexible enough to provide suitable results for different goals. They conclude that the volume rendering method provided a more holistic representation of the uncertainty. A more rigorous task analysis and performance evaluation was not reported.

**Bertin** — Transparency performs a blending of the quantities in Bertin's value and colour variables as in the limit both are reduced to the background instantiation of these variables. Thus using transparency for uncertainty provides a form of redundant encoding (value and colour) of this information and so may be more easily perceived. MacEachren [1995] classifies transparency as an additional variable, and has put it in a subgroup of three "clarity" variables: crispness, resolution, and transparency, that he suggests may be the most useful for encoding uncertainty.[6]

**Tufte** — The data density of these visualizations is high, especially in the volume renderings in which the entire probability distribution is represented. However with the volume rendering identifiable structures became less clear, therefore it could be useful to have the option of integrating text labeling for the atoms, or atom chains (of course layout management may be difficult).

**Ware** — The authors also compare the transparency effect to motion-blur, and as Ware discusses blur is an additional preattentive (Bertin's selective) visual variable. Interaction with the visualizations was not described, and sadly the lack of information on this aspect is not unique to this paper. The authors briefly mention the fact that using the ball and stick model and controlling the opacity allows the image to be perceptually divided into stable and dynamic regions. It is not clearly stated in this paper whether this could be performed interactively as a dynamic query [Ahlberg et al., 1992]. Once generated, the iso-surfaces could likely be interactively rendered, but this precludes dynamic query-like behaviour.

### 3.3.3 Archaeological Reconstructions

Strothotte et al. [1999b] discuss aspects of non-photorealistic rendering and how they

[6] Originally the "clarity" variables were different aspects of what was proposed as a "focus" variable that had edge crispness, fill crispness, resolution, and transparency [MacEachren, 1992]. Kosara et al. [2001] put forward these same characteristics using a photographic metaphor as a variable for focus and context they called "semantic depth of field".

might be applied to representing uncertainty in virtual reconstructions. They show how sketch-like renditions and the use of variable transparency can express the speculative nature of archaeological reconstruction. Figure 3.4 shows some of their results in which a theoretical reconstruction with various levels of uncertainty is integrated into a photograph of the current excavation site. The authors found that photorealistic detail distracts from the fundamental questions of the domain experts. They conclude that more methods of visualization and interaction are required for expressing the appropriate level of uncertainty. No evaluation of their methods was reported. Earlier related work discusses the software that was used in more detail (AncientVis) [Strothotte et al., 1999a].

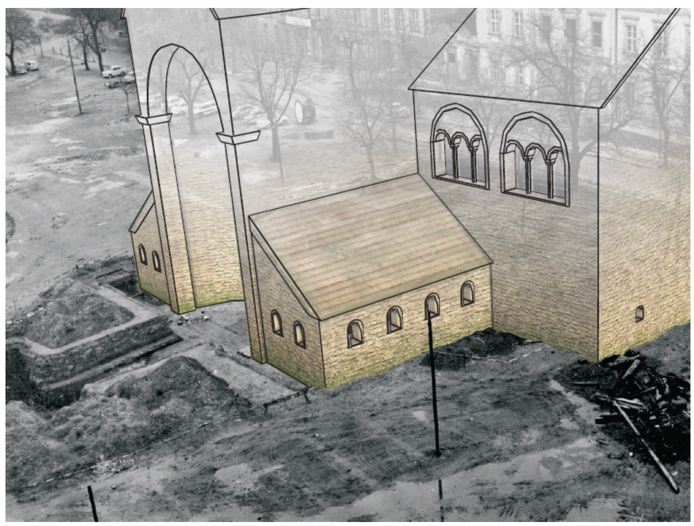


Figure 3.4: Virtual reconstruction using transparency and line drawings to convey uncertainty [Strothotte et al., 1999b]. [©1999 Strothotte et al.]

**Bertin** — As discussed in Section 3.3.2 the integration of Bertin's value and colour concepts in transparency had the potential for effective uncertainty encoding. Bertin states that it is difficult to disregard part of the signifying plane and so an absence of signs indicates absence of data. A line rendering is consistent with this idea, and so is appropriate for the illustration of uncertainty.

**Tufte** — Graphical integrity is applicable to the goals of this type of visualization. The authors point out that the researchers in this domain are very careful to choose verbal descriptions that convey levels of uncertainty. Photorealistic renderings are only potential interpretations of the archaeological data and so using Tufte's Lie Factor [Tufte, 2001], in which a graphic's size should relate only to actual data:

$$K_{lie} = \frac{size(effect_{graphic})}{size(effect_{data})},$$

photorealistic renderings could have potentially huge Lie Factors. Thus portraying the uncertainty is essential to the integrity of the visualization. Line renderings also maximize the data-ink ratio.

**Ware** — Ware's presentation of various cognitive models for objects is applicable as well. Silhouette and contour information may be key aspects used in forming cognitive models [Halverston, 1992, Marr, 1982] and so these may be all that is needed to visually express an interpretation. Perceptual theories more directly related to non-photorealistic rendering can be found in Strothotte and Schlechtweg's [2002] textbook.

### 3.3.4 2D Stochastic Simulation

Various methods for visualizing 2D probability distributions have been presented by Kao et al. [2001]. With their data at each pixel (cell) probability density functions exist based on the different realizations (outcomes) from multiple stochastic simulations. They claim that the spread of a distribution is the most obvious way to summarize uncertainty. Kao et al. provided visual renderings of statistical measures such as mean, median, and quantiles, on a per-pixel basis. These visualizations used colour, surface, and spatial bar charts for presenting various statistical measures and let the users choose the mapping. An example pixel based analysis view is illustrated in Figure 3.5 with the user selected mapping detailed in the annotation. To reduce clutter they provided thresholds for the filtering of insignificant uncertainty representations. A feature-wise analysis tool based on clumps (similarly behaved region) was also described. Kao et al. also present a histogram cube to visualize the distribution of the data. Each histogram bin is represented by a slice for which the pixels contain the counts at the corresponding location. They found it was helpful for understanding the modality of the distributions. The authors conclude that their visualizations were useful based on initial user feedback during the design and development phase, but no formal evaluation was done. Navigation techniques for the 3D view shown in Figure 3.5 were not described and this would be important due to the amount of detail present.

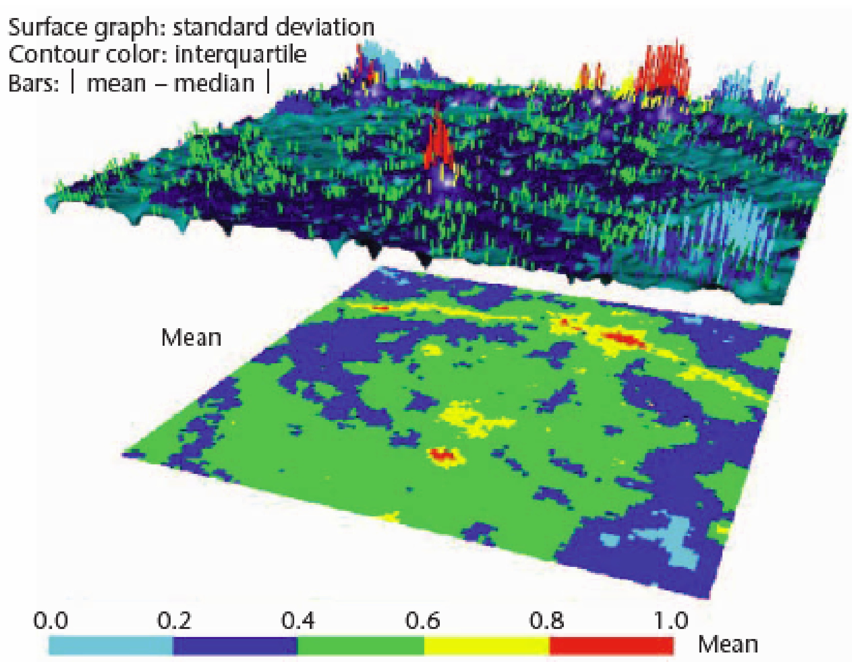



Figure 3.5: Pixel-wise analysis of data distributions [Kao et al., 2001]. The upper surface is deformed by the standard deviation field, coloured based on interquartile range, and has vertical bars indicating the absolute value of the difference between mean and median fields, coloured the same as the lower plane. [©2001 IEEE]

**Bertin** — The use of only the colour variable which is unordered is not helpful for numeric data (as used in Figure 3.5), however, colourmaps were changeable. For spatial data the use of a colourmap that varies in both value and colour still leaves size, grain, orientation, and shape variables for additional information. These variables may be easier to cognitively integrate than the additional 3D surface (that uses the plane, size, and colour variables).

**Tufte** — It may be worth considering Tufte's integrity principle: do not show data out of context. Complex classifications that do not reflect topographical or other known spatial distributions will be difficult to cognitively integrate into the correct spatial context. Text annotations or symbol landmarks could help with this integration by labeling extrema (on low pass filtered data) and showing these same landmarks on an adjacent terrain map.

**Ware** — As can be seen in Figure 3.5 the implementation allows the two representations to be viewed simultaneously. While the small multiples design pattern is not directly applicable, if the goal is to understand relationships then orthographic projection would maintain size consistency and simplify cognitive integration. Ware discusses various issues relating to context and cognitive integration. Numerous other aspects relating to navigation and maps would be applicable, such as Mackinlay et al. [1990] point of interest navigation.

### 3.3.5 Grid-based Annotation Lines

Cedilnik and Rheingans [2000] have presented procedural rendering of annotation overlays that indicate uncertainty. The authors show how procedural variation of width, sharpness, noise, and amplitude modulation can indicate uncertainty. The illustration of uncertainty only on the annotation (grid) lines allows the data to remain largely unobscured, as can be seen in Figure 3.6. The authors state that their method preserves perceptibility across various levels of uncertainty. No formal evaluation was reported.

**Bertin** — Bertin's variables of the plane are mainly used for the amplitude modulation although it crosses into use of grain. The size and value variables are used for the width and sharpness techniques. The noise-based annotation, which was made up of distributed spot noise rather than a continuous line, is more ambiguous as it has aspects of size, value,

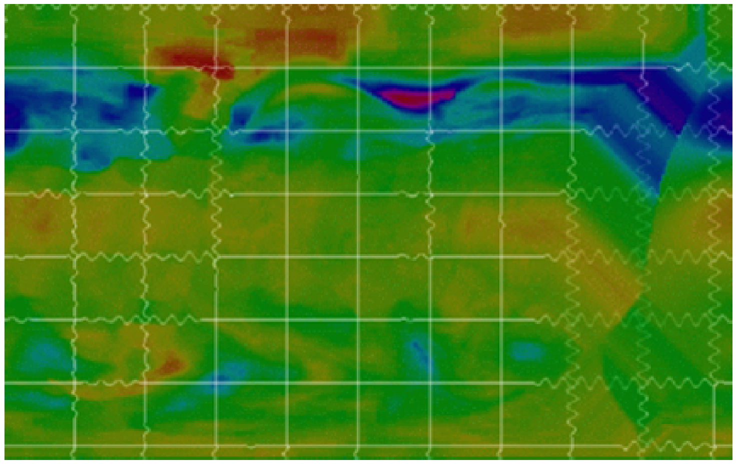


Figure 3.6: Procedural grid overlay with sine wave amplitude modulation indicating uncertainty of data at that location [Cedilnik and Rheingans, 2000]. [©2000 IEEE]

and grain. From this one would expect the largest number of levels of uncertainty would be discernible with the amplitude modulation.

**Tufte** — Tufte's data-ink maximization rule would suggest that the amplitude modulation would also be the best of their methods. The data-ink ratio in the noise based method could easily be increased by only showing random points along the maximum displacement, but this would violate the authors' energy conservation scheme in which perceptibility (via overall intensity) was preserved.

**Ware** — The authors state for all methods they attempt to perceptually normalize the amount of energy present at every place. For energy conservation they integrate an annotation intensity value for normalization. However they map it to saturation from HSV space, which Ware points out, is only crudely linear in perceptual space. This approximation is a trade-off that must be made against run-time speed. Therefore they conserve perceptual energy by trying to transfer perceptibility from the reading of Bertin's value and colour variables to the size variable (assuming Bertin's line implantations). It would be interesting to more formally evaluate how well this works. Ware also points out that size of an object impacts the perception of colour so this is a difficult perceptual balancing act.

### 3.3.6 Particle Movement

Methods for the visualization of uncertain 2D and 3D particle movement over time were presented by Lodha et al. [2002b]. Size (spheres), transparency, and colour were used to visualize the resultant probability distributions. They found that the resulting visualizations could often be categorized by form as can be seen in the Figure 3.7. The addition of colour to transparency was found to better delineate high probability regions. The authors found that their algorithm and subsequent visualizations were useful for understanding probabilistic movement and distributions. No formal evaluations were discussed.

**Bertin** — The authors found that the combined use of transparency and colour more clearly showed the high density regions in the center. The addition of the colour variable adds length beyond the range of perceptual steps available from transparency which is already a value and colour hybrid[7].

**Tufte** — A rule of graphical excellence suggests close integration of statistical and verbal descriptions of the data. It would likely be beneficial to add numeric, textual, or graphic (principle component axes) annotation directly on the visualization. This would be especially true in 3D.

**Ware** — Again Gestalt theory [Koffka, 1935] comes into play for the perception of shapes. Analysis of these laws may provide the validation that the shape being perceived is capturing all the relevant aspects of these probability distributions. The authors found that adding colour helped to understand regions in the uncertainty, as it created clearly separable regions within high probability areas that were not distinguished only on the basis of levels of transparency (possibly due to insufficient length in that variable). Even before adding pseudo-colours, transparency effectively acted as a colour saturation variable due

[7] When the interaction of a combination of variables is perceptually non-linear the addition of a separable variable may provide assistance in perceiving a threshold in a desired range.

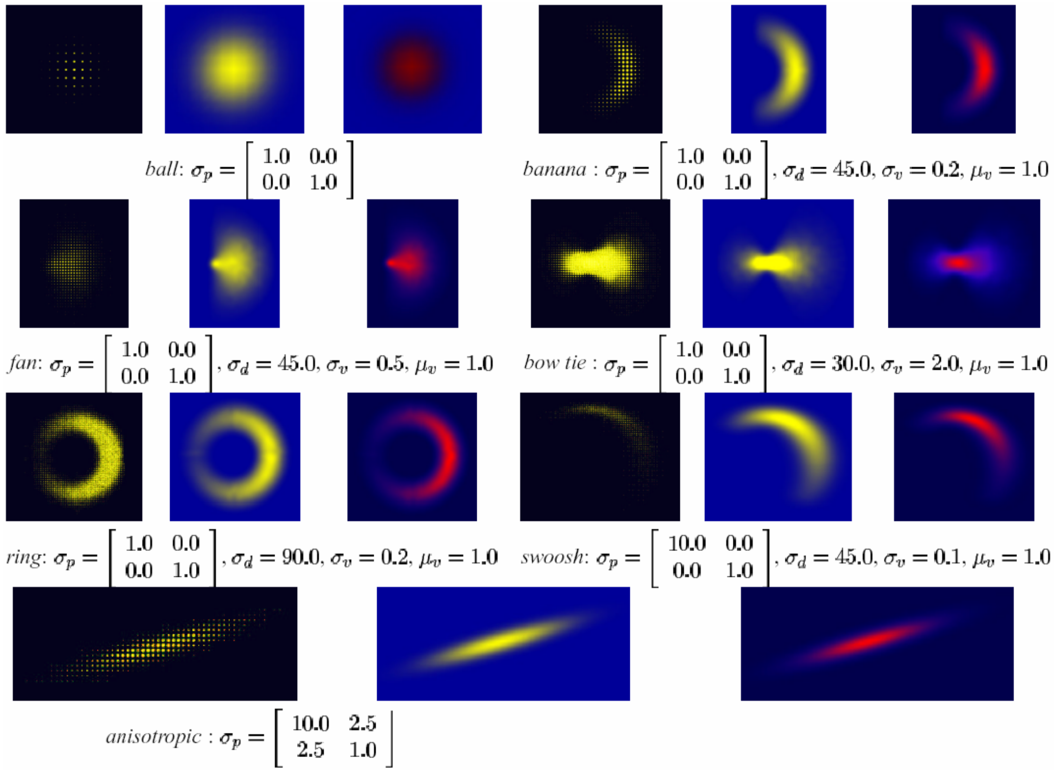



Figure 3.7: Seven different 2D probability clouds. Three different representations for each cloud shown in subimages from left to right: spheres, transparency, and transparency and colour (the colour differentiates high opacity into 2 regions). Cloud forms are assigned shape names such as ball, banana, fan... [Lodha et al., 2002b]. [©2002 Lodha et al.]

to the background colour. Other colour sequences could be used to more clearly delineate more than two probability regions. Ware presents various research that may be useful in this area. Gray scales (value) do provide the highest spatial frequency sensitivity [Ware, 2004], and so could be valuable if spatial frequency content is high. However, Ware [1988] has shown that errors in reading gray and saturation scales can be as large as 20% of the scale.

### 3.3.7 Air Traffic Flow Decision Support

In quite a different domain, Masalonis et al. [2004] discuss the visualization of uncertainty in air traffic flow management. In this domain the user's task is making decisions based on probabilistic alert levels. They have a discrete probability density function representation of the uncertainty and looked at a design option for providing it in a more detailed drill-down view. The authors performed a qualitative task analysis for which they carried out a user study. The user study covered various aspects of the cognitive issues related to operational needs of the uncertainty display. They then proposed multiple views that have various levels of detail and meta-data related to the uncertainty modelling. One of these views relating to alert likelihood monitoring is shown in Figure 3.8. As work was still in the design phase they did not get to the point of evaluating their proposed designs or prototypes.

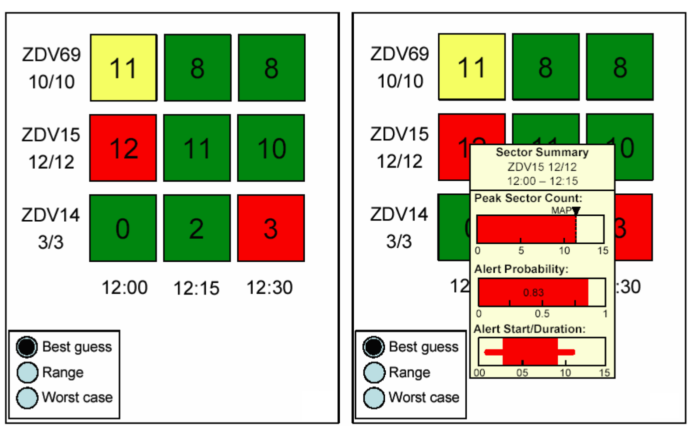



Figure 3.8: Mock-up of alert display with colours indicating probability of exceeding an alert threshold. Left to right images illustrate result of mouse rollover, or hover query [Masalonis et al., 2004]. [©2004 MITRE Corporation]

**Bertin** — As Bertin's colour variable is used to show the probability of an alert (3 levels are used: green, yellow, and red), using this variable there is enough length for even more alert levels. Bertin states that colour has no perceptual order, so it may not be appropriate for the ordered levels, but the symbolic reading of the colours (i.e. stop, caution, go) provides an order. Similarly colour saturation (and value) could provide more levels within each of the three probability regions if needed.

**Tufte** — On the overall display, Tufte's data density measure appears to be quite low. This might suggest that more data could be presented in the Overview Display (not shown in figure). The numbers could possibly be removed completely if more colour levels were utilized. This might also change the scanning strategy if users tried to anticipate the alert changes using the numbers (i.e. green changing to yellow).

**Ware** — The choice of pure red and green colours excludes a large number of colour blind people from performing the task. Around 10% of the general male population and 1% of the female population are colour deficient [Ware, 2004], with red-green being the most common. Ware also discusses that large regions of colour should use low saturated colours to avoid visual stress. Therefore depending on the size of the display (a prototype had an 11x12 matrix of cells similar to those in the mockup in Figure 3.8) the green and yellow colours should be very low saturation. Another of Ware's colour design guidelines is that a text to background luminance ratio of 10:1 is preferred (3:1 is the ISO 9241 part 3 minimum recommendation [Ware, 2004]). The luminance ratio of the black text on green would also increase after replacement with lower saturation colours, making the black numbers more easy to read.

As this visualization involves a visual monitoring task, Ware's coverage of attention and scanning strategies theory such as Wickens [1992] should be useful. Motion and flicker are visual variables that extend further in the user's useful field of view. Depending on the final display size they could be used to help avoid missing significant uncertainty changes [Ware, 2004]. Charbonnell et al. [1968] and Sheridan [1972] have proposed that monitoring behaviour is controlled by growth of uncertainty in a channel and the cost of sampling a channel. Prolonged viewing in the case of monitoring may also lead to over polling of low frequency data [Moray, 1981]. Implications from other monitoring research are also discussed [Moray and Rotenberg, 1989, Russo and Rosen, 1975]. Interaction issues surrounding the use of hover queries are summarized by Ware, such as Rutkowski's [1982] principle of transparency in which the tool itself disappears and one can focus single-mindedly on the task. All this suggests that there might be alternative visualizations to help with the monitoring nature of the task.

### 3.3.8 Surfaces

Grigoryan and Rheingans [2004] have shown how points and lines can be used to represent uncertainty in a 3D surfaces position. Starting with a surface segmented from medical data, a large number of points are pseudo-randomly displaced along the surface normals according to the uncertainty. An example of their visualization method using a tumour segmented from an MRI scan is provided in Figure 3.9. Lines can also be drawn from the zero displacement surface to the point. Their method supports both a uniform distribution or when available a probability density function (PDF) based distribution. Results from a preliminary evaluation comparing their point displacement to a pseudo-colour representation were reported. The task was determining if an object was within a specific error margin around a surface. The point-based scheme showed an average increase in accuracy of 20% ($p < 0.01$) and made judgments faster, although this had lower statistical significance ($p < 0.1$). Subjective ratings of ease, confidence, and satisfaction were all also higher for the point-based representation ($p < 0.01$) over pseudo-colouring.

**Bertin** — The displacement of points in the plane and the use of size provide the quantitative aspects required for this application domain. They also tried using a neutral colour and transparency to encode uncertainty. As the length of the colour variable is small compared to the plane and size it was appropriate that it was only used as binary threshold to switch from a specific colour to a neutral colour (e.g. gray) based on the uncertainty.

The accuracy and speed results from the user testing are predicted from Bertin's theory. As the objects' spatial extents are represented in the plane, if the uncertainty is also

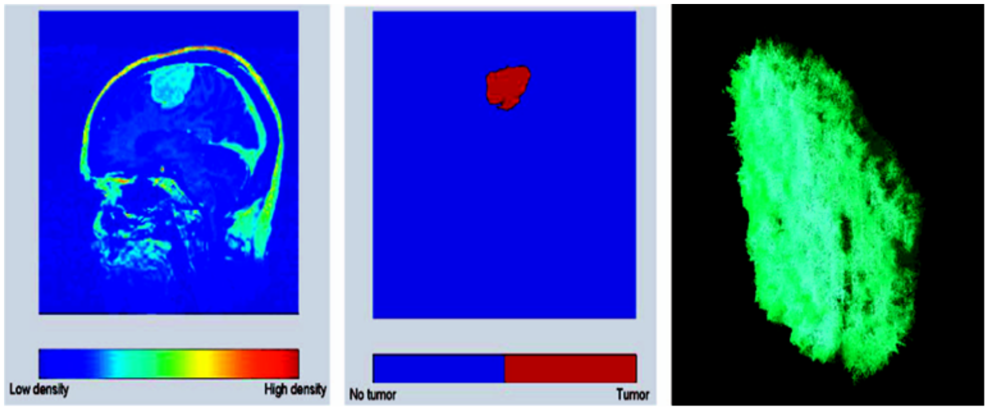



Figure 3.9: Images from left to right: MRI scan, segmented tumour, tumour surface in which points are displaced along the surface normal based on an uncertainty distribution [Grigoryan and Rheingans, 2004]. [©2004 IEEE]

represented in the plane, direct reading can provide the level of uncertainty (the plane also allows quantitative reading). The authors describe a user controlled scaling factor for the displacement, but if this does not match the domain of the PDF it violates a uniformity in the interpretation of space. The statement from Bertin is that certainty of the uniformity of the plane entails a presumption of uniformity in the conventions adopted within the signifying space [Bertin, 1983]. While Bertin was describing 2D representations, this may still be valid in 3D (i.e. a uniformity of 3D space), especially when considering any cross-sectional slice plane through a 3D volume. Thus a violation could occur as the point displacements and the zero displacement surfaces are both represented in the plane, and the non-uniformity could lead to misinterpretation.

**Tufte** — Graphical integrity is clearly in question with the presentation of an uncertain surface as a clean, precise, polygonal surface. This is why the authors have attempted to build a more imprecise, and thus more accurate, representation. The user controlled scaling factor mentioned in the previous sub-section could also relate to a potential increase in the Lie Factor. As the rendering was interactive it could benefit from additional text annotation, perhaps based on a user controlled probe. Another guideline for excellence that might be applied is to reveal the data on several levels of detail. As the point based display is full of fine detail the user should be able to toggle it with the uncertainty free surface, or provide a mouse draggable inset for this level of detail.

**Ware** — While the use of transparency has good properties for representing uncertainty, Ware reviews its limitations. The use of lighting along with colour variation (the more general definition of colour) may also be problematic as this overloads the value and colour variables to the point of potential misinterpretation. Value (luminance) therefore should not be part of the colour variation, and this was not explicitly stated by the authors. As the visualization was interactive it is assumed the user could manipulate the viewpoint or object. Therefore understanding the context of the data is important. In cases such as the tumour dataset, understanding the uncertainty in relation to the surrounding tissue is of vital importance. This could be done by merging the visualization with an interactive slice planes from the original volumetric scan data.

## 3.4 Summary of Bertin, Tufte, and Ware's Heuristics

The theories provided by Bertin, Tufte, and Ware were relevant to all of these uncertainty visualizations. It is important to note that these visualizations were chosen as a representative cross-sectional sample of uncertainty visualization before selecting the three author's theories, and so this choice was not in any way based on the potential applicability of these theories. Most of the visualizations did not mention these theories, and Wittenbrink et al. [1996] was the exception which explicitly utilized Tufte's and others' theories, such as Carswell [1992] and Cleveland [1985], to refine and analyze their solution.

Analysis using these theories can be considered as a form of heuristic evaluation [Nielsen and Mack, 1994, Shneiderman, 1987]. We summarize a subset of the applied theories, in the form of possible heuristics, in Table 3.1. These heuristics are extreme simplifications, but still, the application of them may raise important issues. The relevance count in the table is only provided to summarize applicability for the eight visualizations reviewed and not intended to imply the relative generality. Later, in Section 3.5 we will discuss applying these perceptual and cognitive heuristics to more visualizations to better determine their generality and usefulness.

Table 3.1: Potential Heuristics. The "Heuristic" column presents the simplified forms of the theory. The "Relevant" column indicates the total number of visualizations, of the 8 just reviewed, for which the heuristic was pertinent.

| **Heuristic** | **Source** | **Relevant** (n/8) |
|---|---|---|
| Ensure visual variable has sufficient length | Bertin & Ware | 7 |
| Preserve data to graphic dimensionality | Tufte & Bertin | 2 |
| Put the most data in the least space | Tufte | 2 |
| Provide multiple levels of detail | Tufte & Ware | 2 |
| Remove the extraneous (ink) | Tufte | 4 |
| Consider Gestalt Laws | Ware | 2 |
| Integrate text wherever relevant | Tufte & Ware | 6 |
| Don't expect a reading order from colour | Bertin & Ware | 2 / 6† |
| Colour perception varies with size of coloured item | Ware & Bertin | 2 |
| Local contrast affects colour & gray perception | Ware | 2 |
| Consider people with colour blindness | Ware | 2 |
| Preattentive benefits increase with field of view | Bertin & Ware | 3 |
| Quantitative assessment‡ requires position or size variation | Bertin | 4 |

† Counting aspects beyond the uncertainty components, including those that were not adequately described.
‡ Perceiving an accurate approximation of the ratio between two signs or grouping of homogeneous signs [Bertin, 1983].

Often the authors stated that future work would be in evaluating their new methods in the form of user studies, and Wittenbrink et al. [1996] and Grigoryan and Rheingans [2004] did perform and report their evaluation results. The Masalonis et al. [2004] research also reported analysis from a user study done during the initial stages (task analysis & design) of creating a visualization. The use of studies at the design phase is important and we will provide our study from this phase in Chapter 7.

The amount of work involved in evaluation often forces the two part presentation of research: development and then evaluation. Obviously it is the second part that may not get done, and even when performed may not make it into publication. This suggests that potentially more light-weight evaluations, in a manner similar to what we have done here, could more often be included in current work. The need for greater application of human factors research to visualization has also recently been noted by Tory and Möller [2004]. Following this lead further, we should examine the cognitive psychology literature dealing with uncertainty and probabilistic reasoning (such as Kahneman et al. [1982], Gilovich et al. [2003], Sloman et al. [2003], Kirschenbaum and Arruda [1994], and Finger and Bisantz [2002]); as the uncertainty, if correctly understood, must then be integrated into a decision process [Watkins, 2000]. This decision process adds cognitive load, which may restrict the resources available for the visualization process. Developing this further is the content of the next chapter.

Uncertainty visualization should not be considered unique; we expect the theories would be similarly relevant to most other visualization problems. While a few of the uncertainty paper authors discussed and applied the theories, it appears that they have been under utilized. We would even suggest that detailed analysis from their perspectives should be more strongly influencing the work in the field of visualization. Recent research continues to develop new frameworks, such as Amar and Stasko's [2004] knowledge task-based framework for design and evaluation of information visualizations. In the next section we will discuss how this framework and other theory might complement the lower level perceptual and cognitive theories we used for analysis of the uncertainty visualizations.

## 3.5 Heuristic Evaluation of Visualizations

Heuristic evaluation is a discount evaluation method commonly used to find usability problems at different development stages of a product. A heuristic evaluation involves a small number of evaluators inspecting a system according to heuristics or guidelines that are relevant for the system. Heuristics exist as shared or general knowledge on design. They often can act as instructional guides for the teaching of novices and can evolve into design patterns for construction such as those that exist for software engineering. They aid in the communication of ideas by providing a common language and promote reuse of proven methods or concepts [Gamma et al., 1994]. Other heuristics can be more general and act as a check on design choices. As heuristic evaluation is a light-weight process that can be cheap, fast, and easy to apply, it has potential for integration within development iterations. It can be used both in design and evaluation phases of development and can even be applied to paper-based designs before the first working prototype is created.

While heuristic evaluation has been part of the HCI set of evaluation tools for some time [Nielsen and Mack, 1994], it has not been utilized or examined for evaluating visualizations to the same extent. Granted usability issues also arise in these systems, but they are not the only problems that these systems may have. We discuss issues that call for different or supplemental sets of heuristics for the discount evaluation of visualization systems. Utilizing a few sets of previously published design principles (advice) for visualization we create a possible set of heuristics for evaluation. Using these heuristics we analyze *LuMPB Key* (Landscape unit Mountain Pine Beetle Key [Schlesier et al., 2006]), a visual decision support system that is used to examine simulation data, as a case study to demonstrate their application. We assess the value of the suggested heuristics by applying them to *LuMPB Key* and discuss implications for further research of the process of heuristic evaluation of visualizations.

## 3.6 Determining a Set of Heuristics for Evaluation

The field of information visualization is influenced by many different research domains including psychology, semiotics, graphic design, and art. The goal of an information visualization is generally defined as providing useful tools and techniques for gaining insight and understanding of a dataset, or more generally to amplify cognition [Card et al., 1999]. These are high-level cognitive issues that are hard to measure with quantitative user studies. Tory and Möller [2005] in their summary of expert reviews recommend the use of heuristic evaluation for analyzing visualization systems. While usability heuristics, as known from HCI, encompass a wide variety of issues pertaining to visualizations and the interaction with them, we have found that more specific heuristics are of value, in particular since a wide variety of research fields are concerned.

Previous evaluations in InfoVis have proposed heuristics specific to a certain data domain, e.g. for ambient displays [Mankoff et al., 2003] or multiple view visualizations [Baldonado et al., 2000], for a specific cognitive level based on knowledge and task [Amar and Stasko, 2004], or based on perception and cognition [Zuk and Carpendale, 2006]. Shneiderman's [1996] well known "Visual Information-Seeking Mantra" has also been used for heuristic evaluation of information visualizations based more on task and usability (for an overview see Craft and Cairns [2005]). Tory and Möller [2005] propose to use heuristics based on both visualization guidelines and usability. These all have their own list of heuristics. Although there are several lists of usability heuristics which do apply to visualization tools (not just to the user interface) [Tognazzini, 2006, Nielsen and Mack, 1994, Kahn and Prail, 1994], there are fewer that are specifically tailored to them [Amar and Stasko, 2004, Zuk and Carpendale, 2006, Shneiderman, 1987].

At this stage of development of heuristics for visualization we have reached a similar problem as described by Nielsen and Mack [1994]. It is a difficult problem to assess which list(s) are better for what reasons and under what conditions. This leads to the challenge of developing a “top ten” list that comprises the most important or common visualization problems, or alternatively a series of lists for specific purposes. Visual representation, presentation, and interaction and manipulation of the parameters that build a visualization play a role in the success or failure of the overall high-level goal to amplify cognition. The above mentioned evaluations used different heuristics and methods to evaluate their criteria. They also suggest that data or visualization types and domain specific information processing tasks are a factor for the evaluation of visualization systems. Whether it will be possible to find a small set of heuristics that find the most common visualization problems, similar to Nielsen and Mack’s [1994], is an exciting open problem for the community. How to decide the optimal or even appropriate heuristics is the question.

A hierarchical or taxonomic way of grouping may aid in selecting an appropriate set of heuristics. A tree-traversal-like approach could be used in which a depth-first search is performed with pruning occurring if the more general heuristics are not appropriate. Morse et al. [2000] also pruned an extensive task taxonomy to create a test set (for evaluation-question generation) using the rationale “sample as broadly as possible rather than deeply, and select those which varied significantly” [p. 644]. This organization could lead from a more general heuristic, such as *consider the implications of colour encoding*, to more specific heuristics such as *colour perception varies with size of coloured item* [Ware, 2004], or *don’t expect a reading order from colour* [Bertin, 1983]. The heuristics at the leaf level would likely be “chunked” by experts so that they only need to descend to the more general heuristics to trigger the set of considerations they feel appropriate, but would serve a teaching role to novices. One such possible tree organization is shown in Figure 3.10.

Another approach is to empirically determine a *minimal* set of heuristics. Nielsen and Mack [1994] describes a method of refinement of a large set of usability problems into a small set of 10 heuristics that are intended to be general and easily understandable. At this initial exploration stage, however, we will only probe some potential heuristics to estimate their applicability.

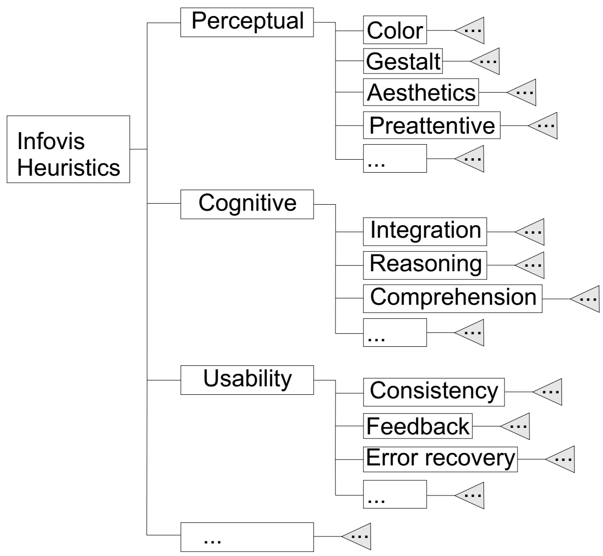



Figure 3.10: Evaluation tree for hierarchically organizing heuristics.

## 3.7 Determining a Process for Heuristic Evaluation

The process of heuristic evaluation may evolve just as the heuristics themselves can evolve over time. The original presentation of heuristic evaluation for usability proposed at least two passes of an interface: the first pass to provide a general feel, and the second pass for the application of all heuristics to each interface item [Nielsen, 1994]. While we initially want to learn from accepted practices we do not want to limit ourselves to that process as the nature of the problem is in some ways fundamentally different. Usability mainly deals with interaction which is only a single, but important, component of visualization. Visualization and uncertainty visualization bring to the table numerous perceptual and comprehension issues beyond usability.

HCI studies showed that using five people as evaluators may be enough to find most usability problems, adding more would reduce the benefit to cost ratio, and suggested that three may suffice [Nielsen, 1994]. More recently Spool and Schroeder [2001], and a CHI conference panel [Bevan et al., 2003] reviewed how many evaluators are required for web site usability analysis. They found for some problems more than five are likely needed to find the majority of problems, and the exact number will likely be product specific. Because the use of heuristics in visualization evaluation has not yet been fully studied, it is still uncertain if this knowledge will transfer. We can only suspect that for evaluating information visualizations, the required number of evaluators to guarantee finding most problems may also be visualization specific. In heuristic evaluation for usability, as performed in HCI, the evaluators are commonly usability specialists. It still has to be determined, however, what is required of an "visualization specialist" when applying a heuristic evaluation. Tory and Möller [2005] suggest using both visualization (data display) and usability experts. What knowledge is required of a "visualization specialist" will have to be discovered. We would also suggest a domain expert should likely be involved whenever tacit knowledge is required.

While evidence has shown that a small set of heuristics can find a majority of basic usability problems with specific applications [Nielsen, 1993], we as yet have no evidence for a similar potential from visualization heuristics. Craft and Cairns [2005] recently undertook the process of analyzing the heuristics of the "Visual Information-Seeking Mantra". They reviewed others' use of the "Mantra" and found a lack of empirical evidence validating the heuristics. They noted that even though the heuristics were presented as descriptive in nature they have been used prescriptively [Craft and Cairns, 2005]. They conclude by calling for a more rigorous design methodology that: takes into account the useful techniques that guidelines and patterns suggest, has measurable validity, is based upon a user-centered development framework, provides step-by-step approach, and is useful for both novices and experts.

Kahn and Prail [1994] have provided a set of design heuristics to help design the evaluation process itself. These are: minimize time cost to engineers who are on the critical path, maximize involvement of engineers who will implement changes, create a method that is an "event" in the usability life-cycle, team-based approach, adapt existing methods (i.e. help do what is done better), leverage the language and structure of well-established methods solving similar problems, task orientation, and clear potential integration with other parts of the usability engineering life-cycle. There may be a danger in assuming too much in reusing the process of heuristic evaluation from usability for more general visualization evaluation, as perceptual and cognitive issues (e.g. domain knowledge) are more internalized and may confound this style of evaluation. Therefore we should consider using Kahn and Prail's [1994] or other process heuristics to re-evaluate the process in its application to information visualizations. To further explore aspects of both process and heuristic selection, the next section describes a case study in which we heuristically evaluate a visual decision support tool and provides a meta-analysis of the results.

## 3.8 Case Study: The *LuMPB Key* System

In order to study the understandability and applicability of a set of heuristics and explore a methodology, we performed a heuristic evaluation of a visualization of simulation data for measuring the impact of mountain pine beetles (MPB) on forests.

### 3.8.1 Method

Our method involved applying three different and distinct sets of heuristics to a single visualization, then analyzing the evaluation results individually, followed by a discussion between all evaluators. The discussion included both an analysis of the individual findings and a meta-analysis of the heuristics and process. The discussion was based on the specific findings, but actively considered the ability to generalize. Rather than making considerations for pursuing a high-quality evaluation (high percentage of all problems found), our methodology was chosen to support the meta-analysis.

### Evaluators

Four computer science graduate students in the Interactions Lab at the University of Calgary each independently performed a heuristic evaluation of a single visualization in the *LuMPB Key* tool. One student was the developer of the visualization tool, two were Ph.D. students in information visualization and the fourth was a Ph.D. student in human-computer interaction. Being one of the evaluators I had experience applying the heuristics listed in Table 3.1. Note that these evaluators were chosen for the purpose of generating valuable discussion in the meta-evaluation and not to appropriately evaluate this specific system.

### System

The *LuMPB Key* simulation tool [Schlesier et al., 2006] can be used to visualize complex simulation data created with the Spatially Explicit Landscape Event Simulator (SELES) [Fall and Fall, 1996, 2001]. In these simulations mountain pine beetle impact on forest is observed for various conditions. One goal of the simulations is to see which forest management strategy is best to protect pine trees under the particular conditions in each specific forest management region. *LuMPB Key* was created to assist with the uncertainty in reasoning around forestry management decisions.

One of the sets of views that *LuMPB Key* provides is shown in Figure 3.11 A stacked bar chart is used to display the relative proportions of tree types (e.g. amount of cumulative logged pine trees) in the forest over different management scenarios for a given year (upper left part of Figure 3.11). Bar charts are used to display a single tree type over management scenarios for a given year (lower left area of Figure 3.11), or to show a time series for a tree type for one or more scenarios (lower right area of Figure 3.11). Furthermore, text describing management scenarios or tree types can be brought on to the screen. The visualizations we analyzed were the two views on the left side in Figure 3.11. The user has the ability to swap the positions to bring either one into focus (in order to get more

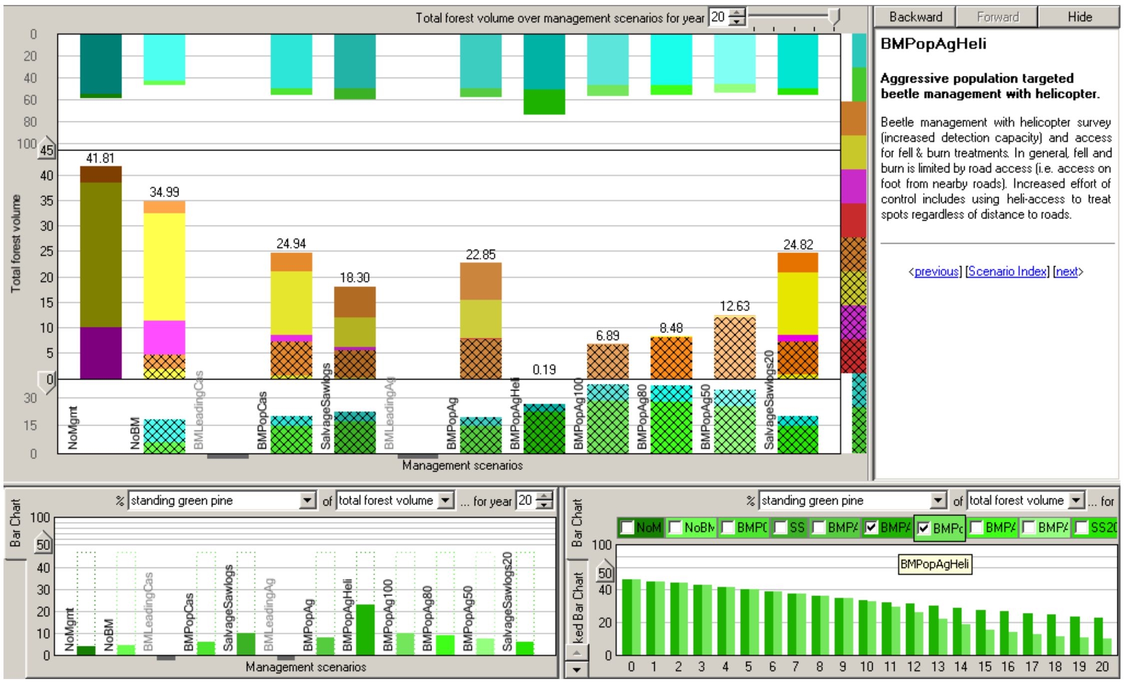



Figure 3.11: Screen shot of the *LuMPB Key* tool. Shows different views on a data set and a text view providing context information about a management scenario.

details).

### Procedure

Each evaluator was asked to identify both positive and negative aspects of two specific visualizations (views) in the LuMBP Key system, based on three sets of heuristics. The first set of heuristics were Zuk and Carpendale's [2006] selection of perceptual and cognitive heuristics. These heuristics were chosen because they were designed to be used as heuristics for evaluation, and have been used for this purpose in practice. Shneiderman's [1996] "Visual Information-Seeking Mantra" were chosen as the second because they too have been used to evaluate information visualizations, even though they were not designed this way. Amar and Stasko's [2004] knowledge and task-based framework was chosen as the third set because they were designed to be used to evaluate (and design) information

Table 3.2: Heuristics applied in evaluation of *LuMPB Key*

| Zuk and Carpendale's [2006] *Selection of perceptual and cognitive heuristics* |
|---|
| Ensure visual variable has sufficient length [Bertin, 1983, Ware, 2004] |
| Don't expect a reading order from colour Bertin [1983], Ware [2004] |
| Colour perception varies with size of coloured item [Ware, 2004, Bertin, 1983] |
| Local contrast affects colour & gray perception [Ware, 2004] |
| Consider people with colour blindness [Ware, 2004, Tognazzini, 2006] |
| Preattentive benefits increase with field of view [Bertin, 1983, Ware, 2004, Healey, 1998] |
| Quantitative assessment requires position or size variation [Bertin, 1983] |
| Preserve data to graphic dimensionality [Bertin, 1983, Tufte, 2001] |
| Put the most data in the least space [Tufte, 2001] |
| Remove the extraneous (ink) [Tufte, 2001] |
| Consider Gestalt Laws [Ware, 2004] |
| Provide multiple levels of detail [Tufte, 2001, Ware, 2004] |
| Integrate text wherever relevant [Tufte, 2001, Ware, 2004] |

| Shneiderman's [1996] *"Visual Information-Seeking Mantra"* |
|---|
| Overview first |
| Zoom and filter |
| Details on demand |
| Relate |
| Extract |
| History |

| Amar and Stasko's [2004] *Knowledge and task-based framework* |
|---|
| Expose uncertainty |
| Concretize relationships |
| Determination of Domain Parameters |
| Multivariate Explanation |
| Formulate cause & effect |
| Confirm Hypotheses |

visualizations, but (to our knowledge) evidence for their use in evaluation has not been published. The heuristics are listed in Table 3.2; detailed descriptions are available in the original papers. Each set of heuristics was to be considered separately in the order shown in Table 3.2.

### 3.8.2 Discussion

Our analysis (meta-analysis) was performed by reviewing as a group all of the individual evaluation results. We proceeded through the heuristics in the order that they were applied looking for commonality, discussing problems found, problem solutions, and to a lesser degree positive findings. At a higher level we also discussed problems and generalizations and what could be improved in the heuristics and the evaluation process.

#### Heuristics for Communicating Patterns

One aspect of heuristics as design patterns is the communication of ideas. However, we found there existed a variety of interpretations of the heuristics across the four evaluators. Placing Bertin's definitions in the perceptual-based heuristics was particularly problematic, as the strict separation of perception from cognition and/or symbolism was not usually maintained. While the heuristics were described in more detail in Zuk and Carpendale [2006], Shneiderman [1996],and Amar and Stasko [2004], only the summary heuristic was provided as a cue for the evaluation. As the heuristics will likely evolve along with the considerations they evoke, tying a concise description to a heuristic will be helpful. Creating consistency of definitions across the community of practice would help in general usefulness and in the possibility of meta-comparisons. This will also aid in the communication and transfer of knowledge from the findings.

The generally high specificity of Zuk and Carpendale's heuristics was also called into question. Loosely defined terms and more general wording in a heuristic may allow the flexibility in interpretation needed to catch a broader range of related problems. For example, the "preattentive benefits increase with field of view" heuristic was considered too narrow, with a potential replacement being "use preattentive visual variables wisely".

#### Redundancy

The three different groups of heuristics did at times find the same problem from different perspectives. If the main goal of the heuristics is to identify problems then redundant coverage goes against the goal of a minimal set of heuristics. However, if the intention is to also indicate possible solutions to the problems, then finding the same problem via different heuristics can suggest different solutions. Instead of redundancy we can consider that heuristics may support each other by revealing the same problem from different standpoints. In our case study, *details on demand* and *integrate text where relevant* are an example where two heuristics pointed out the same problem and the same solution. Both revealed that tool tips could be used to display the mean values and standard deviations in the stack bar chart.

**Conflicting Heuristics**

Heuristics, especially from different sets, may also in some ways contradict each other. This leads to the consideration of trade-offs in the design and it needs to be determined which heuristic has a higher priority. Stakeholders (commonly the domain experts) may also have the right to override heuristics based on domain knowledge or other constraints. For example, colours for the stacked bar chart in the evaluated system were chosen by the domain experts to reflect common usage, and could therefore not be changed to account for colour-blindness. This domain-dependent weighting of heuristics also creates the variability which adds difficulty in producing a minimal set.

**Heuristic Taxonomy**

Our case study was a preliminary exploration of how we might develop a set of appropriate heuristics for evaluation of visualizations, including uncertainty visualizations. We are not yet at the stage of producing a taxonomy, but our combined evaluations led to a discussion of how best to organize the heuristics to provide experts with an improved structuring of potential problems to look for. One suggested categorization was to organize the heuristics according to their applicability to *perception*, *usability*, and *discovery process*. In particular, we found it useful to think of the *LuMPB Key* system by separating our criticism into these three aspects. Specifically, Zuk and Carpendale's [2006] were most useful for evaluating perception, Shneiderman's [1996] heuristics were most useful for evaluating usability, and Amar and Stasko's [2004] heuristics were most useful for evaluating the discovery process. However, there was significant overlap between these sets in terms of this categorization.

### Generalizable Problems

Our preliminary exploration also involved significant discussion of some problems with the *LuMPB Key* system that may be common to other information visualizations. Some of these problems included difficult-to-see visual components due to contrast issues, assignment of colour value resulted in confusion or difficulty to perceive relationships, and lack of detailed information in "tool-tips". In the same way that Nielsen [1994] refined a set of usability problems into a small set of heuristics, both to cover all problems found and to cover all serious problems found, repeating our process with several other information visualizations could provide this same data set and allow the same form of analysis.

### Process

Amar and Stasko's heuristics were found by most evaluators to be difficult to apply without extra domain knowledge. It may generalize that one set of heuristics will benefit most from domain expert involvement, or a particular part of the design life-cycle. Broader heuristics such as Amar and Stasko's may also lend themselves more toward use in design than evaluation, as they may have major implications for system requirements that need to be addressed earlier in the development process in order to reduce costs.

Higher level heuristics such as Schneiderman's and Amar and Stasko's tended to require consideration of additional visualizations the system provided, or the system as a whole, for proper application. Therefore, in our attempt to restrict evaluation to a couple of views, the use of these heuristics led most evaluators to questions about the views not analyzed. One evaluator commonly included another view to aid in the application of the heuristics, while the system developer could not help but consider the entire system.

Lower-level heuristics may thus work better when analyzing a decomposed larger system. In order to minimize learning both a complex visualization tool and the related domain knowledge, one could borrow from Extreme Programming [Astels et al., 2002] and have a domain expert and evaluator work in a pair.

Usability issues were often tied to a detected visualization problem, so a set of usability heuristics would have been a useful addition (e.g. minimize user memory load, clearly marked exits, ... [Nielsen, 1993]). With the addition of other sets of potential heuristics some organization may be necessary. This leads to the problem of heuristic selection and whether partitioning a larger set of heuristics is useful, both of which will require further research.

One of the evaluators used supplementary software while applying the heuristic “Consider people with colour blindness”. Screen shots of the charts were automatically recoloured to test how a colour blind person would see them [Dougherty and Wade, 2006]. This finding raises the question of if and how tools may support heuristic evaluation. The use of tools for evaluation is related to the automatic design of visualizations based on heuristics, such as Mackinlay’s [1986] system for relational information using formal expressiveness and effectiveness criteria.

## 3.9 Conclusions

Our meta-analysis has added to the understanding of using different sets of heuristics for evaluation of visualizations. The approach of using three different sets of heuristics provided practical guidance for the *LuMPB Key* system some of which the designer planned to integrate into the next version. The approach also revealed some characteristics, such as redundancy and conflict, that may be generally useful when comparing different heuristics. We found value in using visualization-specific heuristics, as problems were found that would not have been discovered by usability heuristics. Similarly the uncertainty visualizations analyzed at the beginning of the chapter demonstrated the value of heuristic evaluation even with one set of heuristics.

Many problems we found crossed theoretical and knowledge boundaries, and therefore the evaluation process would benefit from including experts from visualization, usability, and the domain area. Information visualization's focus on amplifying cognition means that heuristics related to higher level cognitive tasks such as Amar and Stasko's [2004] delve into issues that only the domain expert may understand. These higher-level issues also require a holistic evaluation of entire systems and so do not lend themselves to a strategy of divide and conquer.

Both finding an appropriate taxonomy of heuristics and finding a minimal set of heuristics that can find the majority of problems or provide the best guidance will require a large amount of research. During this research, it may be useful to continually look at different organizations of heuristics and different processes which may be more efficient in finding problems and suggesting solutions. Uncertainty visualization along with all visualizations should benefit from these types of heuristic evaluations. This heuristic approach may also be useful as a more general tool to assist the design process in creating new and effective uncertainty visualizations. In Chapters 5, 6, and 8 we return to the heuristics in Table 3.1 in order to evaluate the domain specific visualizations we developed.

# Chapter 4

# Visualization Support for Reasoning Under Uncertainty

> I found it peculiar that those who wanted to take military action could - with 100 per cent certainty - know that the weapons existed and turn out to have zero knowledge of where they were.
> – Hans Blix (1928 –)

Uncertainty in data is paralleled by uncertainty in reasoning processes, and while uncertainty in data is starting to get some of the visualization research attention it deserves, the uncertainty in the reasoning process is thus far often overlooked. This chapter gathers and consolidates the issues involved in uncertainty relating to reasoning and analyzes how uncertainty visualizations can support cognitive and meta-cognitive processes. Uncertainty has been mentioned in the previous chapters often in regard to decisions. Any uncertainty in decisions may arise from uncertain data, uncertainty in reasoning, or often a compounding of both. While concurring with the importance of incorporating data uncertainty into visualizations, we suggest also developing closely integrated visualizations that provide support for uncertainty in reasoning[†].

## 4.1 Introduction

Uncertainty and its complement certainty are fundamental parts of any analytic or reasoning process and relate to important cognitive constraints in using any visualization. To inform the design process we review and coalesce many important aspects of reasoning under uncertainty and discuss these with regard to implications for visualization. For each of these aspects we consider reasoning and representational requirements and assess the

[†]Portions of this chapter have been previously published in [Zuk and Carpendale, 2007]. Thus any use of the word "we" may refer to Torre Zuk and Sheelagh Carpendale

potential for exploiting visual support. Based on our analysis of the impact of uncertainty in the reasoning processes, we propose that these receive increased consideration in the design of visualization systems. For instance, when appropriate this could include an additional visual component focusing on reasoning uncertainty and support for introspection. For this reasoning support we contribute design considerations and touch on an example system for medical diagnosis, which is described in detail in Chapters 7 and 8.

In the analytic reasoning process, often choosing the visual representation drives exploration for an iteration of searching, comprehension building, or hypothesis testing. The inability to transform or change this representation is the representational primacy that Amar and Stasko consider a limitation of many current visualizations [Amar and Stasko, 2005]. In addition to options for alternate representations, it is important to provide representations of uncertainty in order to allow potential interpretations of the data to be considered. Hepting has described an analogous process for visual interfaces as "begin with an incomplete articulation of a context and allow the user to interactively develop and refine it" [Hepting, 2002]. Leaving uncertainty out of a data visualization promotes assumptions that lead to more uncertainty in the reasoning process and the viewer may not be aware of this uncertainty. With insight problems (e.g. the 9-dot problem [Novick and Bassok, 2005]) searching representation space may be key and Gestalt may even hinder the process [Novick and Bassok, 2005]. Thus providing cues about uncertainty in representation may promote consideration of other representations and help further the exploration. Based on and extending the impact of data uncertainty visualization, we suggest that representing the reasoning process may aid in determining both the next reasoning step, and the assessment of the solution. Further, this visual representation specifically designed to support the reasoning process should also incorporate uncertainty to provide transparency of confidence.

One cornerstone of reasoning uncertainty is the relationship of ignorance and knowledge. For many problems accurate assessment of the completeness of knowledge can

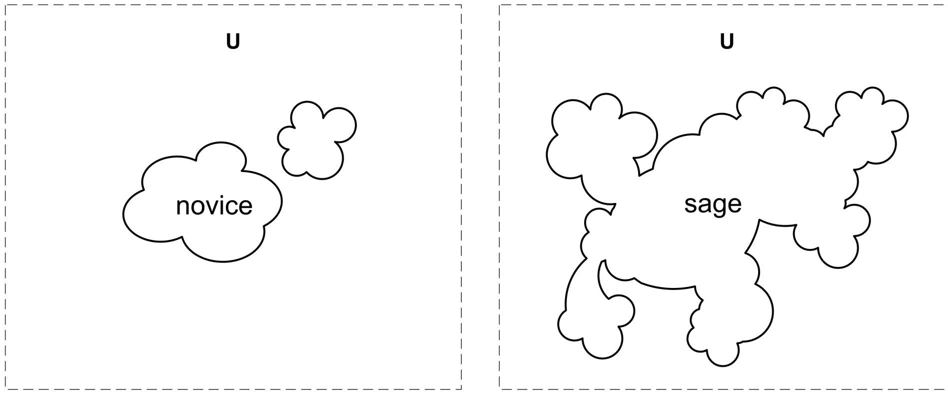



Figure 4.1: Knowledge and ignorance shown in Venn diagrams of novice and sage. Self perception of ignorance, or confidence, may be a function of knowledge set perimeter.

never be known. Therefore completeness of relevant knowledge is based on one's perception. The statement by a wise old sage that they are more ignorant than novices who are just beginning their studies, may be interpreted based on self-perception. Figure 4.1 reveals a diagrammatic rationale for this statement. Notice that while the sage's knowledge covers a larger area than the novice's, the length of the sage's perimeter is much longer thereby increasing the number of points in which the sage is aware of not knowing some aspect. Self perceived ignorance (or uncertainty) may be based on the perimeter length of their domain knowledge, or number of associations into the unknown.

Perceived uncertainty may be considered a function of incomplete knowledge. While one attempts to minimize this function through exploration or data gathering, it may actually increase. Unlike other optimizations, a previous minimum may not be returned to, as knowledge once gained is not so easily lost. Local minima may also provide a false sense of confidence, and truncate the search of solution space. This leads to the notion that there may be significant second order uncertainty (uncertainty in the uncertainty) and unquantified uncertainty.

Given that both knowledge and representation are coupled to uncertainty, we will present arguments to illustrate that uncertainty of reasoning as well as uncertainty in data should be visualized and if possible integrated in a manner that supports the reasoning process. Even well-defined problems such as proving a premise using predicate logic usually requires an external aid (visualization, such as hand drawn sketches) due to the limits of working memory. When adding the complexity of uncertain data or actions, one would expect Bayesian reasoning or some form of satisficing[1] would also benefit from visualization support.

## 4.2 Cognition, Uncertainty, and Visualization

In this section we have gathered together the central components of several discussions of reasoning and cognition and discuss them in light of uncertainty visualization. For our discussion we define reasoning very loosely and consider how *knowledge constructs, heuristics and biases*, and *temporal constraints* impact reasoning and discuss the potential for uncertainty visualization. The cognitive psychology definition of the term heuristic used in this chapter is any non-algorithmic cognitive process used to perform a calculation or make a decision[2], which is somewhat akin to its meaning with regard to heuristic evaluation in Chapter 3. We close this section by delineating types of reasoning uncertainty.

### 4.2.1 Knowledge Constructs

Thomas and Cook [2005] describe three higher order knowledge constructs: *arguments*, *causality*, and *models of estimation*. Arguments are "logical inferences linking evidence and other reasoning artifacts into defensible judgments of greater knowledge value" [Thomas

[1] A strategy of seeking "good enough" over optimality, due to complexity from, among other things, uncertainty [Simon, 1956, Stirling, 2003]. Gigerenzer et al. [2003] separates bounded rationality into two types: satisficing and constrained maximization.

[2] This definition is vague but is intended to include imprecise or ad hoc methods which are not guaranteed to find optimal solutions.

and Cook, 2005]. Causality is an understanding of the action-reaction relationship. Models of estimation provide for the use of powerful abstractions in interpreting the data and providing estimates of solutions. We will discuss these three constructs in terms of their relationship to visualization.

**Arguments and Visualization**

Visualizing an argument formalizes it for introspection and collaboration. *Arguments* are one of the reasoning steps of problem solving, and the presence of uncertainty is what creates an ill-structured problem. Paraphrasing van Bruggen et al. [2003], an ill-structured problem has:

1. an ambiguous and incomplete problem specification,
2. a lack of clear stopping criteria,
3. multiple information sets and representations with no clear indication of relevance, and
4. incomplete knowledge of operations or solution path.

Solving ill-structured problems often requires non-linear progression, partial solutions, and representational refinement [van Bruggen et al., 2003], for which extra cognitive support will be beneficial.

Complex problems and arguments are also more likely to require external assessment or benefit from collaborative refinement. Without a representation of the current uncertainty in different analytic strategies resource management is difficult, as expected values (probability weighted returns) are needed to determine trade-offs. By visualizing which areas have uncertainty and are making the problem ill-structured, users may more easily monitor progress and decide to divert resources to reduce the most significant uncertainty. While various visualizations for argumentation exist [Kirschner et al., 2003], it is an open question how they can be integrated into specific task-oriented decision processes, and visualization tools in general.

### Causality and Visualization

More causality may be perceived than was intended in the visualization. *Causality* is often perceptually linked to temporality. Michotte [1963] found that with the movement of patches of light, the relative timing of motion could create the strong perception of causal relationships. Likewise with less abstract occurrences people will often assume causality based on temporal relationships. Due to this perception, animation may enhance the communication of causality and should be used carefully if causality is not to be inferred.

Reasoning about causality under uncertainty may also utilize heuristics that are prone to error and bias. Tversky and Kahneman found that if one event ($C$) was naturally viewed as a cause of another ($E$), then even if they had equal probabilities their participants would be biased in favor of causal inferences over diagnostic inferences (i.e. believe the conditional probability $P(E|C) > P(C|E)$ even though $P(C) = P(E) \Rightarrow P(E|C) = P(C|E)$) [Tversky and Kahneman, 2003a]. Furthermore they found that people were biased toward weighing evidence for causal impact in the future versus diagnostic reasoning about the past. Kahneman and Miller hypothesize that alternatives to the effect are more available to the mind than alternatives to the cause [Kahneman and Miller, 2003], and so leading the user to consider more causes could reduce this bias. When there is an effect with an uncertain cause this might be visually indicated by the use of visualization artifacts such as showing additional dangling links back from the effect.

### Models of Estimation and Visualization

A visualization is a model which adds its own uncertainty. Applying any *models of estimation* requires a jump from the concrete to the abstract. This may likely increase uncertainty by requiring assumptions, introducing translation errors, or adding in the stochastic variability of a model. Any uncertainty this abstraction process introduces should be visualized to keep under consideration when interpreting the model results. The propagation of errors is also important to consider when using models as the input uncertainty will often be increased, potentially by something as simple as the addition of variables or rounding [Wilkinson, 1994]. The propagation of uncertainty through a model with regard to visualization has been discussed by Davis and Keller [1997]. In their domain they claim that the impact of uncertainty visualization will be difficult to evaluate and that "a shift in 'spatial understanding' regarding uncertainty can only be judged through its effects on policy, resource decisions, scientific hypothesis generation, or other 'bottom-line' items" [Davis and Keller, 1997, p.406].

### 4.2.2 Reasoning Heuristics and Biases

An exemplar of reasoning heuristics and biases may be found from user prediction calibration. Griffin and Tversky [2003] state in the assessment of evidence that overconfidence often resulted when the evidence strength (extremeness) was high and weight (predictive validity) low. For example, there may be a bias toward rejecting the null hypothesis when the means are very different even when there are large standard deviations. Under-confidence often resulted when the strength of evidence was low but the weight high (i.e. a moderate impression based on extensive data) [Griffin and Tversky, 2003]. An example may be the failure to confidently communicate the need to address climate change. One might help address these biases by showing the merged strength-weight visually.

For information systems Turpin and du Plooy [2004] review the decision making heuristics and biases: *representativeness*, *availability*, *adjustment and anchoring*, *problem or decision framing*, and *automation*. Their literature review found real world examples providing some evidence for each of these types. They touch on the role of how information systems may elicit biases as well as aid in debiasing, and also suggest innovative representations and decision process support may reduce bias. They conclude by calling for more field research to better quantify the effects of these biases in relation to other problems such as data quality.

The debate continues as to how frequently these individual heuristics and biases occur outside the laboratory [Griffin and Tversky, 2003, Klein, 1998], but they are likely relevant to design when considering user constraints, as evidence of their presence in the field has been found [Turpin and du Plooy, 2004] (also see Chapters 7 and 8). Klein argues against practical weaknesses of the heuristics used in expert "naturalistic" decision making and argues for their combined strengths as part of his recognition primed decision model (RPD). "The core of the RPD model is a set of heuristics previously described by Amos Tversky and Daniel Kahneman: the simulation heuristic (1974), used for diagnosis and evaluation, and the availability and representativeness heuristics (1980), for recognizing situations as typical [Klein, 1998, p.298]". However, even the heuristics and biases found only in the laboratory may reveal insight into reasoning processes, just as optical illusions may aid the understanding of perception.

We provide a subset of these heuristics and biases, most from the foundational collections on the subject [Griffin and Tversky, 2003, Kahneman et al., 1982], and others as cited. We have organized these into three categories based on visualization strategies that may potentially mitigate them. The categories are: *associations*, *ignorance of rules*, and *application of rules*. Mental associations have a conscious and subconscious influence on reasoning. Rules encompass the simple cognitive constructs for inferring information (e.g. a theorem) all the way up to methods for forming arguments. We will describe each in turn along with visualization strategies that may be beneficial.

**Associations and Visualization**

A visualization is impacted both positively and negatively by associations it triggers. *Associations* may bias the reasoning process in various ways. One major type is the *affect* or reliance on the associated "good" or "bad" response to a stimulus [Slovic et al., 2003], which Norman has recently discussed in relation to its impact on design [Norman, 2003]. *Availability* of instances in the mind for estimating probability form another type of associative heuristic impacting the interpretation of visualization:

- retrievability of instances is important when the size of a set is estimated by availability of instances [Kahneman et al., 1982];
- effectiveness of a search set in which availability of contexts may not relate to instances [Kahneman et al., 1982, Howell and Burnett, 1978];
- if instances are not available, the ease of imagining them will act as availability [Kahneman et al., 1982];
- *illusory correlation* when the frequency of co-occurrence may be estimated based on strength of association [Kahneman et al., 1982], and
- *recency bias* results in the overweighting of recent events [Tufte, 2006].

Visualizations can provide access to huge amounts of data and thereby reduce the biases of one's own limited associations. Using high data density visual queries that can be quickly modified, one may be influenced less by expectations, and be more amenable to let the data provide its own associations. Similarly, the use of a computer to analyze the data and make a visualization based on a fixed set of rules may in itself reduce these types of biases.

**Ignorance of Rules and Visualization**

If a visualization does not convey to the viewer the meanings of its representation(s) the user may fail to form the correct interpretations and arguments. *Ignorance of rules* (often statistical) can also lead to poor reasoning and the *representativeness* heuristics [Kahneman et al., 1982] in which how well an instance represents a set is used to estimate probability rather than set sizes. These include:

- insensitivity to prior probabilities (e.g. Bayes' rule not applied);
- small sample expected to be as representative of population as a larger sample;
- failure to consider regression to the mean;
- misconceptions of chance (e.g. representativeness of a random process as a whole expected in short sequences);

- irrelevant data may be used as a predictor; and the
- *illusion of validity* where redundancy in inputs reduces accuracy but increases confidence.

While visual representations themselves may not promote statistical ignorance, they rarely go the one step further to aid statistical interpretation. Even the basic box and whisker plots tailored for hypothesis testing are in rare use. Visualizations have the potential to alleviate these issues by integrating realizations[3] of other potential outcomes (e.g. using stochastic simulation), and integrating direct access to more detailed statistical information.

Heuer [1999] provides both analytic cognitive strategies, some of which could be classified as heuristics (in the category we describe next), and discusses the applicability of the lower level heuristics and biases (our ignorance of rules category and the previous associations one) to intelligence analysis. Ignorance of rules should be kept in mind for all heuristics in the next category, as facilitating the use of new strategies may have additional value for inexperienced users. As an example of this, Cluxton and Eick's [2005] hypothesis visualization tool has integrated Heuer [1999]'s "Analysis of Competing Hypotheses" method with some additional uncertainty parameters.

**Application of Rules and Visualization**

Direct visual support for reasoning may assist with the application of rules. Any particular reasoning strategy or *application of rules* may provide approximate results (i.e. less than an optimal solution), as is possible with the *adjustment and anchoring* set of heuristics. The two aforementioned categories of heuristics and biases may affect any of the heuristics or strategies in this category. An illustrative subset of the application of rules category are:

- insufficient adjustment when an initial estimate is weighted too strongly during sub-

[3] In this dissertation *realizations* are defined as specific potential outcomes from a set of probabilistic outcomes.

sequent revisions (and may be based on irrelevant data) [Edwards, 1982, Kahneman et al., 1982];

- adjustment from single event probability produces overestimates of the probability of conjunctions of events ($P(A \cap B)$) and underestimates of disjunctions ($P(A \cup B)$) [Kahneman et al., 1982];
- a tendency to be overconfident in decisions or estimates [Fischhoff, 1982, Howell and Burnett, 1978];
- *automation* heuristic or technology dependency leading to errors of omission and commission [Cohen et al., 1998, Skitka et al., 1999, Turpin and du Plooy, 2004];
- overestimated confidence in the ability of a priori predicting past events (i.e. hindsight is 20:20) [Fischhoff, 1982]; and
- *escalation/entrapment* in which the decision maker spends more resources than justifiable (e.g. Vietnam War[4]) [Matlin, 1983];
- the *recognition primed decision model* describes how experts can quickly make decisions. Experience, and when necessary diagnosis, are used to judge typicality. Evaluation of an atypical action may be performed using mental simulation [Klein, 1998].

This application of rules category in general relates more to the reasoning process than the data. Similar to this category, the use of heuristics in software programs dealing with complex problems is also common-place. These heuristics need to be understood by the user in order to avoid potential interpretation errors.

Many visualizations do not include visual explanations of the mapping of data, algorithms and uncertainty, but this is crucial for avoiding these types of biases. Reasoning shortfalls in this class will be greatly aided by a visualization of the reasoning process

[4]One is not hard pressed to think of a more recent example. A heuristic such as escalation/entrapment may be an over generalization for a series of decisions in a governing body, however when a single person has controlling influence these heuristics could feasibly be a major factor.

itself. Any reasoning visualization may provide grounds for review, analysis, and collaboration; thereby opening up what might be a hidden and flawed decision process. Just as MacEachren noted for visualization errors [MacEachren, 1992], we can group reasoning errors into Type I, reaching conclusions that are not supported, and Type II, failure to reach conclusions that are supported.

When biases or problematic heuristics are likely to manifest in a user's reasoning, we can make attempts to debias or provide alternative heuristics (or algorithms). Fischhoff reviewed some of these attempts for *overconfidence* and *hindsight* bias, and found only partial success [Fischhoff, 1982]. The review was organized around three categories: faulty tasks (attempts such as raise stakes, clarify instructions, ...), faulty judges (warn of problem, train extensively, ...), and mismatch between judge and task (make knowledge explicit, search for discrepant information, ...). There is greater potential for cognitive support with visualization systems as the offloaded tasks may use algorithms that do not suffer from these issues, and may dynamically attempt debiasing, but the danger of the *automation* heuristic also needs to be considered.

For many problems, heuristics can provide fast and accurate enough approximations for the task at hand. Gigerenzer et al. compared some satisficing methods (fast and frugal heuristics) against some "optimal" algorithms (e.g. Bayesian networks) representing unbounded rationality [Gigerenzer et al., 2003]. With complete knowledge and across 20 real-world scenarios[5] some simple heuristic strategies (*minimalist* and *take the best*) were found to perform comparably to the algorithms [Gigerenzer et al., 2003]. If specific heuristics are accepted for use as standard operating procedures we may look at providing visualization support to enhance them further or to reveal when they can not be trusted. Identifying decision requirements and constraints can be used to guide visualization design and [Klein, 1998, p.108] describes a case where using decision requirements to refine

[5] Some of these scenarios were estimating: high school dropout rates, stoichiometric products, and numbers of eggs in fish, each based on training data with multiple sets of cues [Gigerenzer et al., 2003].

an existing system led to improved task performance.

Arnott [2006] has provided a taxonomy of biases and proposed a general decision support design methodology utilizing these theories. Watkins [2000] also reviewed many cognitive heuristics and biases and believed that they are worth considering for uncertainty visualization. While we agree that they are an important design consideration, especially when providing a decision support tool, we should be wary of their potential impact on the analysis and discovery process, and so should perform research on their role in visualization in general.

If we assume two cognitive models of reasoning working in parallel: associative and rule-based [Sloman, 2003], then some issues may be more related to one model. The associative system may be directly affected by Gestalt and a visualization's ability to convey the required uncertainty for immediate processing and consideration. There may be the flexibility in rule-based reasoning to use methods that avoid the drawbacks of potential heuristics and biases. With the more general rule-based reasoning we have the potential to learn and utilize problem solving heuristics that have been validated to some extent, but perhaps at the cost of sacrificing creativity and imagination (associative). A graphical or visualization system should try to provide assistance to both systems but avoid leading users to the *automation* heuristic.

### 4.2.3 Relating Uncertainty to Temporal Constraints in Reasoning

One fundamental constraint on the reasoning process is time. Time stress and other situational attributes can distort our perception leading directly to biases [Mandel, 1979]. This distortion adds uncertainty, confounding the uncertainty that may have led to the time stress. Strategies will vary based on the amount of time resources available. At a high level it may be similar to game strategies in which search space (e.g. minimax tree) is pruned based on the time allowed. Cognitive models such as Cohen et al.'s [1996] Metarecognition model have been proposed for time limited decision making. Driven by factors from cognitive models, visualizations may assist by illustrating uncertainty of the data, but visual support of meta-reasoning may be the area where they can contribute the most.

Watkins [2000] created and analyzed an uncertainty glyph to depict three types of uncertainty and their sum in a decision support system. One interesting finding was that all analyst participants (5 National Air Intelligence Center analysts) agreed somewhat or stronger that in general "uncertainty visualization would degrade the ability of most analysts and decision-makers to respond to or 'interpret a scenario' in a timely manner" [Watkins, 2000, p.181–3]. Participants' rationale for this rating referred to the issues of complexity and overload. The majority thought, however, it would not overload decision-makers in less time-constrained situations, and were not comfortable adding data with associated uncertainty to a knowledge base without an uncertainty visualization.

Delay is Lipshitz and Strauss's [1997] first conceptual proposition: uncertainty in the context of action is a sense of doubt that blocks or delays action. They cite Dewey's statement that problem solving is triggered by a sense of doubt that stops routine action [Dewey, 1997], but dropped the important aspect that it is uncertainty that triggers problem solving, which necessitates neither blocking or significant delay. One should note that changes in uncertainty may trigger action, and that delay can be the optimal "action". An example of this may be the space shuttle Challenger disaster, for which the criticality of data quality has been discussed by Fisher and Kingma [Fisher and Kingma, 2001]. Delaying the launch of the shuttle until further analysis removed the uncertainty about the safety of the O-rings under cold temperatures may have averted the disaster. Tufte has also analyzed the space shuttle Challenger and Columbia disasters from a visualization point of view [Tufte, 1997, 2006], and one may argue the most significant uncertainty was not in the data but in the reasoning.

Table 4.1: Extending Thomson et al.'s [2005] typology of uncertainty to reasoning.

| **Uncertainty Category** | **Reasoning Definition** |
|---|---|
| Currency/Timing | temporal gaps between assumptions and reasoning steps |
| Credibility | heuristic accuracy & bias of analyst |
| Lineage | conduit of assumptions, reasoning, revision, and presentation |
| Subjectivity | amount of private knowledge or heuristics utilized |
| Accuracy/Error | difference between heuristic & algorithm (e.g. Bayesian) |
| Precision | variability of heuristics and strategies |
| Consistency | extent to which heuristic assessments agree |
| Interrelatedness | heuristic & analyst independence |
| Completeness | extent to which knowledge is complete |

### 4.2.4 Types of Reasoning Uncertainty

There are many taxonomies of uncertainty to be found in different domains. Lipshitz and Strauss found in a study of military decision makers that they distinguished between *inadequate understanding*, *incomplete information*, and *undifferentiated alternatives* [Lipshitz and Strauss, 1997]. Different strategies were employed based on these types of uncertainty. Thus task considerations may dictate the types of uncertainty that are significant. Hence we would suggest a user and task centered approach be taken with uncertainty issues.

Thomson et al. have constructed a typology for visualizing uncertainty in geospatially referenced data [Thomson et al., 2005]. They considered Pang et al.'s low-level classification [Pang et al., 1997] and Gershon's high-level taxonomy [Gershon, 1998] and provide a typology to be instantiated based on task, giving examples from intelligence analysis. They advise a hierarchical approach for instantiating this typology across multiple domains or tasks. We extend the definitions of their typology to the reasoning process in Table 4.1, demonstrating how their typology is useful at the level of reasoning as well. Considering how this typology applies to reasoning can extend its intended purpose of guiding the development of visual representations for uncertainties.

Dynamic data is one of the main reasons why *currency / timing* is tied to uncertainty. Thereby the error between prior observations and the current state generally increases over time. In some cases the duration of observation allows for a trade-off between uncertainty in one attribute and another related or meta-attribute (e.g. the attributes' derivative). For example, Heisenberg's Uncertainty Principle dictates the tradeoff between accuracy in position and momentum at the quantum scale. Temporal constraints are a major reason why completeness of knowledge can not be fully attained. Past decisions, assumptions, and arguments often form the a priori knowledge base. Visualizing the impact time constraints had on this prior information can greatly influence its usage. Opacity is often used for temporal encoding where data fades out over time as it becomes dated.

For *credibility*, *lineage*, and *subjectivity*, all levels from data gatherers to decision-makers should be considered in the reasoning instantiation of the framework. When the decision processes span multiple levels of management or government these aspects are especially important to consider. As an example of this decision scenario, we can look at when the director of the NASA Goddard Institute for Space Science (a climatologist) had the qualitative certainty and causality in his report on climate change strongly diluted by the U.S. White House Council on Environmental Quality [CBS Broadcasting Inc., 2006] (See Figure 4.2). In this case the reader weights the judgments based on the assumed credibility and subjectivity of the scientist authors, with no way of knowing that a non-scientist had revised the scientific judgment. The final form of edited paragraph is shown in Figure 4.3. On the single page that contained the paragraph shown, eleven changes were made to reduce scientific certainty, nine of which made it into the final version (subjective but conservative analysis). The final decision makers (U.S. Congress) would benefit from visualizing the uncertainty in credibility, lineage, and subjectivity of reasoning. Ignorance of any of these types of uncertainties may directly impact the ability of decision-makers to make good decisions, and therefore guidelines mandating the visualization of such uncertainty should be considered.

To visualize *accuracy / error* one must consider the effects of potential heuristics and biases, as discussed in Section 4.2.2. The visualization of reasoning accuracy will likely not be possible unless tools are used for the reasoning in which heuristics and strategies are made explicit. Error itself is not usually known a priori and so would be visualized as a postmortem task. Visualizing *consistency* and *precision* in heuristics or strategies is important for decision confidence. Precision of a single heuristic may be difficult to assess as cognitive strategies themselves may not be precisely defined. The same visualization of reasoning heuristics that provides an estimate of precision, could likely reveal inter-heuristic consistency.

Visualizing *interrelatedness* may allow results from analysts working in teams to be collectively considered. It may be useful for the interrelatedness of heuristics and analysts to be visualized using preattentively processed visual cues. For example, connectedness (from Gestalt theory) may allow one to consider linked reasoning artifacts holistically, potentially reducing the risk of over weighting redundant findings. Our reasoning instantiation of *completeness* includes comprehension (ignorance) some aspects are dependent

natural variations in ocean temperatures and currents, all cause variability and changes in climate conditions.

Many scientific observations ~~point to the conclusion~~ indicate that the Earth ~~is~~ may be undergoing a period of relatively rapid change on timescales of decades to centuries, when compared to historical rates of change on similar timescales. Much scientific evidence indicates that these changes ~~are~~ are likely ~~may be~~ the result of a complex interplay of several natural and human-related forces.

Although humans are relative newcomers in the vast scale of the Earth's geological history, we

Figure 4.2: Draft copy showing hand editing of scientific confidence. Changing of definite wording "is" to speculative "may be" among the 3 revisions in the paragraph shown.

Many scientific observations indicate that the Earth may be undergoing a period of relatively rapid change on timescales of decades to centuries, when compared to historical rates of change on similar timescales. Much scientific evidence indicates that these changes are likely the result of a complex interplay of several natural and human-related forces.

Figure 4.3: Final version of paragraph shown in Figure 4.2 [Mahoney, et. al., 2003, p.2]. Changes in certainty are hidden in final presentation due to lack of lineage visualization.

on all the other types of uncertainty being visualized. Similar to error, in advance completeness will usually only be estimated. A good example of the cost of unvisualized uncertainty is the wasted resources in duplicated research caused by the lack of publishing on scientific failures.

## 4.3 Visual Support for Uncertainty in Reasoning

Numerous methods have been proposed integrating uncertainty into data for visualization [Pang et al., 1997], and some have been evaluated for specific tasks [Grigoryan and Rheingans, 2004, Masalonis et al., 2004]. However there has been less research into how well these provide decision support. How best to provide reasoning and meta-reasoning support that incorporates uncertainty is an open question.

### 4.3.1 Problem Solving

Newell and Simon [Newell and Simon, 1972] provided a high level organization of a problem solver for a generic information processing system. We have used this organization to highlight aspects of uncertainty in the process of reasoning in general as shown in Figure 4.4. While uncertainty likely exists in some form in all aspects of the organization, the method selection process is important (shown in bold red in the figure) in that it is affected by both data and problem representational uncertainties as well as potential ambiguity in the relationship of methods to goals. Our looser interpretation of their general problem solver allows the method selection to require problem solving (a recursive invocation) and methods would include all heuristics and strategies (top-down, bottom-up, etc.). Visual aids for the method selection process would likely be beneficial as this complex "phase" requires the consideration of sub-goals and the actions related to them, while still considering their context in the overall problem. There is the potential for change in both internal and external representations of the problem and of the data [Scaife and Rogers, 1996].

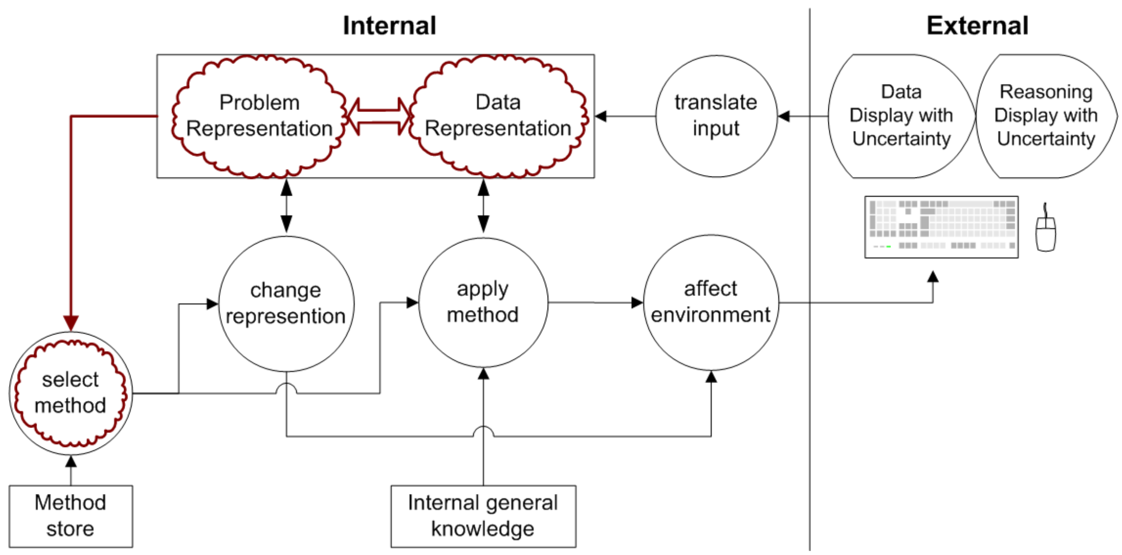



Figure 4.4: Organization of problem solving with uncertainty. Application of Newell and Simon's general organization of a problem solver [Newell and Simon, 1972, p.89] to visualization based problem solving, with additional delineation of where uncertainties likely exist.

While traditional graphics and HCI research has focused on the external part, more considerations need to be made for the internal part. The visualization system should also produce the artifacts that may assist introspection on the cognitive process. As these processes are tightly coupled, the ability to monitor and aid the reasoning process will add additional requirements to the visualization. Visualizations may need to be modified in order to allow parallel support for both data and reasoning process visualization, which might be useful to think of as a larger task context. This support could tie both direct visual artifacts in with meta-data artifacts recording the history of exploration and the discovery processes that were used.

### 4.3.2 Analytic Processes

Analytic reasoning can be generalized as a set of tasks [Thomas and Cook, 2005, p.42]:

1. information gathering,
2. re-represention for the purpose of analysis,
3. development of insight (via observation and interaction), and
4. production of knowledge or decision;

with the repeated iteration of these tasks forming a "sense-making loop". Visualizations may be used to support all four of these tasks. Uncertainty may exist throughout the analytic reasoning process and thus visual support for the process as a whole may provide benefits including providing assistance for meta-analysis.

Amar and Stasko's [2005] precepts for design and evaluation of information visualizations provide a set of principles on how visualizations can better support the analytic process. The three main weaknesses of current systems were stated as: limited affordances, predetermined representations, and the decline of determinism in decision making. These weaknesses or gaps in the analytic process were to be addressed by the Rationale Precepts: expose uncertainty, concretize relationships, and expose cause and effect; as well as the Worldview Precepts: Determine Domain Parameters, Expose Multivariate Explanation, and Facilitate Hypothesis Testing. All of Amar and Stasko's precepts deal with complex issues and appear to pertain to reasoning as a whole, thus providing guidelines for reasoning visualizations and support as well as information visualizations.

Bridging the analytic gaps and extending ideas in current information visualization systems to reasoning visualizations will likely require the linking of these types of tools, or developing additional integrated cognitive support, while ensuring consistent cognitive styles to avoid a huge context switch. We propose that for visualizations that assist with complex problem solving, that support for reasoning with uncertainty be built into the visualization pipeline. This integration could be as light-weight as virtual sticky notes for one's ideas that are colour coded based on certainty. Figure 4.5 shows our extension to Pang et al.'s visualization pipeline [Pang et al., 1997] to include reasoning support with

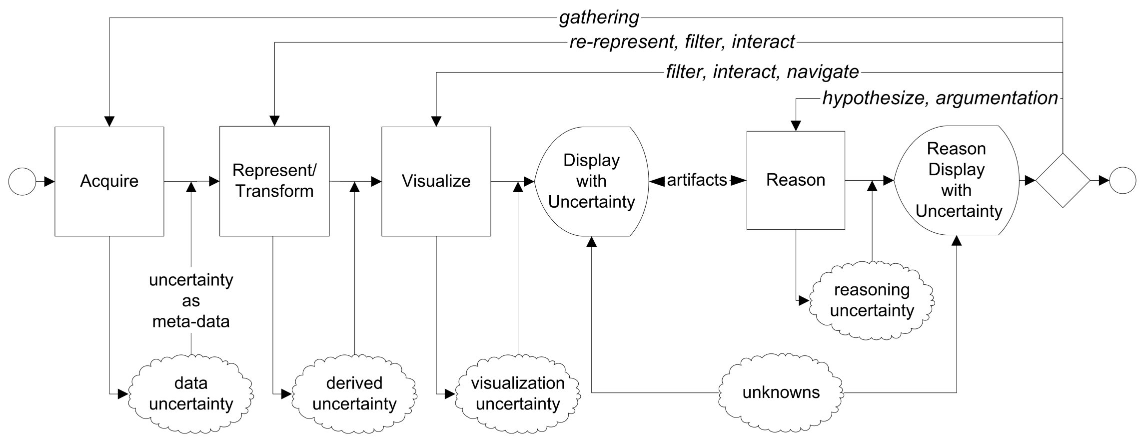



Figure 4.5: Reasoning extensions to Pang et al.'s data visualization pipeline with uncertainty [Pang et al., 1997]. Extensions are on the right side of the artifacts link and provide integrated sense-making loops.

uncertainty. The top arrows in the figure relate to loops in analytic sense-making, which is the process of searching, extracting, and modelling, with the goals of building evidence, schemata, hypotheses, and understanding [Thomas and Cook, 2005]. This integration of data and reasoning visualization support provides benefits by simplifying the backtracking (revaluation and searching) phases of the sense-making loop. Thus uncertainty in one case or hypothesis would be easily reviewable by another user. Visualizations for uncertainty in both the data and reasoning pipelines could use consistent representations and/or metaphors for the same types of uncertainty to reduce cognitive load. The complexity and constant evolution of visualization tools promotes specialized systems to handle specific sub-tasks. Therefore the visualization pipeline may span multiple systems and so providing visual consistency will add design constraints. Independent applications will require support for restoration of data, operations, and viewing parameters.

The link between visualization and reasoning pipelines should be bidirectional to allow for feedback from the reasoning process for potential integration into the visualization tools. This could be as simple as including the goal or larger context in the reasoning process that may be provided with text or graphics. It could also communicate a strategy of exploration which the data visualization tool could then dynamically facilitate. In a collaborative setting this might be valuable to provide awareness of strategy changes when one is focused on a small scale task. There are existing visualizations aimed directly at reasoning support [Kirschner et al., 2003], but there should be further benefits from bridging the gap between them and the information visualizations associated with foraging (search, filter, extract, ...) for information. While this concept has been implemented to a limited extent for integrating links from data/evidence directly into argument structures (e.g. Cluxton and Eick's [2005] DECIDE™, and BAE Systems' POLESTAR), most general information visualizations provide little or no direct reasoning support or are not linked to one that does.

Using a participatory design approach we have developed a prototype system for evidence-based medicine diagnostic support that provides this parallel (reasoning/data) visualization approach. The parallel visualizations are in the form of multiple views, two of which are shown in Figure 4.6. It utilizes a decision tree as a GUI with integrated reasoning and data uncertainty. The reasoning visualization can be viewed along with other data and its uncertainty in multiple other views. This design provides transparency of the uncertainty in the Bayesian reasoning that may assist in this difficult task. This system is only briefly described here for illustrative purposes, and will be fully described later in Chapters 7 and 8.

### 4.3.3 Representations

Visual representations of data uncertainty allow for the amplification of cognition [Card et al., 1999] (i.e. visualizations allow parallel processing, increased working memory, etc.), and when time frames allow introspection, we suggest similar benefits will accrue from visual representations of reasoning uncertainty. Kirschenbaum and Arruda [1994] found an ellipse was more accurate than verbal quantification in communicating uncertainties in

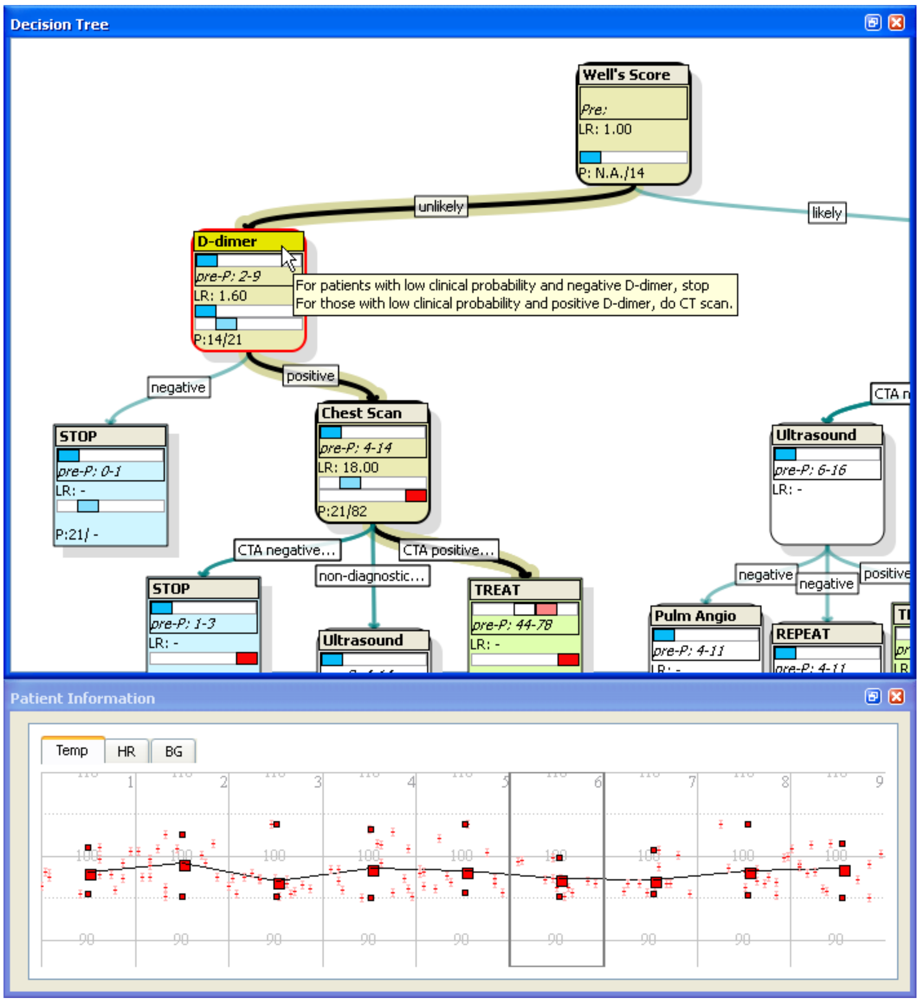



Figure 4.6: Integrated data and reasoning visualizations for evidence-based medicine. Reasoning support view (upper) and test data view (lower).

a spatial problem. With non-spatial uncertainty, Watkins [2000] found his glyph (which distinguished: unreliability, ignorance, and analytical input) was rated well by analysts but with some qualifications. Finger and Bisantz [Finger and Bisantz, 2002] compared degraded icons (levels of blur) against the degraded icons with text probability, and full detail icons with text probability, for the task of hostile/friendly target classification with evolving target probabilities. They found that for their task the addition of text did not provide a statistical advantage, and that the degraded icons without text were in general better. As the number of uncertainty levels that need to be read are task specific, this should drive the representational requirements [Zuk et al., 2005].

In the field of Geographic Information Systems (GIS), which has been at the forefront of uncertainty visualization, frameworks have been put forth that recommend visual representations of the uncertainty based on the data and uncertainty types [MacEachren et al., 2005]. Even though their spatial considerations and constraints limit the general problem, there are still no accepted standards. For general visualization including reasoning, user and task considerations will drive the best way to create uncertainty visualizations. Some representations may be more natural for expressing uncertainty as meta-data such as opacity, fuzziness, and colour saturation [MacEachren, 1992, MacEachren et al., 2005], but when distinguishing different types of uncertainty, or for integration with multivariate data, these options may not be optimal.

Representations ideally should afford a set of methods and actions that allow one to proceed to a solution. Gigerenzer suggested that natural frequency representations of probability (whole numbers and ratios[6]) may have inherent cognitive support in the brain because posing conditional probability questions in the form of natural frequencies helped more people solve the problems [Gigerenzer and Hoffrage, 1995]. Rather than inherent cognitive support for natural frequencies, recent arguments and research have indicated

[6]For example, a natural frequency representation would be two out of the three times some event X occurred, versus the fractional probability representation of $P = 0.66\bar{6}, P \in [0,1]$

that the computational complexity of the problem solving process is a key determinant of a person's ability to find the correct Bayesian solution. This complexity is related to the cognitive transparency of the information structures, which for conditional probability problems may be nested-sets [Sloman et al., 2003]. This does not contradict the finding that natural frequencies may be an efficient representation for promoting Bayesian reasoning [Gigerenzer and Hoffrage, 1999].

Cognitive support may be given by providing uncertainty or ambiguity in representations to provide clues to potential representational transformations or new representations. Vague representations, such as sketches, may allow the deferring of design decisions, or may stand for a generalization (i.e. a set of designs) [Glasgow et al., 1995]. User interactivity in selecting the representation, while often difficult to provide in a visualization, implicitly communicates to the viewer that there is uncertainty in the optimal representation(s). At a meta level, visualizing your own reasoning process can also reveal a bias or suggest a new strategy. Representations of the reasoning process which illustrate uncertainty will help one perform this introspection.

## 4.4 Conclusions

We have described how the cognitive issues of reasoning under uncertainty relate to various aspects of visualization and provided some guidance as to how one may address these issues. As a result of the complexity and uncertainty in the reasoning process we see potential in the integration of data and reasoning visualizations. This integration of the discovery process and sense-making loops, would provide a direct visualization of the entire analytic process, and might facilitate the exposure of analytic gaps. Without this type of cognitive support monitoring the effect of uncertainty in the data and the analytic process will be extremely difficult.

When we create new support there is a potential hazard if the external visualization does not diminish cognitive load, it may in fact raise it, thereby preventing the formation of schemata [van Bruggen et al., 2003]. Therefore when the performance of sub-tasks require complete attention this level of integration may be more useful as an analytic context or an audit trail. Multiple views or the easy movement of reasoning artifacts between the two visualization systems could maintain this context without adding cognitive load. The visualization we briefly introduced (medical diagnostic support) illustrated that for some problem areas a reasoning component can exist as a natural and central component of the interface. As uncertainty abounds in the reasoning process we expect that visualization of the uncertainty will enhance problem-solving and decision making.

## 4.5 Acknowledgements

We would like to thank the Government Accountability Project for providing source material on the editing of *Our changing planet: The fiscal year 2003 U.S. global change research program and climate change research initiative* [Mahoney, et. al., 2003].

# Chapter 5

# Case Study in Archaeological Data

The long unmeasured pulse of time moves everything. There is nothing hidden that it cannot bring to light, nothing once known that may not become unknown.
– Sophocles (495 – 406 BCE)

Uncertainty visualization in the specific domain of archaeology is the first of the three domain explorations to be presented. For this domain I have tried to lessen the distance between myself and the practice, by taking a course, reading texts, attending departmental seminars, and participating on digs. Collaborative discussions with an archaeologist were also part of the methodology in order to understand more of the motivation and goals in this field.

This investigation will explore uncertainty visualization relating to specific needs of archaeology, and while it deals with spatial data, the emphasis is on temporal uncertainty. The uncertainty visualizations in this chapter also focus mainly on data uncertainty, rather than reasoning uncertainty. With archaeological site data in particular, the dating regularly has significant uncertainty. In this chapter we present an application that enables integrating and visualizing the temporal uncertainty for multiple 3D archaeological data sets of a single site with different dating. We introduce a temporal time window for dealing with the uncertainty and review various visual cues appropriate for revealing the uncertainty within the time window. The interactive animation of the time window allows a unique exploration of the temporal uncertainty[†].

[†]Portions of this chapter have been previously published in Zuk et al. [2005]. Therefore "we" refers to Torre Zuk, Sheelagh Carpendale, and W.D. Glanzman

## 5.1 Introduction

Uncertainty in various forms is prevalent throughout archaeology. Archaeological site data can be recorded in numerous formats ranging from hand drawn sketches to ground penetrating radar. All of the recorded data usually represents a minuscule fraction of the information regarding the visual appearance of a site over time and so missing data forms a major component of the uncertainty. Of the data that are available the dating regularly has significant uncertainty.

All archaeological data have a relative chronology value (for example, an artifact's placement within a stratigraphic sequence, or the addition of a wall to an existing building), and some data also have an absolute chronology value (for example, coins bearing mint dates, inscriptions mentioning an event during the reign of a certain ruler) that archaeologists can discern. In both conditions, dating must be thought of as representing either a span of time during which an event occurred, or a point in time before or after which an event occurred. Furthermore, many archaeological sites and their data sets are incomplete or disturbed, rendering their chronological value obscure. All chronology pertaining to archaeological data thus contains uncertainty.

This uncertainty should be integrated into any visualization to improve the cognitive task of spatiotemporal understanding. To aid in comprehension we present a time window for the animated visualization of the temporal uncertainty. We also analyze the applicability of various visual representations that may be appropriate for revealing temporal uncertainty in interactive 3D scene reconstructions.

### 5.1.1 Visualization

Often archaeological data is visualized at a specific time in the past. This can be categorized as a reconstruction, which when using computer graphics is often called a virtual reconstruction. This has been performed on ancient sites such as the Visir Tomb [Palamidese et al., 1993] up to the recent past with the Dresden Frauenkirche [Collins, 1993]. This methodology can even be extended into the future for illustrating models of restoration or deterioration.

Usually within an archaeological site, however, data are collected representing various periods of time. Site data are 3D spatial data acquired during an excavation but the dating of each of the artifacts is not as precise as the spatial location. The 3D position of an object represents either the final position of an artifact and thus its last probable location of use prior to burial, or it represents its original, intended location of use, and is thus *in situ*, in its original placement on a site. A decision must be made as to which location the viewer desires to visualize. Integrating the *in situ* object placement within a virtual reconstruction (of approximate object burial date) can help the archaeologist to visualize the use of an object, or hypothesize why the object came to rest in that position. Two examples of the visualization of last use locations relative to *in situ* architectural reconstructions, are the location of bifaces, scrapers, and debitage[1] within a prehistoric pithouse [Peterson et al., 1995], and lamps and coins inside the Great Temple of Petra [Acevedo et al., 2001].

Reconstructions and their integration with archaeological site data may allow more accurate hypotheses to be made. Virtual reality can allow the archaeologist to understand the past context of the 3D spatial layout of their data [van Dam et al., 2000]. When using a 3D model various lighting or sky/star models can be applied to test other theories as well. For example, would a certain location within a building have adequate natural lighting for the inhabitant to perform a specific task? All of these techniques can provide valuable new tools to aid in interpreting the data.

Using the computational power of current consumer level computer graphics technology, interactive animation of complex 3D scenes is now possible. The animation of time provides a powerful visualization which allows complex 3D spatiotemporal changes to be compared in a natural way. Currently most archaeological visualizations represent

[1] Bifaces and scrapers are primitive tools and fragments are debitage.

spatially static scenes of a speculative nature that represent specific time periods. The following discussion will outline how to extend this type of visualization by adding increased comprehension of the temporal changes and uncertainty using interactive animation.

## 5.2 Time Windows and Interactive Animation

Any artifact or structure may have an estimated timeline based on a creation and destruction date (the destruction may be in the future). Using these dates the 3D scene for a specific date, or an animation frame, can be constructed by simply finding which data sets have a timeline that overlaps the viewing date. However the overlap will be influenced by the uncertainty in the creation and destruction dates. Uncertainty in these dates may be statistical such as from dating technology, or more abstract such as when based on scientific judgment [Renfrew and Bahn, 2000]. This judgment may consider things such as the likelihood of contamination or just be an expert estimate based on seriation[2].

### 5.2.1 Time Windows

The computer generation of an animation frame may use the photorealistic rendering analogy of the shutter speed of the camera taking the picture. This allows effects such as motion blur to be recreated for moving objects, or a moving camera, by sampling the view repeatedly (while the shutter is open) and then blending the pixels together. In our context we suggest that the frame (viewing time) also take into account temporal uncertainty.

In expanding the camera shutter concept to a much larger timescale we create a time window. This larger duration allows events on either side of a specific date to be captured to take into account uncertainty in the actual viewing date. It can provide a visualization to help in answering a question like: what would a person have seen if they visited the site between 200 and 210 BCE? Arbitrarily expanding the time window also enables the viewer

[2] Seriation is a form of relative chronology based on associations [Renfrew and Bahn, 2000].

to see how later and earlier construction relates in an intuitive way. The time window could also be interpreted reciprocally giving all artifacts temporal uncertainty equal to half the time window.

The time window is illustrated in Figure 5.1. The time window's width, or range of time, can be controlled by the user. This window of time allows data that comes within range of the viewing date to be visualized in some way. The time window allows two different types of uncertainty to come into play: the uncertainty in the artifact dating (e.g. deposited between 85 BCE and 20-30 CE), and the uncertainty over the time window (e.g. "around" 22 CE).

Either the time window or timeline uncertainties can be mapped to probability density functions or other schemes. As an example, for the time window the centre can be thought of as absolute certainty, equal to a probability of one, and then certainty can drop off based on a function (e.g. Gaussian) to where one does not want to consider information from that date at all, probability of zero, at each end of the time window. For the time window alone the uncertainty for an object would be the maximum certainty function value that the object timeline overlaps. These certainty functions over the time window and timelines can be used independently or combined. The uncertainty measures can then be used to create visual representations that depict various levels of uncertainty other than the obvious inclusion or exclusion from the scene.

### 5.2.2 Interactive Animation

Archaeological animations often are restricted to a specific reconstruction date and provide a fly-through or a virtual reality experience [Forte and Siliotti, 1997]. In some cases a partially interactive animation over time is created [Vergauwen et al., 2004], but these do not include uncertainty. In these scenarios the rendered frame represents a small window in time (usually infinitely small) in contrast to our time window concept.

As time is experienced in a continuous and unstoppable manner, it is natural to want

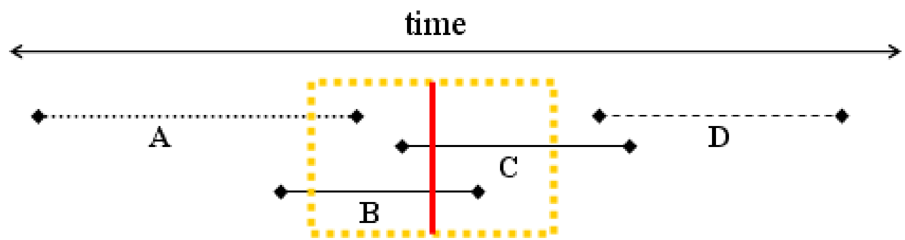



Figure 5.1: Time Window. Segments A, B, C, and D represent data sets and their timelines. The line down the centre of the box represents a specific viewing time, and all data sets that overlap this time are displayed normally (B & C). The dotted-line box extends the standard viewing time to form a time window. The data sets that only overlap the time window and not the viewing time may be rendered in a way to indicate uncertainty (A). All data sets outside the time window would not be displayed (D).

to explore time interactively. We provide a graphical user interface in the form of a slider to allow the user to directly control the temporal position of the time window. By manipulation of the time slider and time window the user can create an interactive temporal animation either forward or backward in time. The user controllable animation along with uncertainty visualizations may provide better temporal comprehension.

## 5.3 Visual Representations

Given an uncertainty metric there are numerous ways to render a 3D artifact within a scene to express the uncertainty. We are concerning ourselves only with uncertainty in time while ignoring the uncertainty in the other dimensions. Obviously the uncertainty in spatial position is relevant, and is temporally dependent, as with the Arrigo VII funerary complex reconstruction [Baracchini et al., 2004], but was not the focus of this project. We are also limiting our discussion to visual integrations into a standard 3D virtual reality scene that can be intuitively understood. Honouring these restrictions creates a visual 3D scene rendering that is compatible with normal virtual reality systems and only slightly reduces the options for uncertainty representations.

Non-photorealistic rendering (NPR) methods have been shown to be able to depict uncertainty required to express speculative designs or constructions [Strothotte et al., 1994, Strothotte and Schlechtweg, 2002]. Strothotte et al. [1999b,a] reviewed aspects of non-photorealistic rendering and how they can be used in representing uncertainty in virtual reconstructions. They show how sketch-like renditions and the use of variable transparency can express the speculative nature of archaeological reconstruction. They also found that photorealistic detail distracted from the fundamental questions of the domain experts. They conclude that more methods of visualization and interaction are required for expressing the appropriate level of uncertainty. Practical aspects of an implementation using these techniques were presented by Freudenberg et al. [2001]. Roussou and Drettakis [2003] have discussed photorealisitic rendering, NPR, and interactivity, and found they all have an important role in the perceived realism.

Reusing the camera shutter analogy and sampling the scene over the time window (and including data timeline uncertainty) generates the visualization. While it would be appropriate to integrate the certainty over the time window, we simply used the maximum certainty in the time window. If the maximum certainty of an artifact was 0.2 as a probability then the opacity could be set to 0.2 to provide a similar effect to motion blur if the object was removed after $2/10^{ths}$ of a frame. Where spatially incompatible artifacts occupy the same space they will intersect each other.

### 5.3.1 Visual Cues

A visual cue can be defined as any visual encoding (colour, size, animation, etc.) and may be used to communicate meta-data. Arbitrary visual cues beyond the motion-blur (accumulated opacity for our purposes) from the standard camera shutter model move us into styles of non-photorealistic rendering. In the current context a visual cue is any visual encoding used to distinguish levels of uncertainty. Some visual cues may be applied to a single artifact while others may cover the entire scene. This change in application can affect how it is perceived. For example if fog is applied to only a single object, it will be perceived as colour blending, similar to a colour saturation cue, rather than environmental fog. Visual cues may also be overloaded in that they have implicit meanings beyond their use as a representation of uncertainty. This is true for cues such as fog and blur/depth-of-field [MacEachren, 1992, Kosara et al., 2001], as a virtual reality rendering may already use these as depth cues [Ware, 2004] (visual encoding of the distance to objects in a scene).

In Pang et al.'s [1997] survey of uncertainty visualization there are numerous applicable methods including: side-by-side views, pseudo- colour, contour lines, blinking, material properties, texture mapping, bump mapping, oscillation, displacement, and blur. They categorize methods for visualizing uncertainty into the groups: add glyph, add geometry, modify geometry, modify attributes, animation, sonification, and psychovisual. We introduce a cue into Pang et al.'s animation category with the use of a rising/sinking animation during continuous time changes (a form of displacement). The rising/sinking animation provides a natural transition animation similar to that of time-lapse photography of construction. A drawback of the rising/sinking cue is that it may be misinterpreted in a static scene.

The two visual cues of transparency and the rising/sinking animation are used to illustrate the time window technique for presenting the uncertainty. Figure 5.2 shows, for simplified illustrative purposes, data sets of single photographs with specific dating assigned matching the photograph's contents. The photographs represent a series of sites which exist at the current time. They are the Giza Pyramids, the Rammaseum, and the Kiosk of Qertassi near the Temple of Kalabsha. The figure shows three snapshots of the window containing the 3D scene view and time slider view. The uncertainty based on the relative position of a timeline in the time window is visible in the top two images. The timeline of each data set (photograph) is shown in a different colour and from top to bottom and corresponds to the photos from left to right.

Visual cues may be classified on various attributes from perceptual to practical. Bertin's

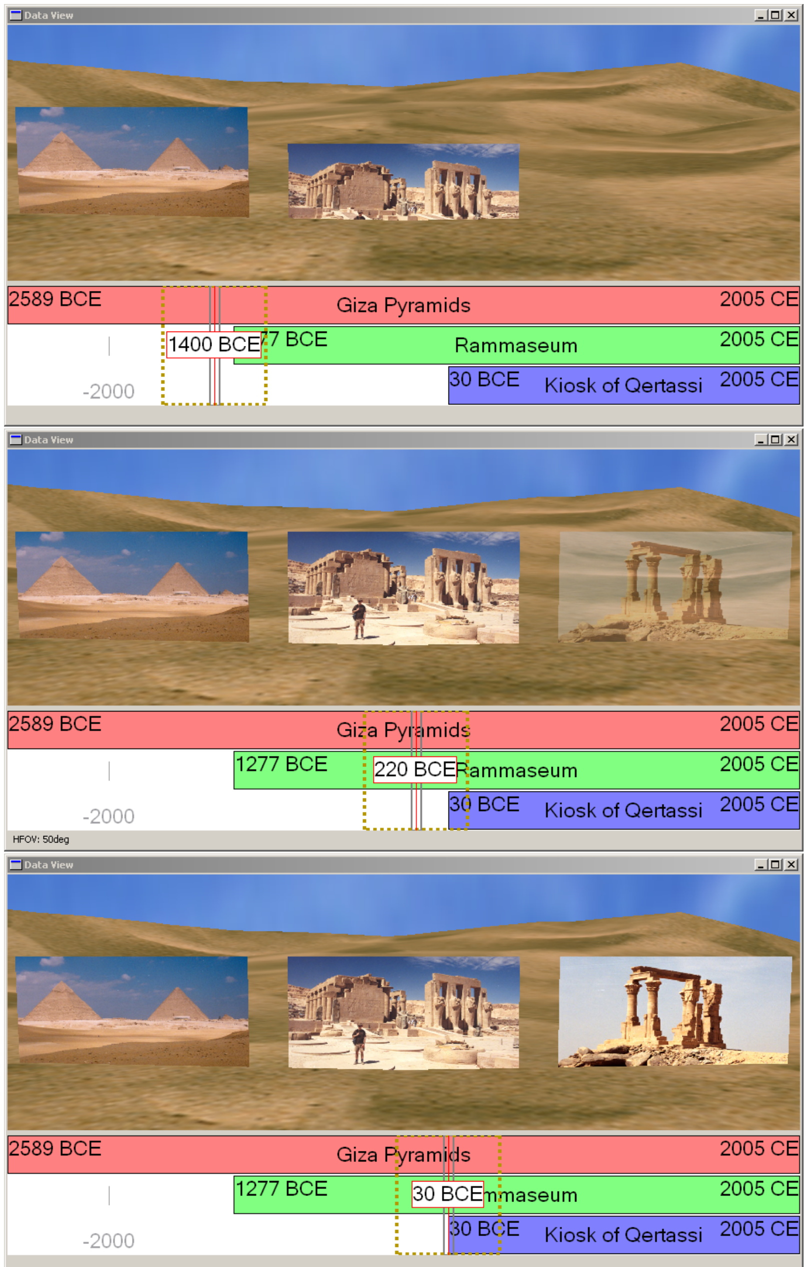



Figure 5.2: Illustration of uncertainty cue animation. Viewing dates (frames) from top to bottom of 1400, 220, and 30 BCE respectively. Time window constant at 300 years. Top image shows rising/sinking cue, middle image transparency, and bottom image no uncertainty.

Table 5.1: Visual Cue Characteristics

| visual cue | length | order | artifact/scene | GPU |
|---|---|---|---|---|
| transparency | small | Y | artifact | Y |
| colour change | medium | N | artifact | Y |
| wireframe | 2 | Y | artifact | N |
| line style (NPR) | large | N | artifact | Y |
| shading/hatching (NPR) | large | Y | artifact | Y |
| floorplan only | 2 | N | artifact | N |
| rising/sinking | large | Y | artifact | N |
| animated warping of surfaces | medium | N | artifact | Y |
| blur | small | Y | artifact | Y |
| fog/haze | small | Y | scene | N |
| rain/snow | medium | Y | scene | Y |

[1983] framework called the Properties of the Graphic System classified visual variables (which often may be used as cues) on the basis of their characteristics such as the potential for immediate perceptual group selection, natural perceptual ordering (not learned), ability for quantitative comparisons, and length (the number of discernible elements that can be represented in the set, i.e. cardinality). A summary of some visual cues appropriate for 3D rendering and relevant characteristics (including Bertin's length and order) are presented in Table 5.1. The table also indicates whether direct programming of the graphics processing unit (GPU) would be advantageous, and this will be discussed in more detail in Section 5.5.2. The practical length of a visual cue depends on the visual size of the rendered artifact in the frame and so the categories of small, medium, and large, are relative generalizations.

## 5.4 Implementation

Our application, ArkVis, was developed for visualizing 3D archaeological data along with their temporal uncertainty. ArkViz allows the user to import multiple 3D data sets and assign various properties to them. The most important of these properties are the dating, or creation and destruction dates, of the physical artifacts or structures composing a data set. Uncertainty may be assigned to each of these dates.

The data may be interactively viewed in a 3D perspective scene. The user selects a date using the time slider and a scene is automatically generated representing the scene of the archaeological site at the given time. The user may also drag the time slider to create a temporal animation. Once a scene is constructed for a specific time window, ArkVis allows the user to navigate (walking or flying) through the site at that specific time in history. They may also interactively manipulate the time window to provide a larger or smaller portal into the near future and near past. Various visual cues for the temporal uncertainty of the data may be selected interactively.

The time window may be shifted along with the time slider or may be specified by directly drawing it. As the concept of vagueness is often tied to uncertainty we also provide the approximate input of values by allowing the time window to be "sketched" out. This process is shown in Figure 5.3.

ArkVis was written in C++ using Microsoft's Visual Studio. Trolltech's Qt library was used for windows and widgets. The 3D scene and visual cues are rendered using OpenGL and Nvidia's Cg language for GPU programming. Model loading was based on Lischke's [2005] 3DS import library, and the sky rendering was based on Sempé's [2005] sky demo.

## 5.5 Results

Archaeological data recorded for the Maḥram Bilqīs sanctuary complex in Mārib, Republic of Yemen [Glanzman, 1998, 1999, 2002] has been used to illustrate the system. The most recent spatial data is of the main oval wall of the temple, provided by a recent survey taking accurate measurements. This data represents a structure deteriorated by looting and time. The earlier data is a theoretical reconstruction of the site at an early date, derived

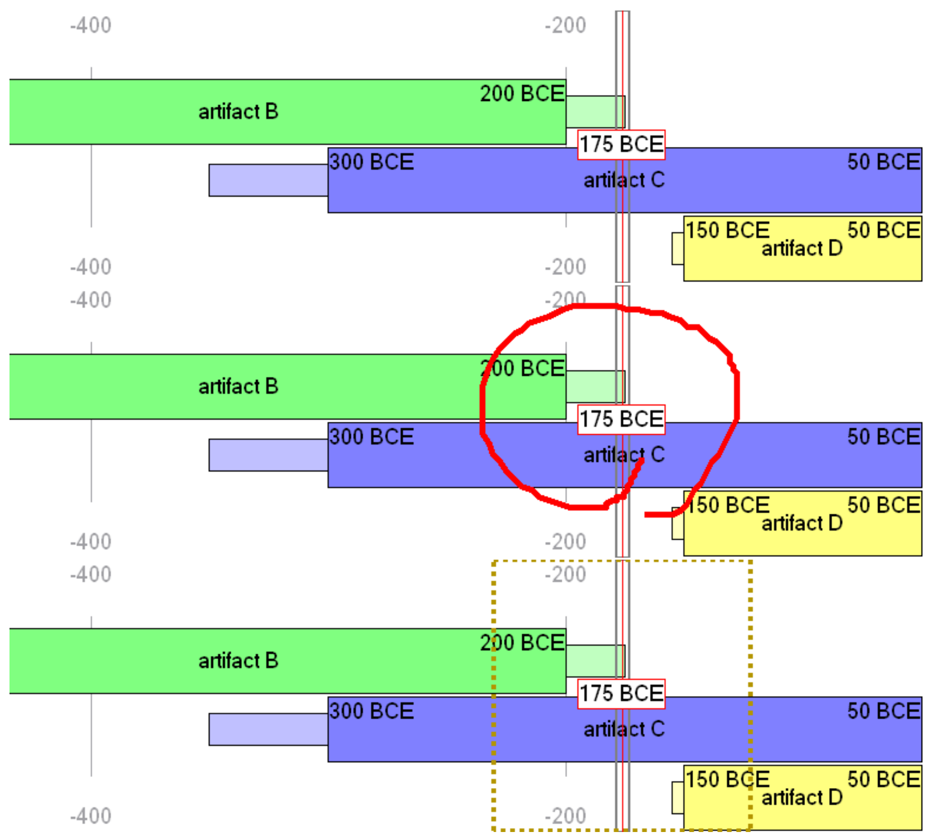



Figure 5.3: Approximate time window specification. Top image: no time window only artifact C visible. Middle image: approximate time window specified with mouse input. Bottom image: new time window based on roughly guided input in which artifacts B and D would be visible but could be rendered with visual cue of uncertainty. Timeline boundaries with uncertainty are indicated by smaller sized extensions with lower colour saturation.

from Albright's [1952] published data. These two data sets are compared using different visual cues in Figure 5.4. Interactive animation provided by the time slider and time window allow smooth transitions between the two data sets. This along with the uncertainty visualization may allow the user to more easily understand the assumptions in the earlier theoretical data set.

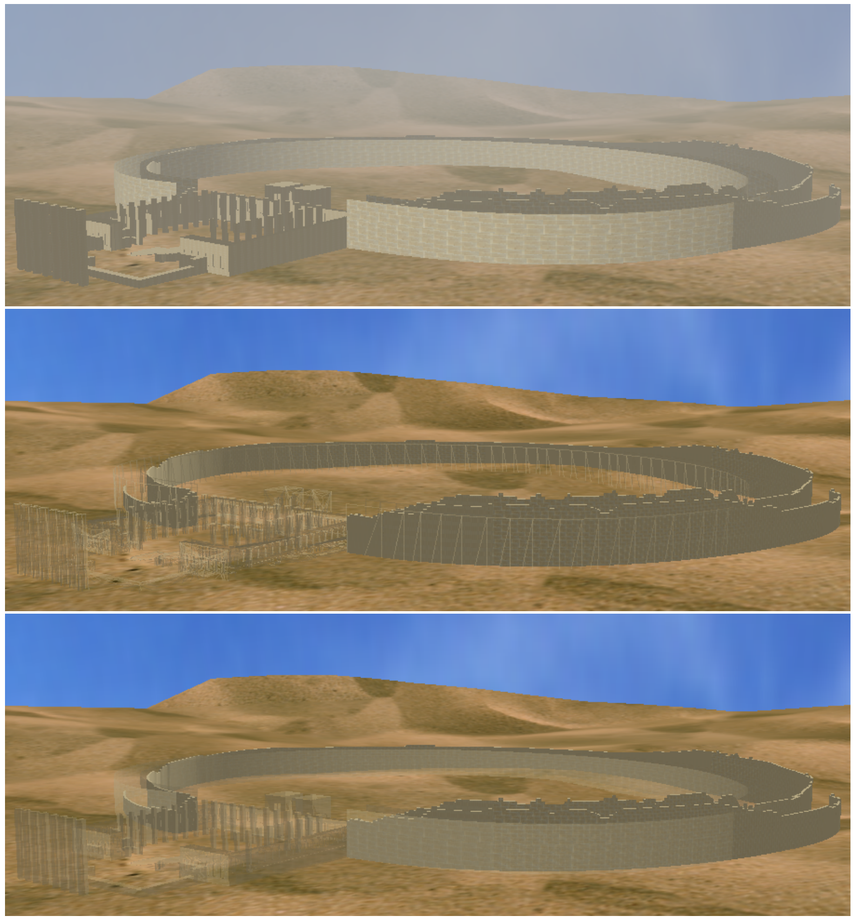


Figure 5.4: Juxtaposition of theoretical reconstruction and survey data. Top image: both data sets with scene haze and no data set uncertainty cues. Middle image: wireframe and transparency uncertainty cue for theoretical reconstruction. Bottom image: transparency uncertainty cue for theoretical reconstruction.

### 5.5.1 Uncertainty Tasks

While simply visually revealing whether there is uncertainty (at the Boolean level), can clearly be achieved, it is not clear what representations are most appropriate for specific tasks. Most of the uncertainty cues in Table 5.1 have a length above a Boolean indicator, but they may not be appropriate for some tasks, or may lead to confusion. For the task of simply eliciting possibilities, however, most of the cues should work.

Amar and Stasko's [2004] general *Rationale-based Task* category of *expose uncertainty* requires both the presentation of the uncertainty and showing the possible effect of the uncertainty on outcomes. Uncertainty cues such as transparency and wireframe directly allow the possible effects on outcomes to be seen, as the user can ignore the data and consider that it did not exist at that time. Once uncertainty is revealed simply providing interactive toggling of a data set also affords this.

Kirschenbaum and Arruda [1994] found that for some spatial problems a graphical representation of uncertainty may improve the judgments of decision makers. We hypothesize that this would also apply to spatial decisions that must account for temporal uncertainty. Future work could determine the cognitive tasks and set of applicable visual cues that could be used to test this hypothesis. For example, assuming Cohen et al.'s [1996] cycle of metarecognition was applicable, then the time window could provide visual queries to aid in the testing of incomplete, conflicting, and unreliable information.

### 5.5.2 Interactive Rendering Considerations

When the time slider is used to create an animation, on each sequential frame the time window moves and so the temporal uncertainty may change for all data sets. The data for a virtual reconstruction may be very large even before adding multiple temporal versions. Therefore any procedural rendering method can reduce resource requirements by simply modifying the single representation of each data set during the rendering process. As interactive animation is required, using the graphics processing unit to its full potential is desirable.

The uncertainty visualization method categories of modify geometry, modify attributes, and animation [Pang et al., 1997] are highly suited for interactive graphics. Using graphics processing unit (GPU) programs to perform procedural rendering, one can work with a single representation of the scene and directly modify the visual appearance based on the uncertainty metric (e.g. transparency can be changed without modifying the model

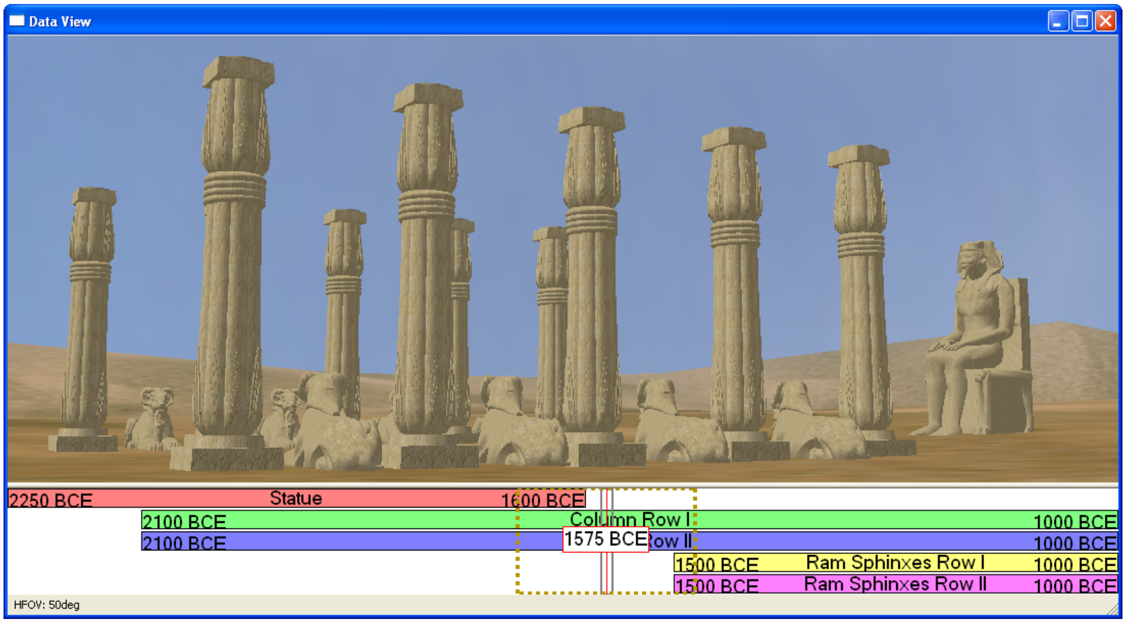



Figure 5.5: Simulated archaeological reconstruction. Rendered with scene haze. No data set uncertainty visualization.

attributes). The uncertainty value assigned to each data set can also be used to determine when a different GPU program is used (e.g. to provide a sketch-like quality).

### 5.5.3 Visual Cue Discussion

We have simulated an ancient Egyptian archaeological site to more clearly demonstrate some visual cues for temporal uncertainty. The site is shown with its associated data timelines in Figure 5.5. This site contains different dating for the columns, sphinxes, and the main statue. Various visual cues are illustrated for the specific viewing date of 1575 BCE and a time window of 100 years (both the statue and sphinxes are uncertain with this temporal configuration) in Figures 5.6 and 5.7.

Cues implemented using standard OpenGL are usually efficient but have limitations. To achieve correct transparency effects with OpenGL (or any Z-Buffer depth sorting) one must ensure that transparent data sets are rendered last and in back to front order. While

Figure 5.6: Uncertainty cues. From top to bottom: no cues, rising/sinking cue, wireframe, and transparency.

this can easily be done at the object (artifact) level it is not usually interactively feasible at the polygon/pixel level. Therefore basic OpenGL transparency is not guaranteed to provide accurate results with complicated objects and scenes. The wireframe cue also has its drawbacks as it may be misleading. Wireframe rendering reveals much of the underlying polygonalization and so is dependent on the manner in which the object was

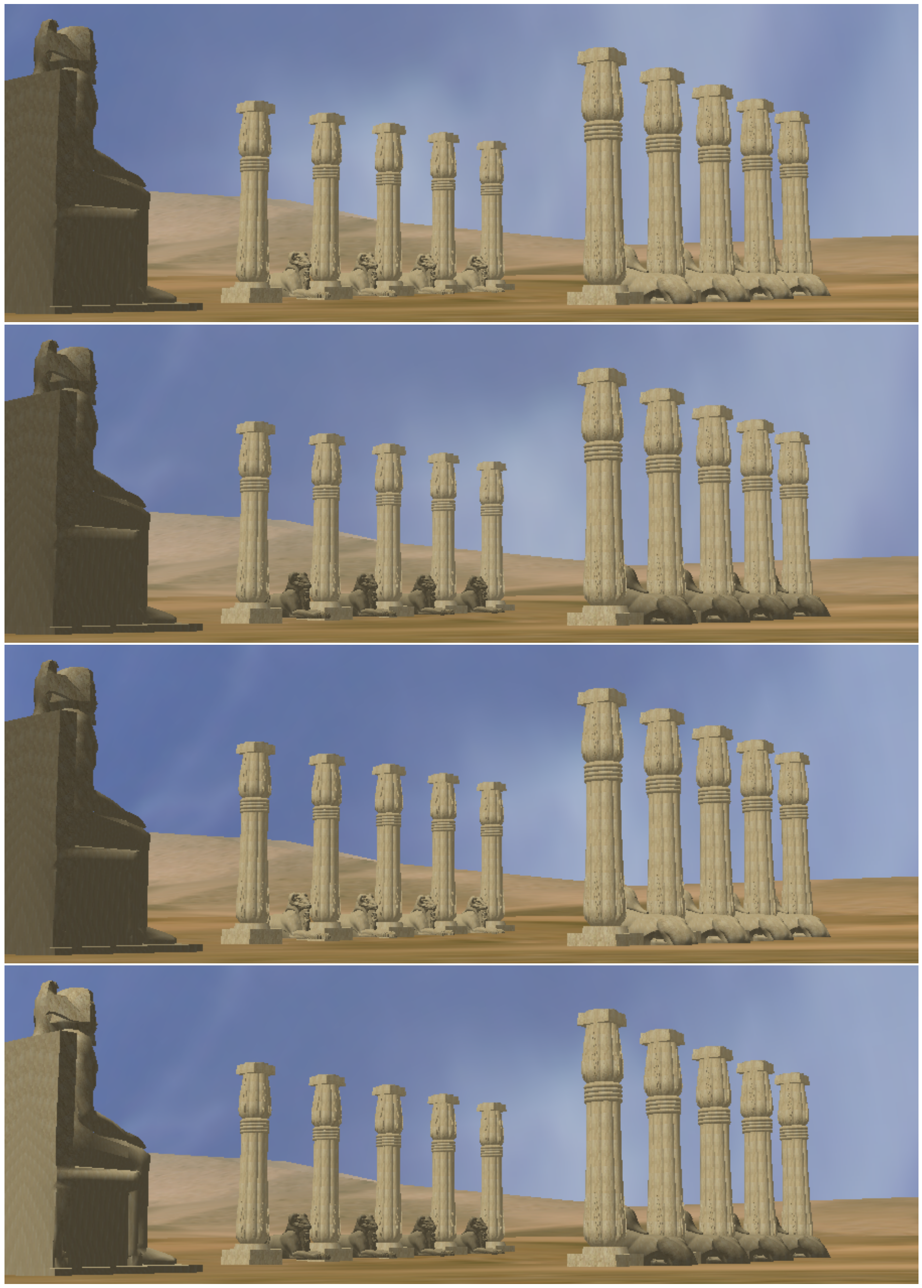


Figure 5.7: Animated shading uncertainty cue (GPU program) with temporal sequence from top to bottom. Uncertainty controls the presence and frequency of lowered lighting conditions. Higher uncertainty has higher frequency and so the sphinxes are in and out of shadow more often than the statue.

created. It may be preferable to determine the silhouette and crease edges of the objects in the data sets and only display those as lines. To do this we could utilize techniques similar to those of technical illustration presented by Gooch et al. [1999]. It may also be possible for the modeller to design objects so that they provide a suitable look when rendered in wireframe mode.

Each visual cue will have its own benefits and drawbacks. Visual cues that can be created using GPU programs benefit from increased flexibility (they are not bound by the fixed OpenGL rendering pipeline) and potentially faster performance. Those that are more intuitive will be more accessible to people in general (e.g. transparency, fog). More complex cues may requiring learning, but then may allow domain experts to express multiple types of uncertainty. Determination of which cues are the most appropriate to use will depend on task and hardware considerations.

## 5.6 Heuristic Evaluation

Applying the heuristics from Table 3.1 (presented in Chapter 3) we perform a light-weight evaluation of the ArkVis visualization system. However, with ArkVis, various encodings may be explored and new ones added, thus the heuristics may be the most relevant in guiding the selection of encodings for particular tasks and for informing the design of new uncertainty representations to be integrated into the system. We now discuss the application of each heuristic in turn by first stating the heuristic and then noting its relevance to ArkVis:

- **Ensure visual variable has sufficient length** – The length of different encodings was discussed in relation to ArkVis in Table 5.1. Interactive selection of the encoding allows the choice of an encoding with sufficient length.
- **Preserve data to graphic dimensionality** – Temporal uncertainty is encoded in ArkVis by modifying attributes of the spatial dimensionality but we do not consider

this contrary to the heuristic. This heuristic may serve to guide the design of novel uncertainty encodings which encode the temporal uncertainty separate from the spatial data. As an example, the timelines shown in Figure 5.3 encode the uncertainty in dating using the width of the smaller horizontal bars.

- **Put the most data in the least space** – Integrating uncertainty directly into the virtual archaeological site is in agreement with this heuristic in that it adds the uncertainty data without increasing the space used.
- **Provide multiple levels of detail** – Uncertainty encodings may be toggled on and off in the virtual environment view, which provides interactive access to one form of level of detail. Precise temporal uncertainty encodings are also provided in the time slider interface. This heuristic suggests providing additional spatial levels of detail may be beneficial. One form that would be appropriate would be a map-like overview of complex sites.
- **Remove the extraneous (ink)** – This heuristic suggests encodings such as wireframe may benefit from further reduction down to silhouette edges or other minimalistic sketch-style renderings.
- **Consider Gestalt Laws** – These influences may need to be assessed based on the interplay of any particular uncertainty encoding and specific archaeological artifacts and scenes in ArkVis. For example, a new uncertainty encoding may potentially change the perception of *figure and ground*, thereby shifting attention.
- **Integrate text wherever relevant** – Text feedback with the precise dating information is provided as a popup window (tooltip) in the time slider based on the cursor position. Further textual information should be added directly into the virtual environment.
- **Don't expect a reading order from colour** – Complying with this heuristic, none of the default encodings for uncertainty in ArkVis utilize colour variation to encode

different levels of uncertainty. If adding new encodings into ArkVis for a public display, this heuristic advises that colour encoding may not be suitable to encode the uncertainty beyond a boolean level.

- **Colour perception varies with size of coloured item** – As ArkVis presents the scene in a perspective projection even identical objects may appear at different sizes when at separate locations. Thus this heuristic warns against relying only on colour to encode the uncertainty. Alternative encodings to colour, such as the predefined encodings in ArkVis, should be used if many objects of differing sizes are to be compared.
- **Local contrast affects colour & gray perception** – This heuristic again warns that precise readings from colour or grayscale should not be expected. The visual context of archaeological objects in the ArkVis scene need to be considered when adding or choosing these types of encodings. For example, if the ArkVis scene utilizes haze or fog, there will be contrast effects related to viewing distance.
- **Consider people with colour blindness** – This heuristic should be considered a requirement for cultural heritage displays. The various encoding options in ArkVis easily support this.
- **Preattentive benefits increase with field of view** – Arbitrary fields of view are supported in the ArkVis virtual environment and with large displays preattentive encodings may be superior for some tasks. For example, preattentive encoding may facilitate changes in uncertainty to be monitored over large fields of view as time is animated, and thus allow for faster site analysis.
- **Quantitative assessment requires position or size variation** – The time slider utilized size encoding for uncertainty. In the virtual environment both position and size are already required for the spatial encoding of the site, however, as illustrated the rising/sinking cue allowed a positional encoding of uncertainty. Quantitatively decoding the rising/sinking cue would require knowing the full appearance of an

artifact, which may be available by animating to a point in time when there is certainty regarding it. Additional representations which use size, such as uncertainty error bars, could also be integrated in a mixed reality manner.

In summary, the heuristics are appropriate to not only the uncertainty components, but can be related to the visualization system as a whole. In this scenario they can additionally serve as guidelines for configuring the visualization during use.

## 5.7 Conclusions

We have described a method, an interactive time window, and an application, ArkVis, that provides visual support for cognitively merging multiple data sets that represent different periods in time. In ArkVis after importing and entering minimal information a scene can be navigated arbitrarily in time and space. By controlling the time window, data from non-overlapping periods in history can be spatially integrated with user selectable visual cues revealing the uncertainty. The animated time window is intended to provide a new look at the progression of time at an archaeological or cultural heritage site.

Visualizations of archaeological and cultural heritage sites serve two distinct user groups: the general public, and domain experts. They can be useful to the general public in providing comprehensible visual explanations and to domain experts by allowing them to explore their data in new ways. While NPR renderings may better serve the cognitive tasks such as hypothesis building [Strothotte et al., 1999a], some tasks may benefit from other types of rendering that may illustrate another person's conceptualization [Roussou and Drettakis, 2003]. For example, at a museum a photorealistic rendering style may best help people conceptualize that an ancient site was a living community. Interactive animation that can allow the user to select the type of rendering style provides the most flexibility.

Similar to problems observed with photo-realistic drawings used in preliminary drafts of architecture [Strothotte et al., 1994], the clean data sets provided for theoretical reconstructions often give the false impression of accuracy and completeness. They may give a viewer the impression that this is exactly how it did look, even though a large portion may be artistic interpretation. Therefore we feel it is important to give the same regard to temporal uncertainty as spatial uncertainty. We hope that the visual differences revealed via uncertainty cues, which allow the controlled comparing and contrasting of data from different times, as well as different sources, can provide new insights, thereby providing an improved understanding of the past.

## 5.8 Acknowledgements

We would like to thank William Glanzman who supplied access to site data and information on the Maḥram Bilqīs sanctuary complex.

# Chapter 6

# Case Study in Geophysical Modelling

To believe with certainty we must begin with doubting.
– Stanislaus I (1677 – 1766)

In the seismic domain we developed two separate uncertainty visualizations for 2D bi-directional vector fields: one based on animated flow and the other based on a static glyph. These visualizations were designed for the task of interpreting and understanding anisotropic rock property modelling in the domain of seismic data processing. Aspects of the implementations are discussed relating to design, interaction, and tasks.

This forms the second of the three different domains to be examined, and is quite distinct from the first domain as it examines volumetric model data. Again the focus is on data uncertainty, but we will return to reasoning uncertainty in the next chapter. This work involved collaboration with domain experts whose geophysical modelling results were being visualized. It also benefited from my own experience having being active in this field for many years[†].

## 6.1 Introduction

While visualizing both the data and its associated uncertainty has been accepted as beneficial for accurate interpretation, the integration of uncertainty information into an existing or new visualization is not standard practice. The practical tasks of maintaining ease of comprehension for both the data and the uncertainty are not straight forward. Hence, in building uncertainty visualizations there still exist many challenges, such as finding good

[†]Portions of this chapter have been published in Zuk et al. [2008]. Thus any use of "we" may refer to Torre Zuk, Jon Downton, David Gray, Sheelagh Carpendale and JD Liang.

representation of errors and uncertainty for 3D visualizations [Johnson, 2004] and understanding how knowledge of information uncertainty influences analysis [MacEachren et al., 2005]. As a result, even choosing an initial design may be difficult.

We provide two new visualizations for bi-directional vector fields with their associated uncertainty. Building upon the image/texture based flow visualizations of van Wijk [2002] and Jobard et al. [2002] and uncertainty extensions of Botchen et al. [2005] we allow user driven exploration of the uncertainty in directionality, orientation, and magnitude. Our method utilizes the GPU to achieve interactive flow visualizations and intuitively handles the ambiguity of bi-directionality as well as orientation and magnitude. We provide a probe for interactive querying of the flow field which affords user controllable directionality enabling a visualization of possible realizations while at the same time revealing the certainty.

With similar goals to Wittenbrink et al.'s [1996] work we also provide a new uncertainty glyph. The new glyph can provide bi-directionality and uncertainty information for orientation and magnitude in a dense field. We describe the interactive controls over both the form and presence of the glyphs that are created on slice planes in a 3D volume. Examples showing of both of our types of visualizations are shown in Figures 6.1 and 6.2.

## 6.2 Related Work

MacEachren [1992] identified tasks related to uncertainty and also proposed what might be appropriate encodings for uncertainty. However, encoding single or multiple types of uncertainty in a way that can enhance interpretation is still a difficult problem. While considering all the variety of theory and representations one should begin by looking what has been already been developed for similar purposes. Therefore, in this section we will review other visualizations for uncertainty in vector fields and flow and summarize some related evaluation work.

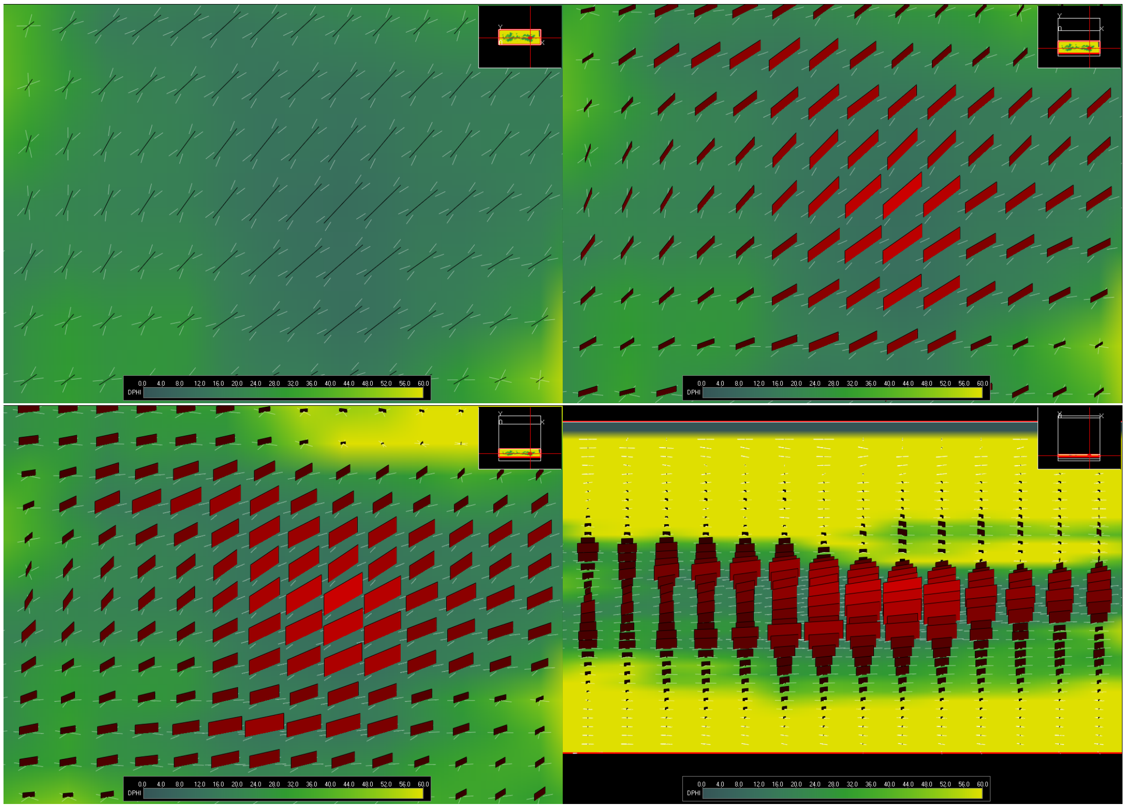


Figure 6.1: Sequence of viewpoints showing rotation of plane at constant time slice with the uncertainty glyph. Inset in top right of images shows the context of the cutting plane in the entire volume extent (animation goes from left to right then top to bottom).

### 6.2.1 Vector Fields and Flow

As vector fields can be used to create flow fields, flow may be a natural, or more “realistic”, representation as it is less abstract. Providing both abstract and realistic representations may benefit users who have trouble conceptualizing the model. Various methods for visualizing the uncertainties in flow and vector fields have been proposed. Adding uncertainty into a visualization complicates the information decoding process for the user, as the additional data is not an independent variate. Similarly, the task is also changed in that it may call for the weighing or modification of interpretations of the visualization without uncertainty.

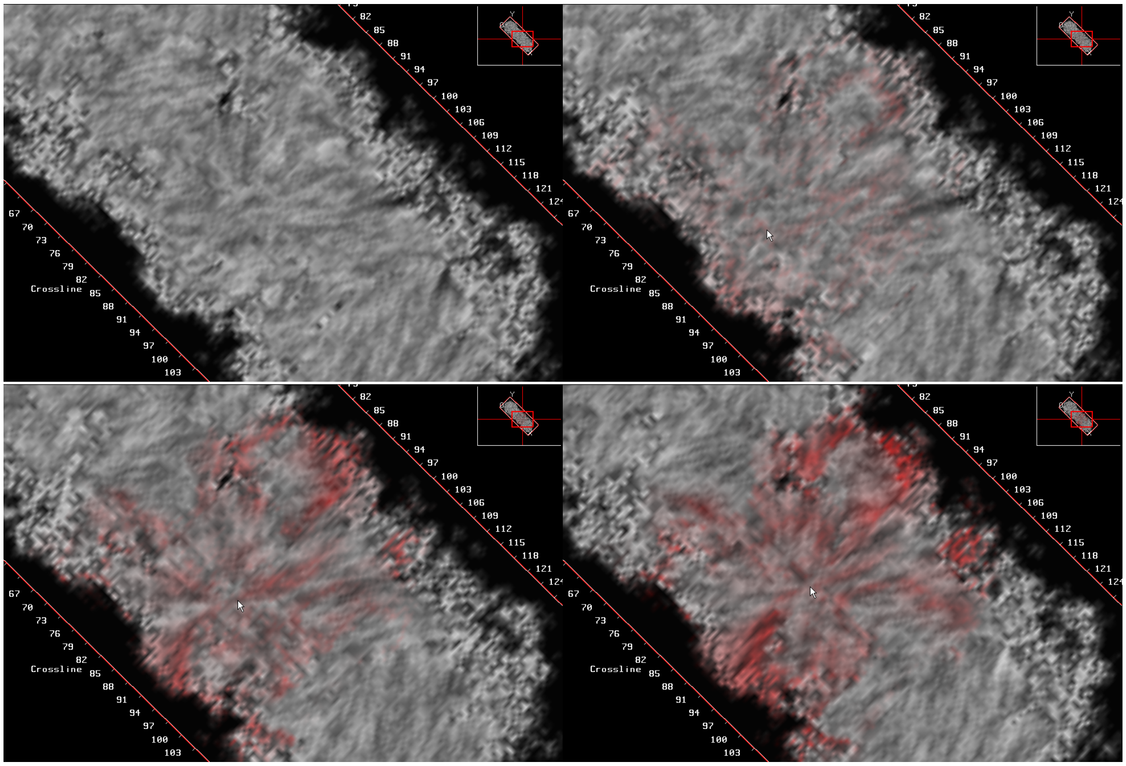



Figure 6.2: Sequence showing flow changes based on query probe (cursor) movement, with uncertainty in direction revealed using red dye. Flow is redirected directly to the cursor if within a fixed tolerance. Inset in top right of images shows the context with the small red rectangle indicating the current field of view (animation from left to right then top to bottom).

Wittenbrink et al. [1996] introduced what they called verity visualizations as a design recommendation for vector field uncertainty glyphs. Verity visualizations used representations of uncertainty integrated with the data without the use of additional graphic variables (e.g. colour, value, ...). This design provided vector glyphs that holistically show uncertainty in magnitude and orientation. Their evaluation showed the verity uncertainty glyph could be decoded with similar error to an arrow glyph decoding, and allowed for uncertainty information decoding (with comparable error).

Lodha et al. [1996b] provided a visualization system for the uncertainties found in flow fields. Variations included glyphs, envelopes of trajectories, and other representations from a stream-line or particle point of view. Sanderson et al. [2004] created a method for visualizing vector fields while potentially encoding uncertainty by using a reaction-diffusion model. Botchen et al. [2005] introduced some novel variations of cross-advection/diffusion and multi-frequency noise to depict the uncertainty in flow using the GPU to realize interactive rates. The uncertainty visualizations we introduce for bi-directional vector fields build upon this previous work with glyphs and flow representations.

### 6.2.2 Evaluation of Designs

User studies may often be appropriate for measuring a specific visualization's performance for a specific task, but it may be difficult to generalize beyond the specific tasks and visualizations that are evaluated [Kosara et al., 2003]. Laidlaw et al. [2005] evaluated static 2D vector field visualizations based on three tasks from fluid mechanics: locating critical points, identifying their type, and predicting particle advection. In comparing time and error measures for six different static visualizations (grid and jittered arrow placement, icons with artistic layering, line integral convolution (LIC), image-guided streamlines/integral curves, and streamlines seeded on a regular grid) they found superiority in: image-guided streamlines for advection tasks, LIC for location tasks, and streamlines seeded on a regular grid for critical point classification. While the LIC performed at the top for location tasks it was at or near the bottom for the advection and critical point type determination tasks, probably due to the ambiguity of direction [Laidlaw et al., 2005]. As the other five visualization types had direction encoded we feel that a more fair comparison would have been against animated flow, although comparing static and animated methods has its own set of problems. We hypothesize, however, that animated flow (e.g. streak-lines) could potentially be at the top for all their measures.

## 6.3 Seismic Domain: Data and Uncertainty

This case study deals with data and tasks relevant to the seismic industry which we will briefly introduce. Seismic wave azimuthal amplitude variation versus angle of incidence has proven to be useful in characterizing fracture distributions and direction for hydrocarbon reservoirs. Downton and Gray [2006] describe a Bayesian process for determining the geological model parameters of anisotropy gradient ($B_{ani}$) and horizontal transverse isotropy symmetry orientation ($\Phi_{iso}$), which are related to rock fracture density and orientation respectively.

The general process to estimate these parameters is driven by the changes in seismic wave amplitude variation over multiple orientations versus the wave incidence angle on a reflector. In their process the determination of uncertainty in $B_{ani}$ and $\Phi_{iso}$ requires joint probability distributions to be marginalized based on the integrals

$$P(B_{ani}) = \int_0^{2\pi} g(B_{ani}, \Phi_{iso}) d\Phi_{iso} \tag{6.1}$$

and

$$P(\Phi_{iso}) = \int_0^{2\pi} g(B_{ani}, \Phi_{iso}) dB_{ani}. \tag{6.2}$$

As it is only possible to evaluate these integrals analytically for a number of special cases, the integrals were evaluated using a numerical approximation. After further transformations the marginalized distributions are approximately Gaussian in shape as can be seen in Figure 6.3 and thus provide estimates of the standard deviations $\sigma_{B_{ani}}$ and $\sigma_{\Phi_{iso}}$.

The results of this process provide a 3D volume for the two parameters and their respective uncertainty: $B_{ani}$, $\sigma_{B_{ani}}$, $\Phi_{iso}$, and $\sigma_{\Phi_{iso}}$. The standard volume mapping is the horizontal dimensions corresponding to space, and the vertical dimension representing time (further stages in processing can map the wave related time dimension to space/depth). The voxels are also highly anisotropic in that the time resolution is very high compared to the spatial resolution. This not to be confused with the rock anisotropy property $B_{ani}$ that

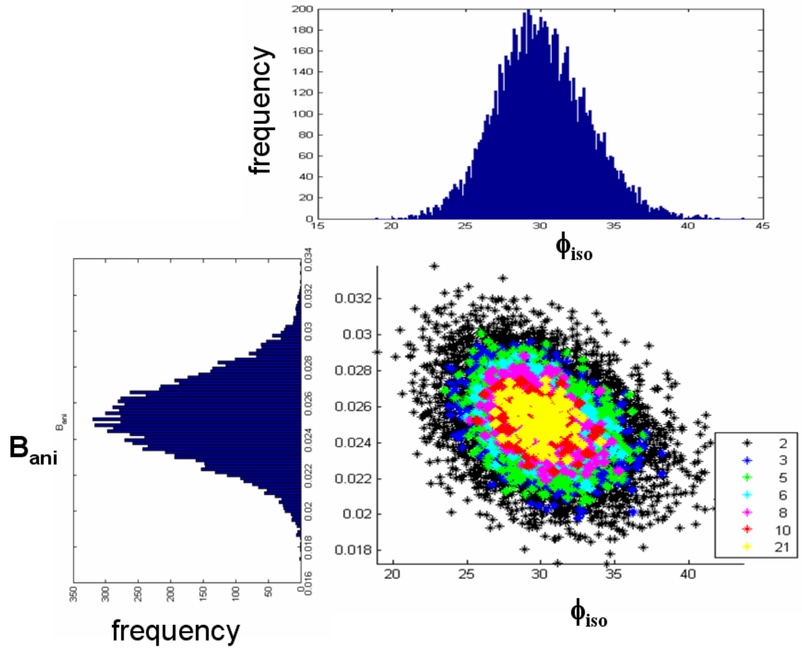



Figure 6.3: Distributions of $B_{ani}$ and $\Phi_{iso}$ used for estimating $\sigma_{B_{ani}}$ and $\sigma_{\Phi_{iso}}$.

is being estimated. Thus, while these volumes could be visualized using volume rendering schemes, they have high frequencies in the time dimension which complicates interpretation. Therefore, simple 2D slice planes are traditionally the preferred visualization. The approximation of the integrals (Equations 6.1 and 6.2) is also done on a time slice basis and the resulting vectors all lie on the time plane. The result is a stack of 2D bi-directional vector fields and their associated uncertainty.

Bi-directional vector fields add uncertainty about the sign of a vector. While directional vector fields can be thought of as vectors starting at locations on a grid, bi-directional vector fields can be imagined as line segments centered on grid cells. Therefore unique orientations are only in the range of 0 to $\pi$ radians (for the 2D case).

## 6.4 Visualizations of Bi-directional Vector Fields

We will describe our implementations of both a glyph-based and an animated flow-based visualization, that fill different niches, but that can also be utilized together.

### 6.4.1 Glyph-based Representation

Wittenbrink et al.'s [1996] static glyph provided a compact representation of the information allowing for specific orientations (realizations) to be imagined within the bounds of the glyph itself. Their glyph however could begin to look cluttered with highly dense fields and large angular uncertainty ($\geq$ +/- 45 degrees), and would have become even more congested with bi-directionality. Our glyphs provide for dense fields while still maintaining readability. The glyphs can be displayed on the currently selected horizontal slice plane, tracking along a surface, or throughout the entire 3D volume. Following Wittenbrink et al.'s [1996] lead we considered multiple possible glyphs. Displaying a dense field was the main use case and so "minimizing data-ink" [Tufte, 2001] was a useful design principle as also considered by Wittenbrink et al. [1996].

#### Implementation

Various viewing points and navigations are common for interpretation of the data in relation to other structures (geological surfaces and well core data). Therefore the glyph would potentially be viewed from all angles. While we considered and prototyped multiple glyphs the final version that shows both magnitude and orientation and their uncertainty is shown in Figures 6.1 and 6.4, and in diagram form in Figure 6.5. The rectangle is used to encode magnitude ($B_{ani}$) and orientation ($\Phi_{iso}$) while the less strong lines reveal the uncertainty in both. For visualization of uncertainty only in orientation an additional glyph was created to simplify the reading as shown in digram form in Figure 6.6 and in a visualization in Figure 6.7.

Both glyphs were designed to provide the clearest reading of all encodings at the top

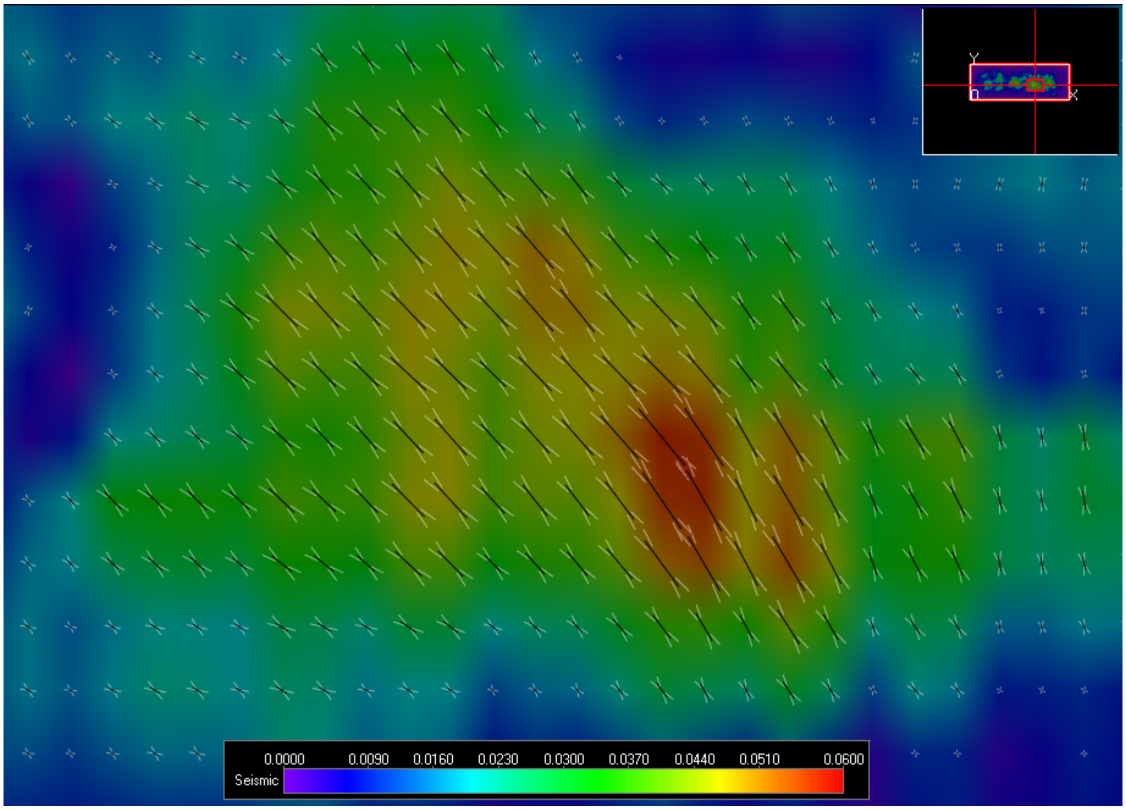



Figure 6.4: Top view of static glyph illustrating rock anisotropy data and uncertainty. Magnitude ($B_{ani}$) and orientation ($\Phi_{iso}$), are shown by black line segment length and orientation respectively, along with uncertainty in each component ($\sigma_{B_{ani}}$ and $\sigma_{\Phi_{iso}}$), indicated by white line segments, with equivalent reading. This can be compared against the colour overlay showing only magnitude ($B_{ani}$), from violet that represents no anisotropy, to red which is strong anisotropy. Larger context of visible data in entire volume shown via crosshairs rectangle in top right inset.

down viewing angle, but also allow reading of the orientation and magnitude at various oblique angles. As noted in the figures the side views (and all non-top views) show some projection of $\sigma_{B_{ani}}$ but this is also true for any non perpendicular view of $B_{ani}$. The magnitude of $B_{ani}$ is encoded in both the length of the rectangle and the height. This design prioritizes occlusion in a dense field based on $B_{ani}$ and thereby overall trends can be observed even when the viewpoint is close to the slice plane as seen in Figure 6.1. In the second glyph the magnitude of $\sigma_{\Phi_{iso}}$ will draw attention at large values while at small val-

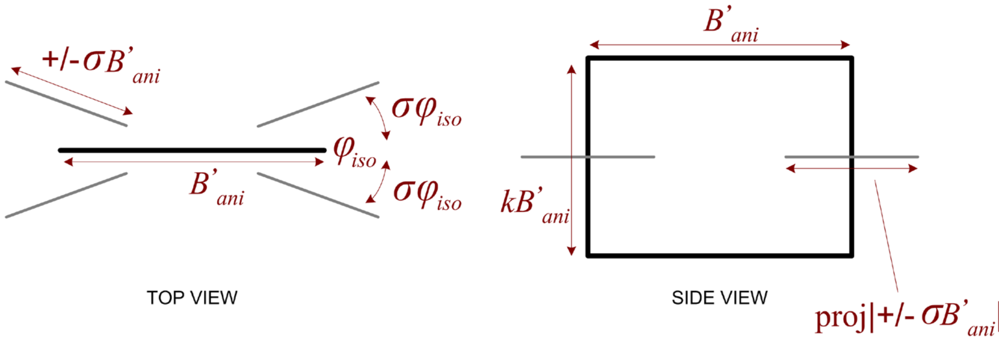



Figure 6.5: Static glyph illustrating bi-directional magnitude ($B_{ani}$) and orientation ($\Phi_{iso}$) along with uncertainty in each component ($\sigma_{B_{ani}}$ and $\sigma_{\Phi_{iso}}$). $B'_{ani}$ represents a user controlled length encoding for the dimensionless magnitude variable ($B_{ani}$) via the equation $B'_{ani} = k_1 B_{ani}^{k2}$. Glyph height can be scaled independently by another user specified constant $k$. proj$||$ denotes the projection onto the plane of the paper.

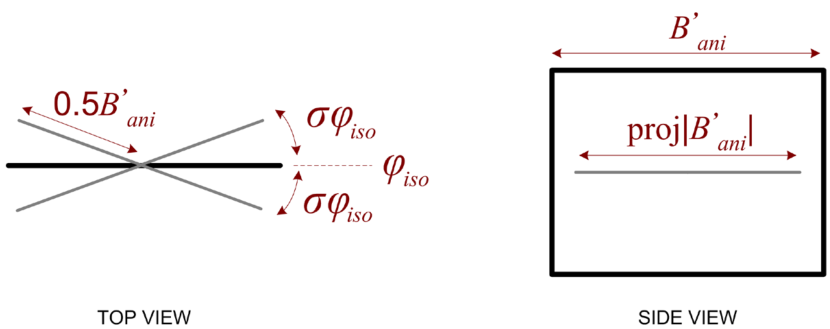



Figure 6.6: Static glyph illustrating bi-directional magnitude ($B_{ani}$) and orientation ($\Phi_{iso}$) along with only the uncertainty in orientation ($\sigma_{\Phi_{iso}}$). Variables have the same meaning as in Figure 6.5.

ues it tends to accentuate the edge detection of the main orientation line segment as shown in Figure 6.7.

### Interactive Controls

The visual appearance of a field of glyphs is quite different from individual glyphs. Therefore we provide interactive manipulation of the mapping of $B_{ani}$ to glyph length (both

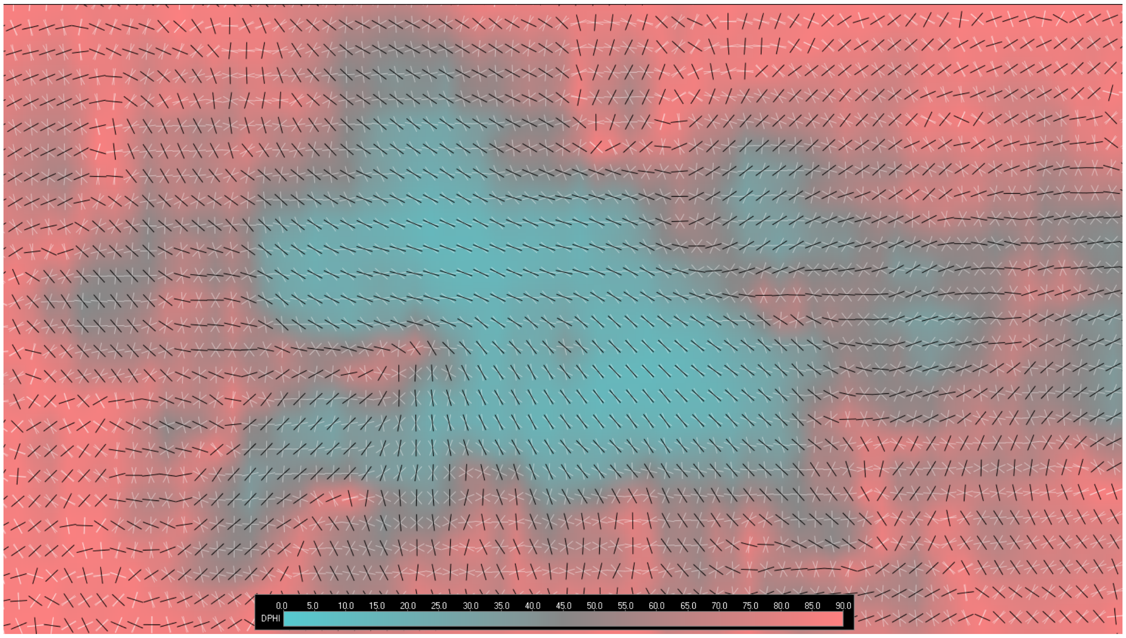



Figure 6.7: Use of the glyph in Figure 6.6 for depicting uncertainty in only orientation. This allows a quality control inspection of a large orientation field. The colour mapping on the slice redundantly shows orientation uncertainty ($\sigma_{\Phi_{iso}}$) from 0 to 90 degrees, an encoding which can be compared against the angle between the black and white line segments.

scalar and exponent factors, i.e. $k_1 B_{ani}^{k2}$) to create various overall field visualization effects. Another scalar, $k$, provides additional glyph height manipulation to provide the user control over the occlusion possible from various viewing orientations. If the size of a glyph is scaled beyond the size of a single voxel they may overlap and appear as a hatching style, thus creating a new regional representation.

The glyph in its natural form emphasizes uncertainty as it enlarges based on the uncertainty. Switching between the two glyphs in Figures 6.5 and 6.6 allows one to vary the focus on the uncertainty. In the first the uncertainty in the anisotropy, $\sigma_{B_{ani}}$, is an overriding factor of emphasis as it affects the viewing size, thus with very small $\sigma_{B_{ani}}$ the uncertainty components may be difficult to perceive even with large $\sigma_{\Phi_{iso}}$. While in the second glyph only orientation uncertainty, $\sigma_{\Phi_{iso}}$, has a role and so is useful for considering angular uncertainty in isolation, due to the possible dominant reading of $B_{ani}$ in the first glyph (reading of length vs. orientation).

### 6.4.2 Flow-based Representation and Animation

Recently Botchen et al. [2005] presented three advection approaches of multi-frequency noise, cross advection, and Gaussian error diffusion, for showing uncertainty in flow fields. Taking a differing approach than revealing uncertainty as a diffusion-like process, we allow user queries using a cursor probe to reverse and reorient flow vectors within the angular uncertainty ($\sigma_{\Phi_{iso}}$) and visually reveal the magnitude of reorientation with the amount of coloured dye injection. With this new approach the user can directly see variation in flow using an interactive query rather than having to imagine it. The user can thus create a simple form of unsteady flow [Bürger et al., 2007] based on their interaction via the cursor probe.

In a pre-processing step the ambiguity of bi-directionality should be resolved to one preferential direction. This should be done based on a spatially consistent scheme, or it could be assigned the most probable direction. These flow vectors can then be reversed and reoriented based on the position of the cursor probe. This process is explained in Figure 6.8 and for illustrative purposes we show the effects on simple left to right flow with increasing angular uncertainty (from 0 to $\pm$ 90 degrees) in Figure 6.9. The distance over which the cursor probe affects flow is set by the user. Reversals are not considered reorientation for the calculation of the amount of dye injection, due to the directional ambiguity.

This flow simulation is then used as an animated texture on the corresponding horizontal slice plane of the 3D volume. The user may interact with the slice using the probe or change visualization parameters, and can drag the slice plane up and down in the volume where the corresponding flow visualization will be started.

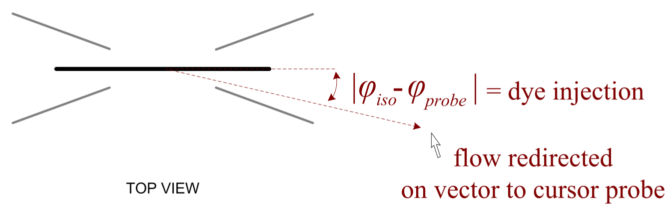



Figure 6.8: Left diagram: Flow redirection and calculation of uncertainty feedback for use in dye injection. Right diagram: No reorientation when query direction is beyond one standard deviation of orientation uncertainty ($\sigma_{\Phi_{iso}}$).

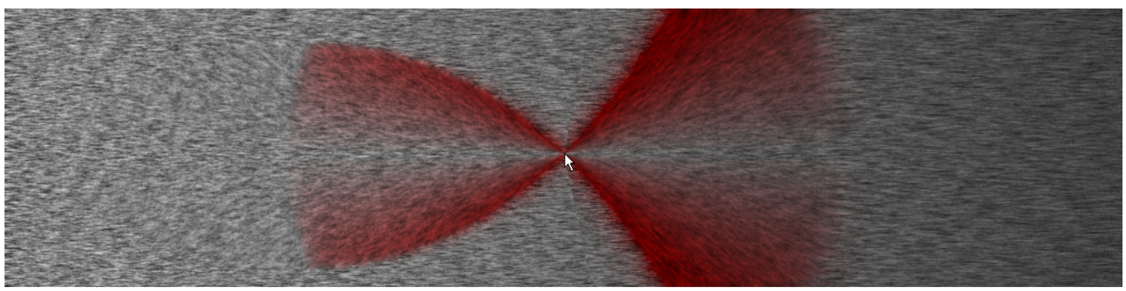


Figure 6.9: Illustration of flow reversal and reorientation query. Initial vectors are horizontal (flow left to right) with angular uncertainty increasing from 0 at the left edge of image to $\pm$ 90 degrees at the right. Red dye shows amount of angular reorientation in redirected flow toward cursor.

### Implementation

With a GPU-based implementation we use textures to store our vector field variables ($B_{ani}$, $\sigma_{B_{ani}}$, $\Phi_{iso}$, and $\sigma_{\Phi_{iso}}$) and a GLSL fragment program to perform the flow advection. Spot noise textures are also utilized to create the streak-lines as described by van Wijk [2002]. The flow simulation state for the current and previous time step are held in two OpenGL framebuffer objects (FBO). The previous time step FBO is used as a texture source and the current FBO as a rendering target to avoid the need for any OpenGL context switching. With this scheme the fragment program uses multitexturing to access all required grid data for the visualization.

Without coordinate variable representations fragment programs compute regularly spaced grid output. Thus our simulation steps backward in time ($t_{n-1}$) to advect toward the grid cell centers at the current time ($t_n$), rather than forward advecting results from the grid centers. This general scheme has been used by Botchen et al. [2005] and Jobard et al. [2002]. Our advection vector ($\vec{v_x}$) is computed as a function of location ($x = (i, j)$), all vector field variables, and the cursor probe location ($x_{probe}$), at the current time ($t_n$) as,

$$\vec{v_x}(x,t_n) = s_x(x, B_{ani}(x), \sigma_{B_{ani}}(x), t_n)\vec{d_x}(x, x_{probe}, \Phi_{iso}(x), \sigma_{\Phi_{iso}}(x))\Delta t, \tag{6.3}$$

with the function $s_x$ being a pseudo-speed function, and function $\vec{d_x}$ providing a normalized direction vector. $\vec{d_x}$ is computed as either $\Phi_{iso}$, $-\Phi_{iso}$, or the direction to the probe ($x_{probe} - x$) depending on its proximity to the probe being less than the user specified threshold and the angular difference between this direction and $\Phi_{iso}(x)$ being less than $\sigma_{\Phi_{iso}}(x)$. The flow speed $s_x$ is a function of $B_{ani}$ computed using user specified constants $k_1$ and $k_2$ as $k_1 B_{ani}^{k2}$. The user also has the option for viewing the flow speed animated over the domain ($B_{ani} \pm \sigma_{B_{ani}}$). For this $s_x$ animation the values oscillate with linear interpolation between $\pm$ one standard deviation over a fixed number of interpolation steps, but using delays computed from a Gaussian function for each step. Thus the flow duration for any given speed varies between a user specified duration of $T_{user}$ at $B_{ani}$ to $\approx \frac{1}{\sqrt{e}} T_{user}$ at the extreme values ($B_{ani} \pm \sigma_{B_{ani}}$). Calculation of the current grid location at our current time step is performed with the equation:

$$f(x,t_n) = f(x - \vec{v_x}, t_{n-1}) + g(x, x_{probe}, \Phi_{iso}(x), \sigma_{\Phi_{iso}}(x)), \tag{6.4}$$

where the grid $f$ forms the final texture image, and $g$ is any dye-like contribution (including spot noise). Uncertainty in magnitude is encoded with the $s_x$ animation that is reflected in $\vec{v_x}$. The calculation of function $g$ includes a linear blend of the spot-noise and a colour which reveals uncertainty in direction. The blend being proportional to the magnitude of the flow redirection. This is not an overlay but modification of the grid cell (RGB texture) contents which are advected. The repeated evaluation of $f$ can be considered an

Euler Forward Method integration of an approximate flow simulation[1]. Use of the non-grid aligned previous time step results (i.e. $f(x - \vec{v}_x, t_{n-1})$) is performed using bilinear texture filtering. The noise injection creates the variations of streak-lines based on the user controlled blending rate of previous time step results.

Our visualization does not try to represent actual fluid flow (e.g. oil, water) through rock, but only is intended to provide relative comparisons of anisotropy between areas. For our use of "flow" only as a graphic variable, the limiting of velocity (and thus advection distance) to around one grid cell width is acceptable, thus maintaining the Courant-Friedrichs-Lewy Condition (numerical stability) and avoiding aliasing of the spot-noise patterns. Therefore we scale the visualization flow velocities down to a unit range, which is equivalent to reducing the flow simulation time step, to avoid these same issues. This enables us to skip any post advection filtering step as used by Jobard et al. Jobard et al. [2002]. If the CPU or graphics hardware cannot attain the user requested animation frame-rate the net animation velocities are effectively slowed down as well. The actual fluid flow velocities, which would be slow and whose estimation would be extremely approximate, along with the scale of data (on the range of kilometres) preclude the value of a temporally realistic simulation. More importantly velocity is only related to $B_{ani}$, and our visualization is not an accurate model of flow. However, flow is a good graphical encoding as actual fluid flow is a realization of interest. The majority of the computation time is spent in the fragment program calculating Equations 6.3 and 6.4 (direction determination, Euler integration, dye additions).

### Interactive Controls

With the flow visualization we allow the user to interactively reverse local flow by moving the cursor probe (a sink or source) over the field. Within a user defined distance of the sink

[1]This approximation will introduce some error but the simulation is not an accurate flow model because there are too many unknown parameters, and thus the uncertainty introduced by this aspect of the representation should not have any significant impact.

cursor probe, any vector that points away from the cursor is automatically flipped and vice versa for the source. In the application domain this may correspond with reality in that an oil field well may either pump in fluids or be used for extraction.

Similarly the user can also explore explicit realizations with the cursor probe as flow vectors are reoriented directly toward the cursor if this new vector lies within their orientation uncertainty as defined by $\sigma_{\Phi_{iso}}$. Our use of colour specifically provides visual feedback indicating the difference between the most likely orientation of each vector and the user requested orientation. The results of this interaction are shown in Figure 6.10.

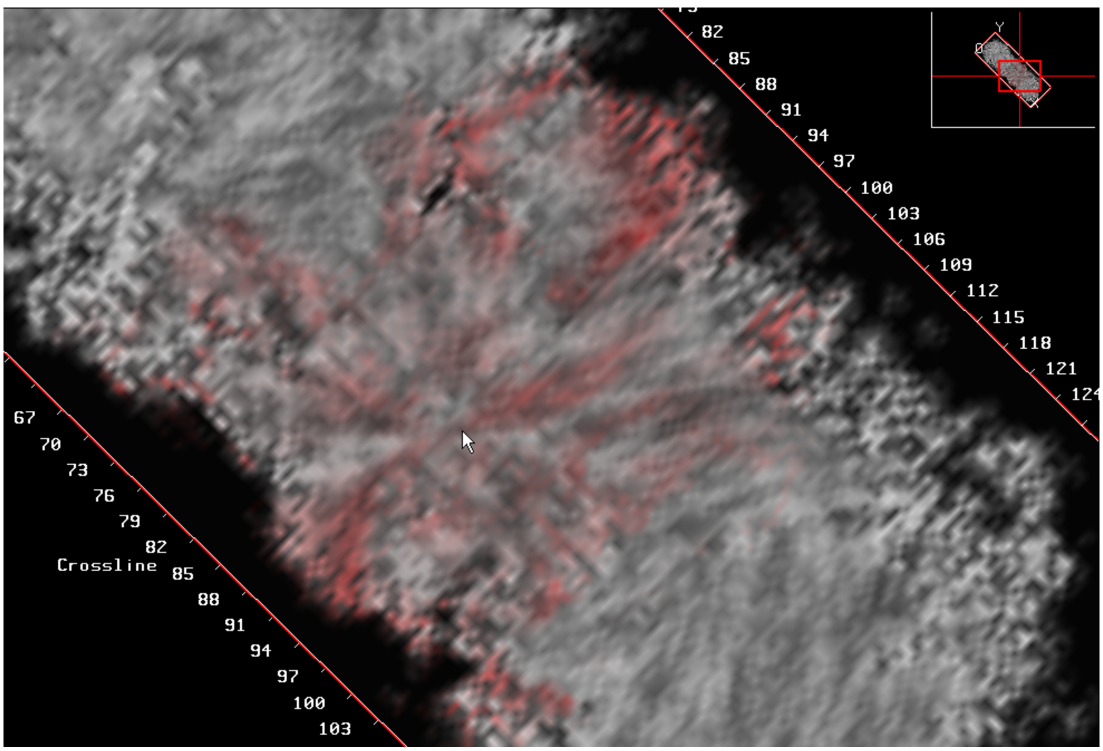



Figure 6.10: Flow visualization showing user directional and orientation query. Red dye injection is based on the difference between user requested flow orientation and the most likely direction. Context shown in the top right inset.

Opposite to the glyph, with the base flow visualization uncertainty is de-emphasized when there is no user interaction. Selective emphasis of uncertainty is provided based on the user's movement of the cursor probe revealing uncertainty in flow direction. The uncertainty is injected like a dye and then flows along streamlines fading with the spot-noise.

Within the flow, user thresholding of regions is also performed by blending highly uncertain areas to black. This is currently based on a lower threshold of anisotropy magnitude ($B_{ani}$), as it is correlated with orientation ($\Phi_{iso}$) in that orientation is undefined at low levels of anisotropy. Thus the user can eliminate areas from consideration; this helps the user avoid watching for patterns in regions of arbitrarily assigned directions.

## 6.5 Visualization Use

These visualizations may be used to interactively explore the data and its uncertainty on one slice plane at a time. The previously described user controllable variations allow their use to be tailored to the specific phase of interpretation and exploration.

### 6.5.1 Glyph and Flow Integration

The user can choose to combine both the flow visualization and the static glyphs. This allows the unique benefits of each to be combined. The user can be guided by the reference provided by one visualization style while interactively adjusting the display parameters of the other. The movement provided by the flow may also enhance visualization with large displays by utilizing the stronger perception of motion in peripheral vision [Ware, 2004]. Figure 6.11 shows the glyph on its own and then combined with the flow visualization, both from a more distant viewing point where the uncertainty encodings are deemphasized.

### 6.5.2 Simplified Visualizations

All parameters including uncertainty can be viewed as standard colourmapped slice planes through the volume, as shown in Figure 6.7. This allows the uncertainty to be treated as

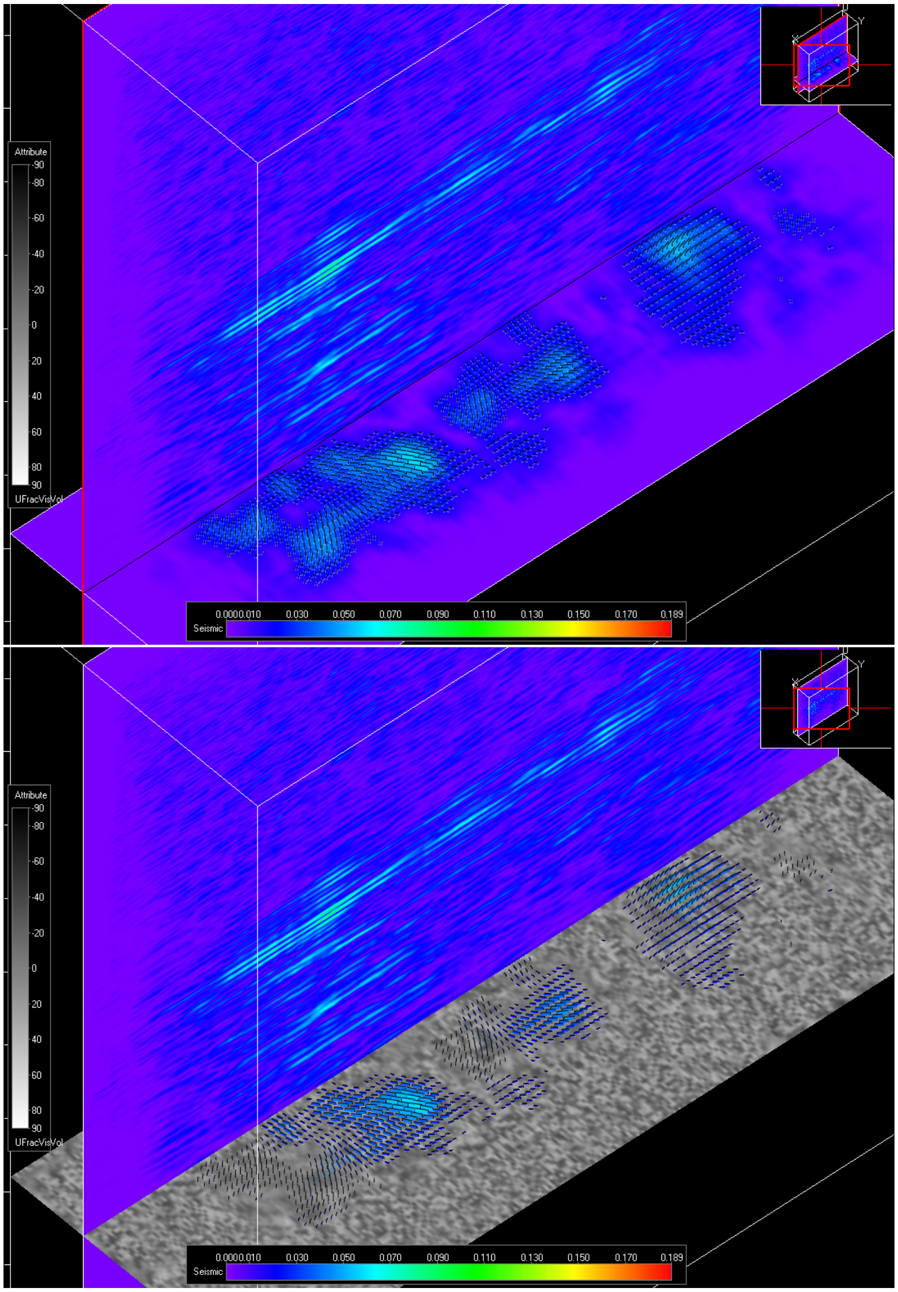



Figure 6.11: Distant views of glyph alone (top image) and combined with flow visualization (bottom image) along one horizontal slice of 3D volume of data. Vertical slice shows the relatively higher frequency data in that dimension.

data rather than meta-data and simplifies tasks that may only concern uncertainty issues. The second glyph representation also provides a form of this in that it allows a simpler reading of the orientation uncertainty on its own as $B'_{ani}$ can be chosen to be replaced with a unit sized vector (also shown in Figure 6.7). This is important as when considering uncertainty information it is helpful to allow the user to decompose and view each component individually to assist in comprehension or interpretation.

### 6.5.3 Tasks

Returning to consider the tasks used by Laidlaw et al.'s [2005] evaluation (locating critical points, identifying their type, and predicting particle advection), our animated flow visualization may aid in locating critical points as the motion may enlarge the user's useful field of view [Ware, 2004]. However an evaluation would be required to determine how easily the motion of specific types of critical points can be perceived, preattentively or otherwise, separately from the other flow motions within this larger field of view. Additionally for both the tasks of identifying the type of critical point and particle advection prediction, the explicit streamline tracing and unambiguous flow direction should be of assistance.

The bi-directionality is important to explore in the seismic industry, and as stated previously new sources or sinks (which would be wells) can be interactively placed to visualize the resulting flow as vectors are reversed and redirected. Uncertainty in the amount of redirected flow was made apparent by the amount of red dye being injected and transported. Bi-directionality of the vectors was also inherent when reading the glyphs, and overall trends may be seen with both the flow and glyph visualizations individually or combined.

## 6.6 Heuristic Evaluation

In order to further the understanding of these visualizations, we return again to the heuristics presented in Table 3.1. The following is a summary of the application of these heuristics to the visualizations in this chapter:

- **Ensure visual variable has sufficient length** – The glyph uses size to encode the vector magnitude, and orientation is used to encode the vector direction. Uncertainties are represented using the same visual variables and they provide sufficient length. The adequacy of these variable's length is partly derived from the enhanced perception of differences between adjacent glyphs for which smaller variations can be perceived to read the local trends. The flow visualization encodes magnitude information based on advection speed and encodes direction using the orientation of the flow advection. Uncertainty in magnitude is encoded with variation in advection speed, and the uncertainty in orientation encoded by red dye. Both these variables for uncertainty may not provide sufficient length for readings at a micro level due to the blending of noise injection patterns: small motions may be lost in the noise, and the blending further reduces the length available to the red value variation. However, even with the reduced length the flow visualization allows reading of overall trends at the macro level.
- **Preserve data to graphic dimensionality** – The glyphs preserves the dimensionality in the plane, but out of the plane the rectangle encodes the vector as a region. The violation of this heuristic was done to simplify reading the orientation at oblique viewing angles. While informal feedback suggests that this is effective, this use of dimensionality might benefit from further empirical study. The flow visualization preserves spatial dimensionality, but the use of animation adds a temporal dimension. If the data is already time varying this violation of the heuristic may not be appropriate.
- **Put the most data in the least space** – Both the glyph and flow visualizations allow for dense encoding of four parameters on a plane.
- **Provide multiple levels of detail** – Interactive controls allow the glyph to be shown

based on sub-sampling the data. The flow representation may be utilized more as an overview, and the glyph for finer detail. Additionally both a magnification lens and an overview inset can be displayed.

- **Remove the extraneous (ink) –** This heuristic was part of the basis for the design strategy of minimizing "ink" in the design of the glyph. The heuristic also suggests that reducing the flow representation to a subset of the streamlines should be considered.
- **Consider Gestalt Laws –** The Gestalt Laws predict the problems with the glyphs if they are made too large. If the uncertainty "whiskers" cross other rectangle lines or other "whiskers" the connectedness caused by overlap may confound reading them. It may no longer be clear if they are a part of the glyph with which they were composed.
- **Integrate text wherever relevant –** Text feedback is provided in the application window status bar, but could also be added at the cursor to provide simplified integrated reading.
- **Don't expect a reading order from colour –** Colour mapping can be chosen by the user for volume slicing and the glyph rectangle and therefore colourmaps with value variation can be used if an order is required.
- **Colour perception varies with size of coloured item –** The glyph rectangle can be coloured to encode anisotropy magnitude and so this heuristic warns us that at some sizes the same colour may be read differently. The size redundantly encodes this magnitude and so may counter this effect.
- **Local contrast affects colour & gray perception –** Alternative non-colour and value encodings are provided to avoid these reading issues. The flow also uses a noise based pattern which should assist with the reading of uncertainty colour value by averaging out the background contrast.

- **Consider people with colour blindness** – Colourmaps may be chosen by the user to avoid these issues.
- **Preattentive benefits increase with field of view** – The preattentive encodings that are utilized allow large fields of view to be reviewed based on visual variables such as motion, size, colour, and orientation.
- **Quantitative assessment requires position or size variation** – The glyph encoding allows this type of quantitative assessment based on size, as position is already used for spatial encoding.

The heuristics revealed aspects of both the visualizations' strengths and weaknesses and this again demonstrates the applicability of the heuristics themselves. One heuristic may have to be balanced against another in making design choices, such as "remove the extraneous" versus "gestalt", but this constraint can be useful in limiting designs that might otherwise go too far in one direction.

## 6.7 Conclusions

We have created two new and differently styled visualizations for uncertainty in bi-directional vector fields. These two new visualizations extended previous work in the area of uncertainty visualization and vector fields for both bi-directionality of the vectors and richness of interactivity. The interactive aspects enabled the glyph visualization to provide a user-adjustable, precise, micro reading in an abstract form. With the flow visualization the interactivity allowed user driven exploration of possible flows, which provided a macro reading of the data and its uncertainty.

While these visualizations focused on the data uncertainty, it may also be worth considering visualizing the uncertainty in interpretation. For example, if the critical points in possible flow fields were automatically detected, such as with Ford's [1997] approach, they could be labelled showing the classification confidence. We expect it may be of value to integrate this or the interpreter's confidence directly into the visualization as a decision aid. Looking to investigate cognitive issues further, we turn to the medical domain provided in the next two chapters.

## 6.8 Acknowledgements

We would like to thank CGGVeritas for the data and also for funding in conjunction with NSERC.

# Chapter 7

# Case Study in Medical Diagnostic Reasoning

# Part I: Problem Analysis and Design Issues

Doubt is not a pleasant condition, but certainty is absurd.
– François-Marie Arouet a.k.a. Voltaire (1694 – 1778)

This chapter introduces the third and final domain investigation, looking at the issues involved in medical diagnosis. This domain is different in that the uncertainty directly relates to the reasoning process and potential decision support. Thus, both reasoning and data uncertainty will be explored. The methodology also follows a more in-depth strategy to formally develop a deeper understanding of the problem and of existing support. The observational grounding, collaboration with a pair of domain experts, and association with an interdisciplinary research group for medicine, were all important for investigating this domain.

The diagnosis of medical conditions can be extremely challenging and motivates us to provide improved decision support tools. Diagnostic reasoning in evidence-based medicine (EBM) relies on updating estimates of probability, but many other uncertainties exist in the task, such as the physician's confidence. To ground the design of new visualization support, an observational field study of existing computer support and contextual interviews were conducted. Based on the study we provide a task model that decomposes and structures the problem. Our discussion exposes the role of uncertainty in the sub-tasks and provides design considerations and recommendations for future computer support for EBM[†].

[†]Portions of this chapter have been previously, or will be, submitted for publication. Therefore "we" refers to Torre Zuk, Sheelagh Carpendale, William Ghali, and Barry Baylis.

## 7.1 Introduction

To shed more light on the best practices for uncertainty visualization we explored the domain of medical diagnosis. Before creating new decision support for the complex problem of medical diagnosis, which is structured around probabilities and managing uncertainty, it is important to create grounded design criterion. To gain this knowledge we performed a series of observational and contextual interviews to analyze current practices in diagnosing pulmonary embolism at Foothills Hospital, Calgary, Alberta. The study and analysis provided a task model and insights on the uncertainties, which are then used to provide design implications for software support.

To better understand the diagnostic decision process we ran a focused observational study involving the particular condition of pulmonary embolism (PE). PE is a potentially lethal disorder accounting for approximately 100,000 deaths annually in the US. It can present with a variety of signs and symptoms and in those suspected with PE the prevalence rate is only 30%. It is important to not miss the diagnosis as the 30-day mortality rate can be as high as 17%. For physicians, this diagnostic dilemma is amplified by the wish to avoid diagnostic tests that are invasive, associated with risk, and expensive. Therefore its detection is typically accomplished through the use of non-invasive diagnostic tests that have imperfect sensitivity $(\frac{true\ positives}{true\ positives+false\ negatives})$ and specificity $(\frac{true\ negatives}{true\ negatives+false\ positives})$. Inherent in this process is the consideration of uncertainty in final diagnostic decisions since false negatives may lead to mortality, and false positives to unnecessary treatment with potentially serious side-effects. Physicians often face uncertainty about the presence of PE as they cope with the difficult challenge of combining clinical estimates of the probability of disease with medical test results.

We observed and interviewed physicians, who use an evidence-based medicine (EBM) diagnostic approach, during their use of existing software tools that have been explicitly designed to facilitate an EBM process of diagnosing PE. EBM is a practice in which a doctor will try to establish an *a priori* probability of a condition and then use probabilistic evidence to determine the *a posteriori* probability of that condition [Jenicek, 2002]. While these software tools were designed to assist clinicians in the difficult process of accurately diagnosing pulmonary embolism; anecdotally they have several limitations that affect their use in clinical settings. The analysis of our study data exposes some of these limitations and provides implications for the design of subsequent software support.

## 7.2 Problem Domain

EBM involves continual weighing of probabilities and uncertainties. For example, a test result can be considered as probabilistic evidence, as rarely do tests provide absolute certainty that a patient has a specific condition. EBM also accepts the fact that many decisions must be made based on best guesses as absolute certainty is not a practical or even reasonable goal. To practice EBM the fundamental statistical components are conditional probabilities. While numerous decision support software and tools have been developed, a recent systematic review of their clinical performance found that the majority have not produced significant benefits in terms of patient outcomes [Garg et al., 2005]. This does imply that decision support systems have not shown value (e.g. training, efficiency gains). For a more general discussion and examples of clinical decision support which are beyond our scope, see Berner [1999].

Study derived statistical evidence does not easily fit within the experiential paradigm of a naturalistic decision maker. If, as in an EBM process, a physician is to apply recent policy or strategies as recommended by the latest medical studies, they are faced with integrating new statistical information with their own experience-based knowledge. This may create a dilemma for the proper utilization of evidence-based protocols as it forces what may be an internalized process to mesh with the world of explicit external probabilities.

### 7.2.1 Bayesian Approach

Conditional probabilities allow for probability revisions based on observations. In its simplest form a conditional probability can be stated as the probability of A given that B is known, which is the probability of the intersection of A and B divided by the probability of B. Expressed in statistical notation, this is much more concise as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{7.1}$$

If this is visualized using a Venn diagram (see Figure 7.1) it becomes more intuitive, as you could count the dots in the intersection and divide by the total number of dots in $B$ to get your answer for $P(A|B)$. The obvious symmetry between A and B in a diagram

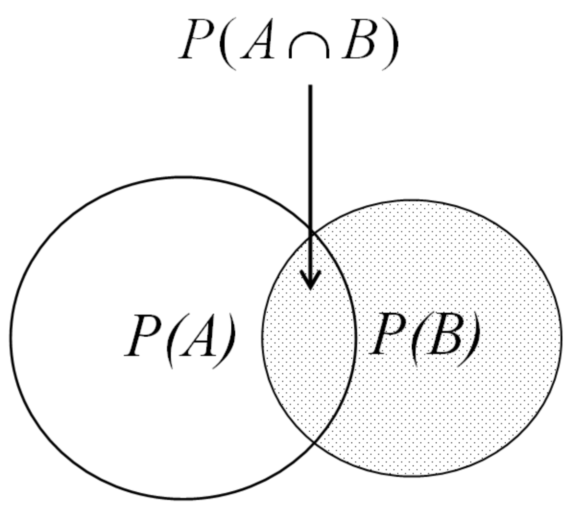



Figure 7.1: Venn diagram to aid understanding of conditional probability.

such as this may even have been the insight which led Bayes to generalize the conditional probabilities equation into its bidirectional form:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}. \tag{7.2}$$

This is Bayes Theorem and it allows conditional probabilities known in one direction ( $P(B|A)$ ) to be used to compute conditional probabilities in the other direction ( $P(A|B)$ ). It can also be used for more than two regions in the form:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i) + P(B|\neg A)P(\neg A)}, where \bigcup_{i=1}^{n} A_i = S. \tag{7.3}$$

An example medical question for the purpose of illustrating the use of Bayes Theorem is:

> The probability of breast cancer is 1% for a woman at age forty who participates in routine screening. If a woman has breast cancer, the probability is 80% that she will get a positive mammography. If a woman does not have breast cancer, the probability is 9.6% that she will also get a positive mammography. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer? [Gigerenzer and Hoffrage, 1995, p.685].

Before reading further you should try and determine if you can answer the question by applying Bayes Theorem above or using your intuition. If you did not bother trying (very likely) and you do not know where to start I have helped you by mapping it to the formula, where $C$ = cancer, $\neg C$ = no cancer, and $M_{pos}$ = positive mammogram:

$$P(C|M_{pos}) = \frac{P(M_{pos}|C)P(C)}{P(M_{pos}|C)P(C) + P(M_{pos}|\neg C)P(\neg C)}. \tag{7.4}$$

and so this provides the solution:

$$P(C|M_{pos}) = \frac{(0.80)\cdot(0.01)}{(0.80)\cdot(0.01) + (0.096)\cdot(0.99)} = 0.078.$$

The initial step of mapping the problem to an equation or algorithm is difficult, as medical students have been shown to have difficulties with this type of question [Gigerenzer and Hoffrage, 1995].

### 7.2.2 Cognitive Heuristics

Tversky and Kahneman [2003b] studied the cognitive heuristic called the conjunction fallacy (giving a conjunction greater probability than either of the two components) in a pulmonary embolism decision task. Participants were asked to “rank order the following in terms of the probability that they will be among the conditions experienced by the patient” [Tversky and Kahneman, 2003b]. Example symptoms were:

- dyspnea and hemiparesis,
- calf pain,
- pleuritic chest pain,
- syncope and tachycardia,
- hemiparesis, and
- hemoptysis,

with the symptoms of interest being the conjunction of dyspnea (typical) and hemiparesis (atypical) versus hemiparesis alone. Two groups (37 and 66) of internists consistently ranked the conjunction of atypical and typical as more likely than atypical alone, even though standard interpretation indicates the former is a subset of the latter. Surprise and dismay were among the responses of another group of 24 physicians when being confronted by their apparent violation of basic rules of probability ($P(A\&B) <= P(A)$ and $P(A\&B) <= P(B)$) [Tversky and Kahneman, 2003b].

While other domains have also shown the existence of conjunction errors with statistically savvy participants [Tversky and Kahneman, 2003b], perhaps insight into why this happens can be found by investigating the diagnostic task as well as uncertainties in the data. If one must consider uncertainty in all observations as is the case with diagnostic tasks, then conjunctions may not be optimally interpreted in an abstract statistical framework. This being said, one should not expect physicians to be immune from this or other heuristics’ potential for error.

### 7.2.3 Visualization Support

As stated previously in Chapter 4, natural frequencies and cognitive transparency of the nested-set information structure can enhance a person's ability to compute a Bayesian solution for a conditional probability problem [Gigerenzer and Hoffrage, 1999, Sloman et al., 2003]. Thus, if a visualization provides access to levels of detail that expose the application of Bayes Theorem, we should try to make this nested-set information structure apparent. A post-test probability problem is illustrated in Figure 7.2 using some natural frequencies. Interpreting the representation in the diagram one can easily determine that $P(Cancer|Test_{pos})$ is the count of those who have cancer and test positive divided by the total count of those who test positive:

$$P(Cancer|Test_{pos}) = \frac{8}{8+95} = 0.078.$$

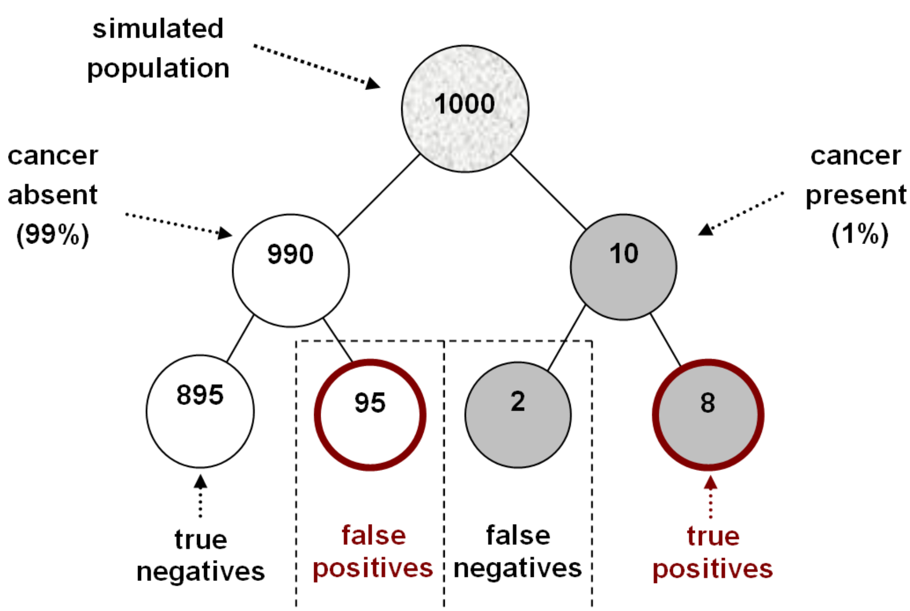



Figure 7.2: Natural frequency diagram to aid calculation of conditional probability. Shows nested sets of positive and negative test outcomes from healthy and diseased populations.

To further the point that visualizations may help people understand a process, Sloman et al. [2003] have shown in a study of probabilistic reasoning that the use of Euler circles greatly reduced the number of people committing the conjunction fallacy.

Klein's [1998] recognition primed decision model (RPD) describes how experts can make quick and effective decisions without consciously comparing multiple options. Klein applied his model to experiences with nurses and paramedics, however, while likely to be applicable it has not yet been specifically applied to physicians' decision making. We can utilize the model to suggest where visualization support might be of assistance, such as under uncertainty, in atypical scenarios, and when cues are difficult to prioritize.

Reducing uncertainty is a fundamental component in the diagnostic process of ruling-out and ruling-in conditions. When a test provides evidence it may be statistically interpreted in the form of conditional probabilities to compute post-test probability of a condition using Bayes Theorem (Eq. 7.2). However, calculation of post-test probability is usually done by converting pretest probability to odds, multiplying by a likelihood ratio (which is the predictive power of a test outcome), and then converting from odds back to probability.

Turning to uncertainty in reasoning there has been little direct visualization of the reasoning process itself, as the focus has been on the data and its uncertainty. Some work has been done on visualizing argumentation [Kirschner et al., 2003], but integrating support for reasoning introspection into information visualizations requires further exploration. Reasoning heuristics and biases have been found to potentially degrade performance when reasoning under uncertainty [Kahneman et al., 1982] and this may be pertinent to evidence-based medicine [Elstein and Schwartz, 2002]. Thus one may expect there are potential benefits from any cognitive support for the reasoning process [Zuk and Carpendale, 2007].

## 7.3 Study Methodology

Our challenge was to improve computer support for a task that is largely embedded in a physician's thought process. Thus it was deemed important to first assess and understand the role of existing software support. An observational field study was chosen as it offered the potential to capture aspects that may not have been explicit out of context and provide qualitative insights into the bigger picture of diagnostic processes.

We began by observing how physicians use the existing computer support for the task of diagnosing pulmonary embolism (PE) at Foothills Hospital. To further flesh out the issues the observational aspect was immediately followed by a questionnaire and then a contextual interview. Our goal for this study was to improve our understanding of the diagnostic task and its associated uncertainties.

### 7.3.1 Participants

We were interested in the full spectrum of medical experience and so participants were solicited from all levels. The study was conducted in a teaching hospital where both residents and staff physicians work. Seven participants, five women and two men, were involved. The participants' formal experience levels were four first year residents, one third year resident, and two staff physicians. Experience with evidence-based medicine varied from 8 months to over 10 years. All were comfortable with computers, each having more than 10 years experience using them.

### 7.3.2 Methods

We observed doctors in situ, performing the task of diagnosing pulmonary embolism with simulated patient data. This was followed by a written questionnaire, and then a discussion style question and answer session. The observations and contextual interview data were conducted by a single experienced software developer. A pilot study was performed with one experienced physician as a check on the process and questions, the results of which, led to slight revisions.

For the observational component of physicians proceeding toward a diagnosis, we utilized simulated patient case histories presented on paper, with some specific and some vague details, that were approved for our purposes by an experienced physician. These simulated patients were also entered into the existing computer system, the Technicon Data Systems (TDS) 7000. Initially a leading statement was used, "given a pulmonary embolism mindset can you consider this patient". Participants were asked to work through the diagnostic process in as realistic a manner as possible, and to use the system to order any tests they thought were necessary to move toward a diagnosis. Test patients were added into the live TDS system, and so could be accessed in the same way as actual patient data. Participants were asked to use a "think aloud" protocol as they worked. Due to the lack of reports on thought processes and speed of data entry, occasionally participants were asked to slow down, repeat what they had done, and sometimes briefly explain their decision process. No choices they made were ever called into question. After the observational component of simulated tasks performed in situ, participants completed a written questionnaire, which was then followed by a contextual interview.

**Clinical Cases for Diagnostic Testing**

Two simulated patients were created to be diagnosed by the participants. The first case "Patient A" was entered in the system as "Pathfinder, Torre A" and had the following description:

- 52 year old Caucasian woman,
- height: 170 cm,
- weight: 61 kg,
- heart rate: 98 beats per minute (bpm),
- temp: 37.5 C,

- feeling weak and short of breath,
- tinges of blood in sputum, and
- no recent medical problems or chronic condition (no previous DVT/PE).

All questions regarding other vitals and tests were answered that the results were non-diagnostic and no causes were suggested. It was expected that this patient would fall into the low probability branch of the PE diagnostic tree.

The second case "Pathfinder, Torre B" (Patient B), had slightly different characteristics with the addition of more diagnostic symptoms:

- 46 year old Caucasian male,
- height: 194 cm,
- weight: 93 kg,
- heart rate: 105 bpm,
- Temp: 38 C,
- short of breath, right leg is swollen with pain on palpation,
- pleuritic chest pain, and
- no recent medical problems or chronic condition (no previous DVT/PE).

Again requests for further information were provided as non-diagnostic. This patient was intended to be grouped into the moderate to high probability of PE category of the PE diagnostic tree.

### 7.3.3 Environment

The study occurred at four different locations in the teaching hospital. Performance of the task required access to the on-line TDS information system. Use of the live system dictated that terminals could not be reserved and so availability dictated the location used for a scheduled session. Sessions occurred based on participant schedules and were often

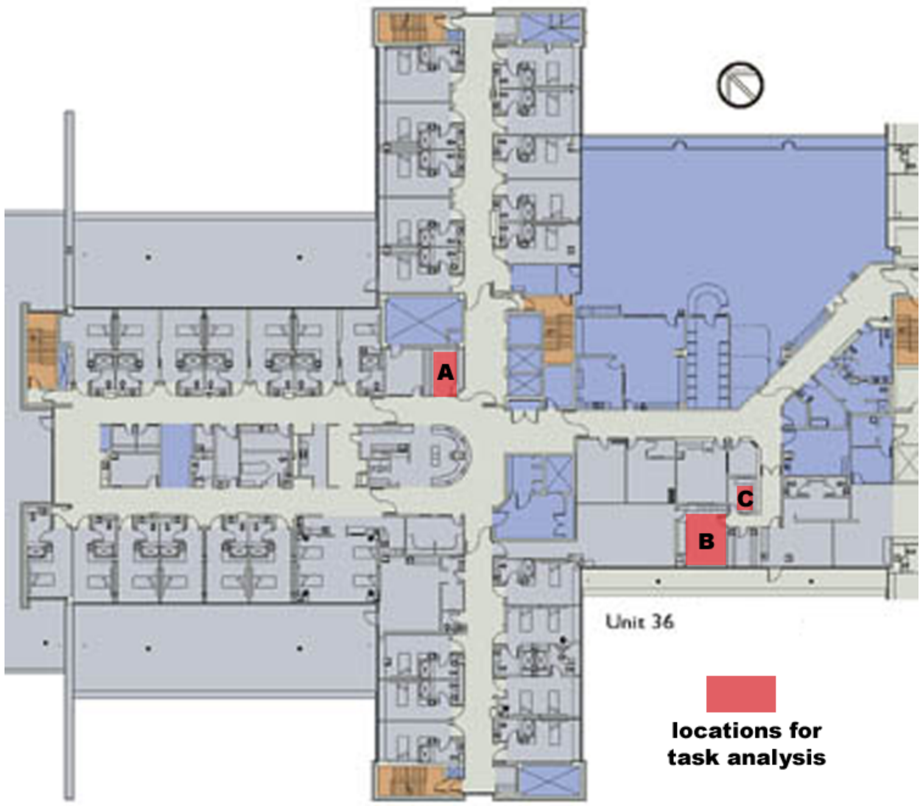



Figure 7.3: Locations used for observations and contextual interviews.

at the end or before their regular work times. This resulted in sessions at various times throughout the day and early evening.

Three locations on one ward (Unit 36) were used and are shown in Figure 7.3. One location was a common array of terminals for general use in close proximity to the main administration desk (A) at the hub of the unit. The second location (B) was a residents' debriefing room with two terminals, a meeting table, and lockers. The third location (C) was a more private "physician's room" with only a single terminal located off of a quiet hallway leading to the second location. The final location was outside Unit 36 on the main floor in the "Doctor's lounge" where a long narrow room has multiple terminals available for general use (all but one terminal lined up along a single wall). Locations A, B, and C were increasing in privacy, while the final location was the least private. All these locations would be possible locations for normal performance of the task.

Each terminal used was a personal computer running Windows XP software. The main application used was the TDS system launched from the desktop icon. Internet explorer was occasionally used for researching information to aid in the diagnostic decision making. Some participants also had Palm Pilots which can be of use for some sub-tasks related to practicing evidence-based medicine (EBM). An example of this type of artifact is shown in Figure 7.4.

Figure 7.4: Example artifact of Palm Pilot with pen for size reference (Note: image blurred for anonymity).

## 7.4 Study Results

### 7.4.1 Observations

For both hypothetical patients, all participants with only one exception, to be noted later, considered PE the top candidate. This determination agreed with our hypothesis given the patient descriptions and the PE mindset. Ordering a test with the TDS system relating to diagnosis of PE inevitably leads to a PE Wells score [Wells et al., 2000, 2001] calculation. The questions for computing the Wells score that are asked by the TDS system are provided in Table 7.1.

Table 7.1: Wells scoring for PE as described by TDS (points given for positive answers).

| **Question** | **Points** |
|---|---|
| Signs and symptoms of DVT: leg swelling (objectively measured) and pain with palpation in deep vein region | 3.0 |
| Pulse >100 beats per min | 1.5 |
| Immobilization, bed rest, or surgery in previous four weeks | 1.5 |
| Previous DVT or PE (objectively diagnosed) | 1.5 |
| Hemoptysis | 1.0 |
| Malignancy and/or<br>A) receiving treatment for cancer,<br>B) received treatment for cancer within last six months,<br>C) receiving palliative care for cancer | 1.0 |
| PE as likely or more likely than an alternate diagnosis (no specific criteria - use hx, physical exam, chest X-ray, EKG & lab results to decide) | 3.0 |
| **Pretest probabilities** | **Total points** |
| Low | <2 |
| Moderate | 2-6 |
| High | >6 |

This scoring system is a form of actuarial judgment for computing the a priori probability of PE, versus a more holistic clinical judgment estimate. The Wells score (score to probability of PE: $< 2$ low, $2-6$ moderate, $> 6$ high) is then used to suggest the next step along the hospital recommended decision tree incorporated into the TDS system. The TDS will recommend a D-dimer[1] test for low to moderate pretest probabilities and a ventilation/perfusion[2] (V/Q) scan for high probabilities. A visualization of the diagnostic decision tree that exists within TDS is shown in Figure 7.6.

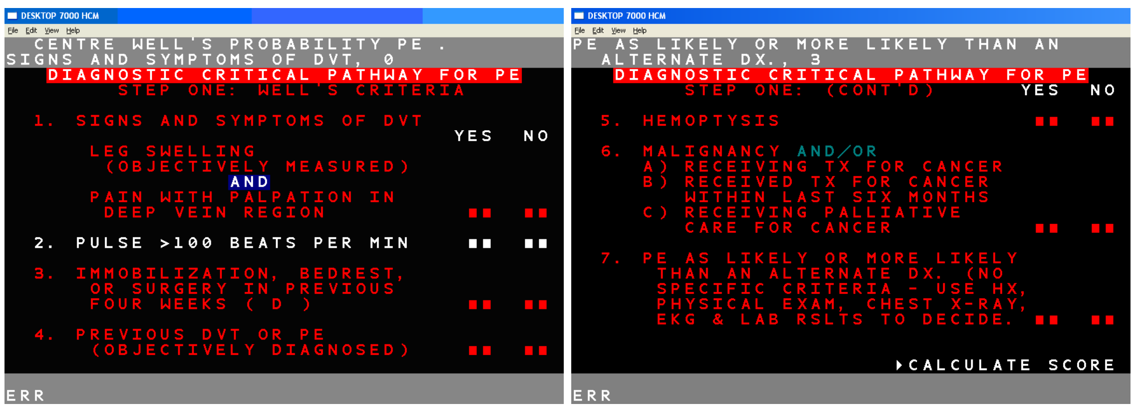



Figure 7.5: Screenshots of the TDS system screens for PE Wells Score. Highlighted question is current question.

### Diagnostic Process for Patient A

All participants determined the candidate condition to pursue was PE. Many considered using TDS to look for other test results, but participants were told no other relevant test information was in the system. Participants for the most part ran through mental checklists of risk factors for all candidate conditions (e.g. drug abuser, recent surgery, asthma, acute bronchitis, ...). In a couple of cases participants stated they would have performed more extensive research, but given the time constraints went with their best guess of PE. Some participants highlighted in pen the key symptoms on the sheet of paper with the pa-

[1]D-dimer is blood test which can detect clot or thrombus, it is very sensitive, but not very specific.

[2]A ventilation/perfusion scan evaluates the circulation of air and blood within a patient's lungs, abbreviated V/Q, where Q represents the perfusion variable.

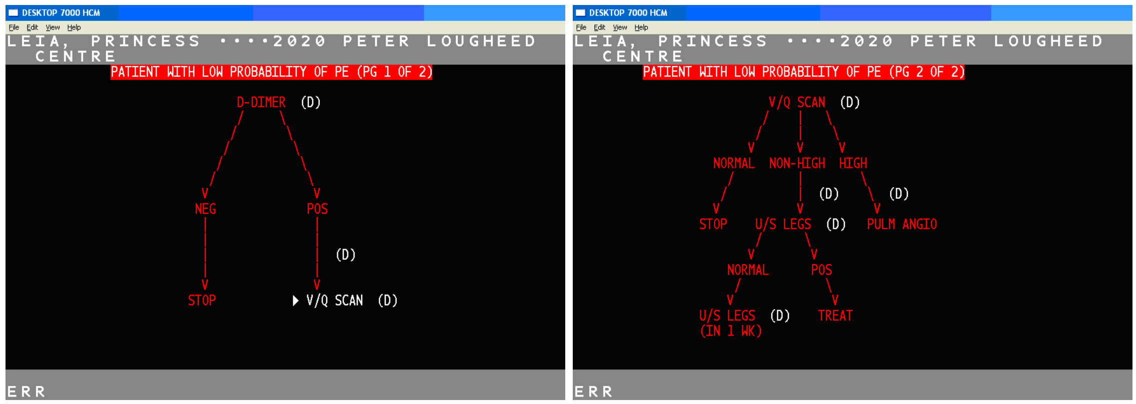



Figure 7.6: Two screenshots of the TDS system used during observational study showing diagnostic tree for low probability PE. Lower branching (screen) in tree is shown after clicking on "V/Q SCAN" in upper screen.

tient's information. No participant used any external cognitive support to manage or order candidate options. The "UpToDate" website was used by two participants for initial PE Wells scoring before TDS was even used. Participants all reviewed symptoms, considered candidates, made a strategy, and decided (preliminarily) on a test before they started using TDS.

For 'Patient A' the Wells score was computed by all but one participant as 4.0 and the other computed it at 5.5. The two participants who used the "UpToDate" website calculated Wells score of 3.0 and 1.0 in addition to that on the TDS system of 4.0 and 4.0. One participant calculated the Wells score of 4.0 but also went back and changed their answer to the last question "PE as likely or more likely than an alternate diagnosis" to determine the score (1.0). This showed a D-dimer test recommendation in both cases, likely building confidence. In summary the test ordering resulted in five D-dimer tests, one V/Q scan, and a CT with a conditional V/Q as a backup test.

The graphic diagnostic tree within TDS was not viewed by any participant. Recommendations based on the tree, however, did suggest to one participant ordering a D-Dimer rather than a V/Q scan. Figure 7.7 shows the form the system recommendation based on a Wells Score. While the recommendation was not considered valid by the participant it was ordered anyway along with the V/Q scan. All other recommendations based on the Wells score reconfirmed what participants had been ordering.


Figure 7.7: Screenshot of the TDS system. Shows form of test recommendation based on score, which may differ from original test being ordered.

### Diagnostic Process for Patient B

All but one participant determined PE to be the candidate condition to pursue, with the other considering deep vein thrombosis (DVT). Again participants reviewed symptoms, considered candidates, made a "strategy", and decided (preliminarily) on a test, and in one case treatment, before they started using TDS. Other candidates were weighed against PE such as DVT, paricarditis, pleuritis, and cellulitis. Further research would have again been performed by some participants before proceeding. One participant used their Palm Pilot to calculate this patient's body mass index (BMI), and stated it would have been overestimated without the tool. Again participants did not use any software tools to manage or order candidate considerations. The TDS system was not used at all by the one participant who immediately decided to treat the patient (treatment was beyond the scope of the observational component of the study).

This patient had Wells scoring of five participants being 7.5, one participant calculated the different deep vein thrombosis (DVT) Wells score of 1.0, and the one participant who began treatment did not calculate a score. Strategies varied more with this patient, five of these strategies and tests were: one ultrasound (U/S) of the leg with V/Q scan, one immediately treating without further tests, one V/Q scan, one V/Q and CT, and one begin treatment and CT with conditional V/Q as backup. Another strategy was initiated as a D-dimer but the system recommended a V/Q for the 7.5 score. Then a CT was ordered with a V/Q as a backup. The final strategy was for DVT with a U/S, D-dimer, and a V/Q or CT. The DVT U/S order required a DVT Wells score which was calculated to be 1.0 and so the system recommended a D-dimer, which was ordered along with the U/S. A V/Q or CT follow-up waiting on the ordered U/S and D-dimer test results.

Again the diagnostic tree within TDS was not directly viewed by any participant. With Patient B the recommendations based on the tree confirmed the tests of five participants. One participant had the suggestion that the D-dimer be replaced with a V/Q. This was accepted, but as a fall-back after replacing the D-dimer with a CT (which was thought to be superior to the V/Q). Another participant considered DVT and when ordering the ultrasound had the system suggestion of a D-dimer based on the DVT Wells score of 1.0. This test suggestion was accepted and ordered before the V/Q or CT which were considered to follow the U/S.

The TDS system again mainly played a confirmatory role. The TDS decision tree corrected a decision in one trial where the more practical V/Q was suggested over the D-dimer, for its superior positive predictive value (an approximately equivalent CT was actually ordered in the end). The decision changing advice was taken by this less experienced participant, showing the system worked as it was designed.

Table 7.2: Questionnaire questions and responses (responses used 5 point Likert scale, SD = strongly disagree, ...).

| | **Questionnaire Question** | **SD** | **D** | **U** | **A** | **SA** |
|---|---|---|---|---|---|---|
| I | When using computers I am comfortable exploring features or options. | 0 | 0 | 1 | 3 | 3 |
| II | I am confident in the system recommendations for ordering a diagnostic test. | 1 | 2 | 2 | 2 | 0 |
| III | The current TDS/OSCAR system helps me practice evidence-based medicine. | 2 | 0 | 3 | 2 | 0 |
| IV | I am confident in my application of evidence-based medicine. | 0 | 2 | 3 | 2 | 0 |
| V | Decision support and test ordering should be integrated into one system. | 0 | 1 | 0 | 3 | 3 |

### 7.4.2 Questionnaire

Following the completion of the task on the two simulated patients a brief questionnaire was given. The questions all used a 5 point Likert scale. The results in Table 7.2 show that all participants did not feel very inhibited about using computers (only 1 was undecided on Question I). Question II reveals some skepticism in the system recommendation, with disagreement on the description of "confident". The next question also shows similar skepticism about the system actually helping with the application of evidence-based medicine. Question IV tells us there is uncertainty in the form of self-confidence in the participants applying evidence-based medicine, which is consistent with the Question III responses. The answers for Question V showed general agreement with the design of integrating test ordering and decision support.

### 7.4.3 Contextual Interviews

After the participant filled out the questionnaire on paper the contextual interview was conducted based around the discussion questions in Table 7.3. These questions were designed to raise the issue of uncertainty in various aspects of the task, as well as pro-

Table 7.3: Contextual interview discussion questions and summarized responses to yes/no questions (U = undecided, - = discursive only response)

| | **Interview Question** | **Y** | **U** | **N** |
|---|---|---|---|---|
| A | *Based on task and use of system* | | | |
| 1. | How would you describe your interpretation of the Wells Score question: "PE as likely or more likely than alternate diagnosis?" | - | - | - |
| 2. | Were you equally confident about all of your answers to Well Score questions? | 2 | 0 | 5 |
| 3. | Did you think about probabilities explicitly as a number during the process? | 1 | 0 | 6 |
| 4. | Did you only want to order a test when using the (TDS) system? | 4 | 2 | 1 |
| B | *Problem domain and use of system* | | | |
| 1. | What would make you more confident in a (TDS) system recommendation? | - | - | - |
| 2. | How would you report the confidence in the diagnosis (so far) to the patient? | - | - | - |
| 3. | Do you feel the system helps you practice evidence-based medicine? | 1 | 6 | 0 |
| 4. | Have you used the diagnostic tree display (in TDS)? | 2 | 2 | 3 |
| 5. | Do you read any additional information provided about tests (by TDS), or do you have it memorized? | - | - | - |
| 6. | Do you use the TDS to share information for consulting others? | 1 | 0 | 6 |
| C | *General characteristics and ideas* | | | |
| 1. | How familiar are you with evidence-based medicine (how long practicing)? | - | - | - |
| 2. | How many years have you used computers? | - | - | - |
| 3. | Have you used software related to evidence-based medicine? | 7 | 0 | 0 |
| 4. | What would like to change about the TDS? | - | - | - |
| 5. | What other information would you like to see to improve a new system? | - | - | - |
| 6. | Where would you prefer to use this system? (i.e. current stations, bedside, home...) | - | - | - |

vide specific design reviews on the current system, and design recommendations and constraints for an improved system. The questions were pilot tested with one experienced physician, after which they were refined. The interviewer often requested clarification of answers to develop ideas further, and if prompted provided clarification of the questions. Table 7.3 shows the questions organized into themes and provides summarized responses, or whether the answer was only discursive. A discussion of the responses follows in the next section.

## 7.5 Discussion

### 7.5.1 Uncertainty in the Wells Score Questions

The Wells scoring questions involve information with associated uncertainty. During the contextual interview the initial discussion considered the question: “PE as likely or more likely than alternate diagnosis?”, which is potentially recursive due to the fact that the answer to this question is being used to judge the likelihood of the PE diagnosis. The observational results showed that almost all participants interpreted it and the data similarly. For Patient A, six answered yes, and the other answered yes and no (sequentially) to see how it affected the score and test recommendation. One of the six who answered yes also compared it against the wording on the UpToDate website. For Patient B, five answered yes, one did not use the system (immediately advised treating the patient), and one answered no. The answer to this question was primarily interpreted to be yes if PE had top ranking of the candidate diseases, although was described by words such as “hard”, “convoluted”, and “confusing”. As one participant considered the deep vein thrombosis diagnosis it is worth pointing out that in the software its Wells score question set has a similar question but the candidate condition is reversed as “alternative diagnosis as likely or more likely than that of DVT” versus the “PE as likely or more likely” wording in the PE question set. This type of inconsistency can also lead to errors.

Five participants stated that confidence in their answers to the Wells score questions varied, and the remaining two said it did not but gave qualifications. Questions with a hard threshold did not allow for uncertainty such as Wells #2 (pulse $>$ 100 bpm) which does not allow for variability in measurement(s) and the effects of any drugs on the heart rate. One respondent stated “... from clinical point 98 or 100 is not different. I should have said yes, but the system told me to say no.” Some participants added margins into these hard numbers. When the system forced the participant to internally resolve the ambiguity it may have left a internal residual, to be carried over. This might be the motivation for one statement about answering yes to the Wells question is PE as or more likely than alternative diagnosis, "..usually say yes to beef up score...", which shows some doubt in the scoring system's ability to reflect their clinical judgment (high pretest probability in this case). Similarly whether tinges of blood constituted hemoptysis (Wells question #5) was not clear to some participants.

### 7.5.2 Uncertainty Representations

Contextual interview Question A3 raised the issue of representation as a probability. Only one participant said they represented the pretest probability as a number in their head. None wrote down numbers and three stated they thought about general categories such as low, medium, and high. One participant said they thought about low, medium, high and how they were mapped to percentages in the PIOPED study [PIOPED investigators, 1990] (and so thought medium likelihood of PE was 20-70%).

Representations for reporting to the patient were discussed in Question B2. As expected from the discussion on A3, no participants reported they would use numbers. Qualitative words and terminology such as "(un)likely", "not an absolute", "ruled-out", "possibility", "confident", "low or high suspicion", "probability high or low", and "primary concern" were used when discussing diagnosis with the patient. Explaining the plan on how to confidently reach a conclusion was the goal of the reporting.

### 7.5.3 Cognitive Diagnostic Strategy Support

While the TDS system's decision tree played for the most part only a confirmatory role in the observational study, this is not a moot point. This confirmation instills confidence in the less experienced user that their decision was a correct one, which is very useful as a teaching mechanism. The more interesting scenario is the system suggestion for the D-dimer test over the V/Q scan to more efficiently rule-out PE. The participant did not believe in the predictive power of the D-dimer in this case, and "did think I'm ordering V/Q no matter". Without extra information provided on these overriding suggestions it is clear the system is viewed as a black box, and uncertainty on the validity of its recommendations exists.

If TDS strongly helped the practice of EBM one might expect more confidence than shown in Question IV. Question V leads us to believe that the participants did accept the premise built into TDS that the test ordering and decision support should be in the same system. However, since this may be that they simply do not like the idea of having to learn two systems, rather than liking the close binding, this response is only taken as a probable indication.

Interview Question B4 asked if the diagnostic tree had ever been seen by the participant. Three responses were a solid "No" and the other weaker responses "may have", "don't recall", "sometimes", and "20%". This ties closely with Question A4 asking if they only wanted to use the system to order a test (presupposing that they had already created in their mind a diagnostic strategy). This question received four definite yeses, two yes and no, and one no. We interpret this as the tendency of one to want to act on a decision once it is made, and the general momentum against changing one's mind.

Question B3 asked if the system actually helped with practicing evidence-based medicine. Only one participant said "yes" while all others gave a mixed "yes and no" qualified response. One responded, "but it almost obstructs me so I have to go back like I made a wrong choice, go back to order the test". The positive side of TDS's support is expressed by one person as "... motives are right, it helps,... trying to." When asked how one could improve this decision support in Question B1, in general the answers indicated that they required more information on what the system was doing to feel confident in it. This is supported by the fact that there was limited awareness of the tree by the participants (Question B4 responses) and so its' guiding principles were not transparent. Scepticism of any system guidance based only on limited questions and answers that did not capture substantial clinical judgment was also stated by participants. References, details of how the current patient matches the patient profiles in studies, effects of tests on pretest probabilities, example scenarios, and other information were given as information that could lend credibility to the system recommendations.

### 7.5.4 Task Model

Based on this study we created a task performance model of the diagnostic process to help structure the creation of support. The diagnostic process can be described as: 1) the collection of information, 2) the interpretation of that information as evidence of suspected diseases, 3) making a plan on how one can optimally determine the true disease, and 4) a decision to either make a diagnosis, consider other options, plan-further, or to collect more information. In some cases treatment based on a likely diagnosis may begin before the practitioner is satisfactorily confident of the diagnosis. Time constraints (patient mortality) force this use of a most likely diagnosis for treatment, which may provide more information as to the accuracy of the diagnosis.

We provide a task model in Figure 7.8 which also delineates aspects of external and internal uncertainty. The sub-tasks are observation and testing, inferring candidate diseases, diagnostic planning, and the decision for the next action (including diagnoses). These processes, in general, follow the temporal ordering shown in the figure, thus the uncertainty is compounded as one step leads to the next. The theoretical flow of the sub-tasks is provided, but in practice there may be no clear distinction between some sub-tasks. As shown in Figure 7.8 the sub-tasks may be repeated by jumping back from the decision sub-task.

Our discussion of the diagnostic task was based on the goal of understanding uncertainties as they arise in the process and as a result there may also be other valid discussions that focus on other aspects of diagnosis. For more detailed information on the complex task of medical diagnosis, one can refer to a text on the subject (e.g. [Knottnerus, 2002]). However, the task performance model as shown in Figure 7.8, can be useful in informing visualization and interactive support.

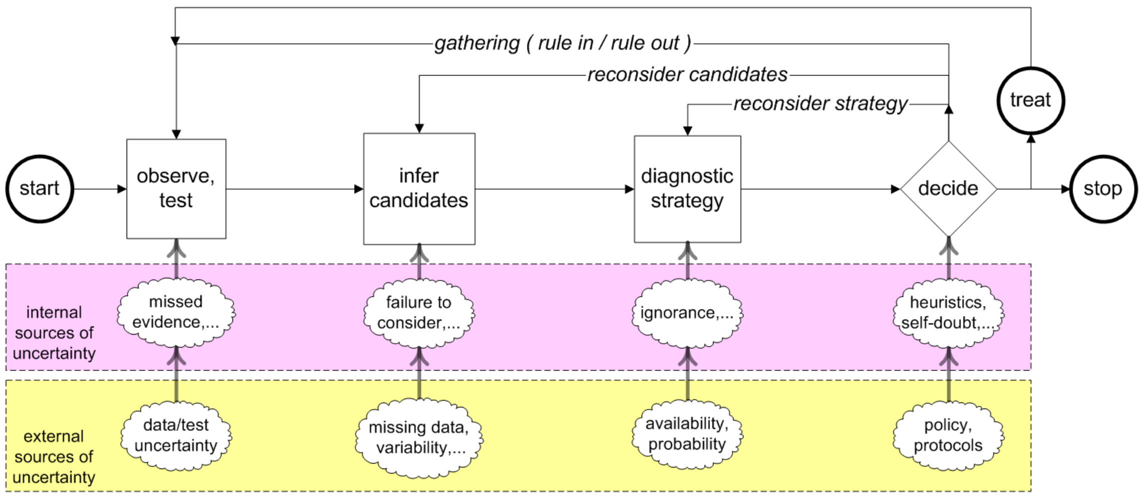



Figure 7.8: Medical task performance model: diagnosis sub-task decomposition showing associated uncertainty.

### 7.5.5 Design Factors

#### Sociological Issues

To understand physician's reactions to uncertainty Gerrity et al. [1990] developed a rating scale from an analysis of 428 questionnaires. Their reaction to uncertainty scale is based on two main components: stress from uncertainty and reluctance to disclose uncertainty. This and the earlier work of Fox [1980], may provide insights into the sociological processes relating to uncertainty in medicine.

Gerrity et al. [1992] provided a model of factors influencing a physician's reactions and behaviour under uncertainty. It included five major components: the patient, the condition, the physician, the test or treatment, and the organizational structure. They categorized reactions to uncertainty into: patient-physician relationships, physician-colleague relationships, professional norms, self-esteem as a physician, bad outcomes, missed diagnoses, malpractice worries, patient referrals, and test ordering. The bulk of our direct observational work relates to the test ordering category, but the contextual interviews touched on issues from all the other categories except malpractice worries. Many of the behavioural and cultural issues extend beyond our current scope, but may still be important factors to consider in any design.

**Cognitive Load and Stress**

Medical diagnosis requires the utilization of a large amount of background knowledge. Managing that amount of information itself may cause stress, particularly when trying to deal with task constraints as well. Visualization may provide cognitive support by offloading some of that burden.

Timely decisions will be required for critically-ill patients and so there is great pressure to diagnose quickly. Mental as well as physical fatigue may add cognitive stress especially for residents doing very long shifts and work weeks. Residents will be learning vast amounts of new information and procedures and thus it may be useful to customize support for use in a teaching role.

**Self-confidence**

The knowledge-base of the physician will be utilized in all the sub-tasks. Accurate introspection on the validity of any internalized rules is therefore important for self-calibration. This will guide the physician to forage for more information at any stage in the process, or to request consultation with another physician. As all but one participant said they did not use the TDS system to share information, facilitating this may be something to explore, but it may have to also overcome the reluctance to disclose uncertainty found by Gerrity et al. [1992].

How best to apply evidence-based medicine is still under discussion [Ghali et al., 1999, Ghali and Sargious, 2002]. Ever changing evidence requires the constant reviewing of new information by the physician. This will naturally drive self-doubt as what was the best strategy yesterday, may have been discredited today. Given this process it is not surprising to hear one of the participants' comments, "I've been doing it longer but don't think I'm good at it".

## 7.6 Design Implications

Using our task model we found significant uncertainty in each sub-task. The over-riding impression from the combination of the diagnostic task observations and the interview discussion was that the physicians welcomed the possibility of support, while objecting to system suggestions that were not fully explained or were provided at inopportune moments in their own decision processes. In light of this, we discuss each sub-task and suggest factors for design consideration.

### 7.6.1 Sub-task 1: Observations and Testing

The observations in this sub-task refer to medical observations during the diagnostic process. Initial medical observations may be from patient histories, patient exams, patient charts, or test results. These observations or measurements may be qualitative or quantitative. Multiple tests and observations will often be acquired based on standard practices, even before considering candidate conditions or diagnoses.

Many types of uncertainty exist in this sub-task. Uncertainty in physical measurements is similar to uncertainties in other scientific areas of measurement (accuracy, precision). Measurements will also have uncertainty from temporal variability. Continuously varying vital signs if represented by a single number without its associated uncertainty may result in uncertainty in confidence regarding the number given.

Verbal patient responses are also full of the ambiguity of conversation, and the same may be said for written information on charts. Misinterpretation of questions may occur, and patients under distress may obviously not report all information accurately. Patient reporting of information is also naturally biased by what they think is important and relevant.

Each diagnostic test usually has a sensitivity and specificity or a likelihood ratio, potentially with uncertainties in these values. These numbers depend on patient population profiles as well as research protocols and so their applicability may come into question. Even highly accurate tests and measurements often do not directly relate to specific conditions and so their indicative weight is uncertain. Many treatments also provide test-like results on the basis of patient response.

In our study we found various questions in the Wells scoring system had ambiguous interpretations. We hypothesize that the truncation of uncertain categories in the mind when mapping to overly specific answers (Yes and No in this case) may result in significant accumulated truncation (or round-off) error. A person may decide to carry this truncation error in their mind, but with the current TDS system there is no way to add extra points at the end. While this might be dealt with by providing more ambiguous answers, one participant described this as a bad solution since it is better to force a yes or no, or likely you would end up answering maybe to everything. Any new system should allow for easy modification of the answers or final score to allow expression of this type of uncertainty.

#### Considerations and Design

Visualizing repeated measurements as a time varying function could avoid uncertainty in confidence, and this type of display was requested, albeit indirectly, by some participants. For example, a heart rate at a single reading is tough to interpret even assuming limited error; tachycardia would only be a confident conclusion given multiple readings. A participant stated that one must also bear in mind any drugs in use that have an effect on heart rate. For the case of heart rate, a graph over time with annotation showing drugs could reduce the uncertainty. Many of the observation and test results are uncertain and a graphical representation may be easier to digest for multiple readings. As an example, a graph of points with error bars may be easier to review than a table of numbers along with confidence intervals, or +/- error margins [Alonso et al., 1998].

Five participants stated they found it difficult to use the TDS system to find older information, and so tests would often be ordered again. As there is too much information for the physician to manage in their head it is important that this data can be reviewed on demand and with minimal effort at any point in the task flow. Therefore current and historic data graphs (with any aggregation showing uncertainty) should be easily viewable in order to accurately understand and weigh each piece of evidence and to reduce any fear of recency bias [Powsner and Tufte, 1994, Tufte, 2006].

The TDS system hid the responses to the Wells scoring questions (Y or N). This type of visual feedback is important to catch data entry errors. Figure 7.5 showed a screen shot after answering question one, with question two highlighted in white. This design may be due to a technology constraint but transparency of both answers and scores for each question should exist. If scoring is thought to be potentially biasing it could be revealed after all questions are answered. This relates to the potential danger of a participant's cognitive heuristics, since, for example, the automation bias [Skitka et al., 1999] may cause a system recommendation that was the result of data entry error to be accepted without prudent scepticism.

### 7.6.2 Sub-task 2: Inferring Candidate Diagnoses

Weighing the initial evidence to form a list of possible causes is the core of this sub-task. The observations are noisy with a large amount of irrelevant data, and the expected symptoms of any condition are often not clearly defined. This can be categorized as fuzzy pattern to fuzzy template matching. It may result in multiple conditions to consider, each with some ranking.

This sub-task involves the comparison of potential candidates and was not supported to any extent by the TDS system. While short lists of candidates are being made system support could provide data management. On-line resources such as "UptoDate", "MD Consult", "PubMed", "Cochrane Reviews", "Medline", "Web of Knowledge", "American

College of Physicians", and others were reported as being used, with frequencies of use as high as hourly by the residents.

Given a set of observations and any *a priori* mindset an associative process may quickly provide a candidate set of inference rules to apply. While in our study we used the leading statement "given a PE mindset can you consider this patient", some participants did make a short list of candidates they were considering besides PE. To what degree cognitive load and other stress affect this associative process would be worth studying. For example one potential associative bias is when the size of a set is estimated by availability of instances to the mind, the retrievability of instances will bias the estimate [Kahneman et al., 1982].

#### Considerations and Design

Simply improved management of on-line resources could be of use for this task. Linking of symptoms to candidates with strengths could aid the formation of top consideration lists. When multiple candidates proceed to the diagnostic strategy stage, managing the ranking of various hypotheses could also benefit from external support. Reducing cognitive load by offloading resource management for this task should aid the reduction of uncertainty by allowing more candidates and potential missing data to be considered.

### 7.6.3 Sub-task 3: Diagnostic Planning

This sub-task involves prioritizing the potential diseases and forming an optimal ordering of tests to rule-in and rule-out candidates. The strategy will be referred to as a diagnostic decision tree or pathway, but the "tree" may in fact be a cyclic graph. The TDS system provides a set of diagnostic strategies built into the test ordering process. This system is an evidence-motivated predetermined decision tree, or protocol, which is triggered when a test tied to a specific condition is ordered. In the case of PE, a Wells Score is required before any test on the pathway can be ordered, and the score may trigger the recommendation of test that is different to the one being requested. This recommendation is slightly out of order in an initial pass through our task model, as this strategy related guidance should have been provided before the decision for a specific test was made, and this will be discussed further in the decision making sub-task section.

Comparing the pretest probabilities of different conditions will be the initial basis of what strategies are to be considered and developed. As these pretest probabilities are only estimates, uncertainty will exist in their values. Practical constraints and characteristics of each test, such as specificity, sensitivity, availability, timeliness of result, patient harm, and cost, must be added to the equation and will need to be considered. While specificity and sensitivity are directly related to the calculation of uncertainty, each of these other test attributes also have uncertainty associated with them.

Decisions about strategies are also included in this sub-task. Given multiple considerations for possible conditions to investigate there may be ambiguity in which to address first. Ambiguity can often be resolved by considering (ruling-out) the most time-critical diseases first, but test availability and cost may also be weighed into any strategy. Similarly given a single condition, (e.g. PE) there are multiple options to proceed toward ruling-in or out the condition. Bayesian reasoning may be performed to compute post-test probabilities for different test outcomes and strategies or pathways. Eddy [1982] has reported on various cognitive problems in applying Bayes rule and confusion between retrospective accuracy and predictive accuracy both in practice and the medical literature.

### Considerations and Design

We suggest that the diagnostic decision tree should be shown (visualized in some form) at any point in which they are expected to be guided by, or conform to, the predetermined strategy. Visualizing the tree may reduce unnecessary uncertainty as to why the system makes suggestions, and increase confidence when following or disregarding recommendations. The tree should also be supported by information to justify the decision recommendations set down in the decision tree's branching structure.

Support could provide alerts based on evolving research, and easily link the physician to newly added information relevant to a diagnostic decision. This could improve the lack-luster confidence physicians reported in the TDS system recommendations. The potential post-test probabilities are fundamental to choosing the most efficient diagnostic strategy; this is because when the post-test probability of all test outcomes does not cross a decision threshold, then the test may not even need to be ordered.

### 7.6.4 Sub-task 4: Decision Making

This mostly internalized step is based on the weighing of evidence from all the previous sub-tasks and choosing the appropriate next action. This final decision relies on the understanding of various information and the uncertainties in them, along with the ability to compare significantly different types of information and uncertainties. Ultimately a threshold may be crossed for which the decision to treat or stop will be made.

This sub-task's role is one of specific actions (test, treat, stop) as well as integrating evidence as the entire task loop is repeated. These action triggering thresholds may not be clearly defined, but are the basis of ruling-out or ruling-in a particular condition. External uncertainty may come from poorly defined hospital policy or protocols, as well as the integration of patient utility. Enhancing the communication of uncertainty to the patient should also be considered.

#### Considerations and Design

Dawes et al. [2003] summarize the arguments from multiple studies that actuarial judgment may often be superior to clinical judgment. However, one major limitation of actuarial judgment is its inability to capture all exceptional cases in the rule base or decision tree. While the statistical strength of actuarial judgment is basically the rationale for providing decision trees (as in the TDS system), it appeared that the participants were not convinced.

One participant stated that if they are to follow guidelines (i.e. a decision tree) based on a population profile they need to be shown how the current patient's profile fits into that population. Exposing this relationship would thereby be one way to increase confidence and reduce uncertainty in applying guidelines or using actuarial judgment. The automation bias was also likely seen in our study as in one case the TDS system recommended a test that was not the desired one, it was ordered anyway to satisfy the system even though it was not considered a useful test.

Information that could lend credibility to the system recommendations were such things as: references, details of how the current patient matches the patient profiles in studies, effects of tests on pretest probabilities, and example scenarios. Thus integrating the evidence behind any recommended decision trees is important for the user to see, or easily access. For similar reasons we think it may be useful that the decision tree is available for context when viewing detail information.

Participants' responses indicated that the decision support should be available before the test ordering is initiated. As was noted earlier, using our task model a decision has already been made when the test is to be ordered. Once a decision is made to order a test there is cognitive context switch needed to go through the steps required to order it, and so the TDS system was then felt to impose a hidden strategy on them and this support may be more of a nuisance at that point. When support is not requested users may in fact work around the system support, as was stated by one participant, "tip on the street is put 3 in to by-pass the pathway".

While participants mostly agreed that the decision support and ordering should be integrated into one system, we believe the lack of confidence in system recommendations is also confounded by support coming too late in the reasoning process. The integration of decision support before the test ordering step could provide efficiency gains by avoiding backtracking and potentially eliminate extra tests that are "committed" to before considering all the options in a strategy. If the option exists to go directly to test ordering without decision support, then it raises the problem of motivating a user to use diagnostic support when it could be beneficial.

## 7.7 Summary

Our study examined some of the complex issues involved in evidence-based medical diagnosis. The observations and interviews provided insights into this difficult task, exposing a variety of uncertainties in many components. From this we derived a task model that provided a basis for decomposing the uncertainty and informing the design of new support.

Data and cognitive uncertainty were found to be important factors to consider when determining what support might be beneficial to the physician. Cognitive support for each sub-task should only be provided on demand as different physicians will only need or welcome assistance at specific points. Support for this task must also be as transparent as possible as accurate confidence is crucial for any system to be clinically valuable.

Communicating the evidence behind any system recommendations is paramount to the physician judging their applicability. This suggests a strategy of providing access to visual evidence at all levels of detail while revealing how it relates to the current context. Utilizing these recommendations for developing new system support will be an area of future investigation.

As this chapter is Part I of a II part series further interpretation of these design implications will come in the next chapter. Some analysis of the aforementioned observational study is also provided in the next chapter to aid in understanding the motivations of the separate visualizations developed for supporting this task.

## 7.8 Acknowledgements

We would like to thank Marilyn Gore for her assistance on the use of the TDS system for the observational component at Foothills Hospital, Calgary. We would like to thank all the participants in our study and the Ward of the 21st Century for providing logistical support.

# Chapter 8

# Case Study in Medical Diagnostic Reasoning

# Part II: Visualization Support

> Probability is expectation founded upon partial knowledge. A perfect acquaintance with all the circumstances affecting the occurrence of an event would change expectation into certainty, and leave neither room nor demand for a theory of probabilities.
> – George Boole (1815 – 1864)

In this chapter we conclude the case study in evidence-based medical diagnosis. Using the results from the observational study, contextual interviews and task analysis provided in the previous chapter, multiple visualizations were developed to provide cognitive support for different aspects spread across the diagnostic task. Visualizations were created that relate to each sub-task and are discussed along with initial evaluation results[†].

## 8.1 Introduction

To ground this work in current practice in evidence-based medical diagnosis, we performed an observational and contextual interview based study, as described in Chapter 7. That study and the resulting task model provided for a structured investigation of the uncertainties involved, and served as a basis for the research presented in this chapter. Focusing on the role of uncertainty we developed multiple visualizations. These visualizations are designed to improve comprehension and performance by incorporating the uncertainty relevant to the task of evidence-based medical diagnosis.

Initial requirements generated from the design implications and the task model from the previous chapter guided the developmental process which utilized multiple iterations

[†]Portions of this chapter have been previously, or will be, submitted for publication. Therefore “we” refers to Torre Zuk, Sheelagh Carpendale, William Ghali, and Barry Baylis

of participatory prototyping. The final visualizations from the last iteration of prototyping will be presented along with the analysis supporting their design. The visualizations are presented in conjunction with the evaluation results from a focus group performing a pluralistic walkthrough of the entire system. The next sections will briefly summarize both the development and evaluation methodologies.

## 8.2 Development Methodology

Initial analysis of the observational study provided a list of potential data and reasoning uncertainties in the process. Their relationship to decision strategies recommended by the hospital system was also noted. Similarly the contextual interview component of the study also provided ideas on what could be visualized. These formed the design implications from the previous chapter which were translated into functional requirements for the system. Based on this a visualization system was developed to reveal the uncertainties fundamental to the task. The visualization was designed to suit a comparable display platform to the existing system (desktop PC).

### 8.2.1 Participatory Prototyping

In multiple sessions with one or two physicians the latest visualizations were presented and discussed. Feedback was used to refine the various components. Other demonstration sessions with Information Visualization experts also provided feedback for iteratively refining the visualizations. The current state of this refinement process will be detailed in the following sections along with important motivations taken from the study findings.

## 8.3 Evaluation Methodology

Prototypes were shown to physicians on multiple occasions to collect feedback on the visualizations and the system. The results of the informal evaluations provided by the collaborating physicians were then used to refine the system design. After three passes of this participatory prototyping process a qualitative focus group based evaluation process was performed.

### 8.3.1 Participants

Participation involved a group of general internists at the Foothills Hospital, Calgary, Alberta. This group was chosen as pulmonary embolism is a condition they likely have to diagnose on a regular basis. The group that filled in the written component was composed of three women and six men. These participants' ages were: three in the range 30-39, three between 40-49, and one in the range 50 and over. All participants were well experienced in practicing evidence-based medicine as shown in Figure 8.1. A few additional physicians were present during the session but did not complete any written component, however, they were free to add to the discussions.

### 8.3.2 Methodology

The chosen form of evaluation was based around a focus group performing a pluralistic walkthrough [Bias, 1994] of the system. This evaluation style involves walking through a user scenario discussing the role of the visualizations at each step. The scenario used was working through the diagnosis of pulmonary embolism. The components of the evaluation that targeted the aspects of reasoning support could also be considered a cognitive walkthrough [Wharton et al., 1994]. Both methods were developed for usability evaluation and hence our use of them was atypical, by not being restricted to only the issue of usability. A walkthrough-based methodology was chosen for its potential to quickly ex-

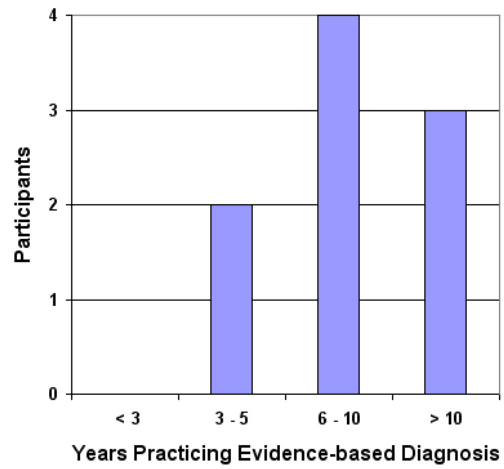



Figure 8.1: Participants' experience in practicing evidence-based diagnosis.

pose the features of the visualization system for the sub-tasks of evidence-based diagnosis. The walkthrough was directed by an experienced visualization designer, and facilitated by another visualization expert and two experienced physicians.

Participants were given a questionnaire booklet with images showing individual graphic components corresponding to the visualizations in the walkthrough and asked to provide their written feedback either as the walk-though proceeded or at the end. Questions could be asked at any time to further discussion about the separate features. In the booklet a final questionnaire component contained questions regarding demographics and general aspects of evidence-based medicine.

The pluralistic walkthrough began with setting the stage for utilizing the system to support the evidence-based diagnosis. The walkthrough proceeded using the system visualizations to support the process beginning with visuals related to sub-task 2, then 3 and 4 before returning to sub-task 1 at the end. This order was chosen as we were more interested in these aspects, and given time constraints were not guaranteed to cover all of our components.

### 8.3.3 Environment

The focus group occurred in a large meeting room in part of the medical center complex at Foothills Hospital. The walkthrough utilized a laptop with the running application projected to a large screen at the front of the room. Participants sat at desks laid out in a "U" shape around the screen. The evaluation occurred at a regularly scheduled group meeting where physicians normally discuss a variety of work related topics and issues. Participants were at work and so the potential existed for individuals to be interrupted by being paged, and a couple did leave for brief periods and then return. The main evaluation lasted just over an hour, but two participants interested in more details remained longer for further individual demonstrations and discussions.

## 8.4 Visualization Design and Evaluation Results

The core of applying evidence-based medicine is the use of tests to statistically rule-in or rule-out a diagnosis. This may be done using Bayes Theorem for the updating of probabilities based on evidence (test outcomes); to provide cognitive support for this we brought this aspect directly into the visual interface. The recommended decision tree and expected changes in probability was exposed as the user interface rather than being a model hidden to the user. This initial design choice was based around working with the hardware constraints of the systems currently used. Dealing with other technology such as very small or very large displays was excluded from our initial design to simplify the process and reduce any deployment and evaluation issues.

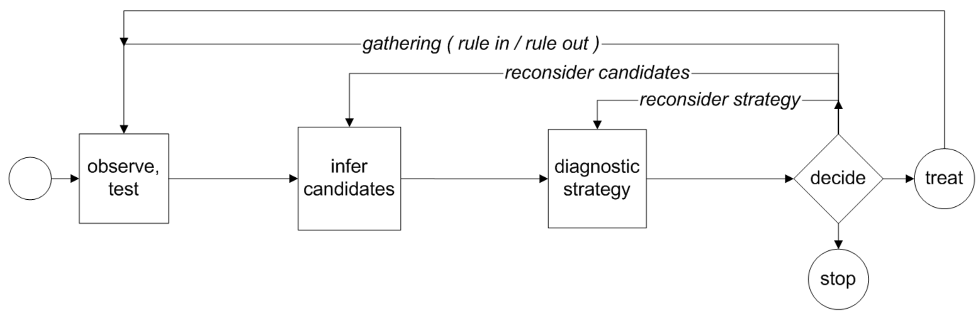



Figure 8.2: Evidence-based medicine diagnostic task performance model.

### 8.4.1 Overview

For this section we will use our task model to frame an overview of the diagnostic task process as observed, stepping through the sub-tasks (see Figure 8.2) we provide informal system requirements consolidated from the design implications from the previous chapter. For each sub-task we will then describe the visualizations that we designed to support them and provide the evaluation results.

#### Design Overview

The system is composed of three main components shown in Figure 8.3. The main view (top-left in the figure) shows a decision tree visualization aimed mainly at Sub-task 3: Diagnostic Planning, but with aspects related to the other sub-tasks as well. The bottom view relates mainly to Sub-task 1: Observations and Testing, and provides graphical review of historic test results. The right view integrates information relating to all sub-tasks, but in the figure shows the nested-set for computing post-test probabilities based on D-dimer test results (sub-task 1). All views can be laid out based on user preferences. Another view for the sub-task Inferring Candidate Diagnoses is not visible, but all views will be described in more detail in later sections.

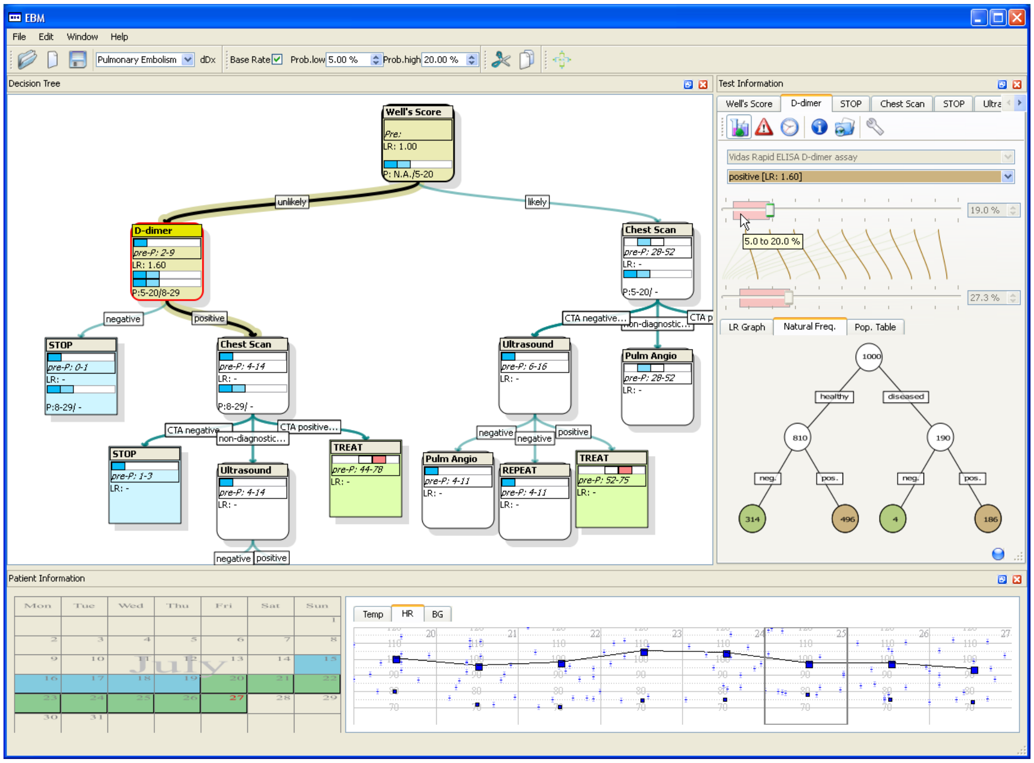


Figure 8.3: Overview of the system to provide support for evidence-based medicine.

### Evaluation Overview

The focus group walk-through proceeded in only a slightly different order than the task model ordering, in that the test result selection and graphing over time was deferred to the end, for reasons stated earlier. Many of the written feedback questions relating to the walk-through used a 5 point Likert scale and I will summarize responses with the following notation: $\frac{n1}{SD}\frac{n2}{D}\frac{n3}{U}\frac{n4}{A}\frac{n5}{SA}$, where $nX$ are the total number of participants who responded to each category, and $SD$ = strongly disagree, $D$ = disagree, $U$ = undecided, $A$ = agree, and $SA$ = strongly agree.

### 8.4.2 Sub-task 1: Observations and Testing

This sub-task involves the gathering of raw data from observations, information foraging, or testing procedures. Based on a translation of the design implications from the study results in the previous chapter, and other related research as noted, we have derived potential functional requirements relevant to this sub-task:

- S1.1 Provide for encoding of, and working with qualitative measurements. *Rationale*: Almost all of the physicians worked with unquantified probabilities such as clinic assessments.
- S1.2 Simplify retrieval of old data. *Rationale*: Old test results were needed but were difficult to find and often had to be reordered, this also counters any recency bias [Tufte, 2006].
- S1.3 Report this historic data graphically. *Rationale*: This was requested, and has been shown superior with related data [Alonso et al., 1998].
- S1.4 Expose the sensitivity of probability estimates to actuarial scoring questions (e.g. Wells score). *Rationale*: This allows the physician to quickly acquire confidence in the recommendation even if they are unsure about a particular answer or judgment.
- S1.5 Make the physician's input, any actuarial scoring system, and its applicability visible. *Rationale*: Allow the physician to develop prudent scepticism and avoid automation bias [Skitka et al., 1999].
- S1.6 Provide support for Bayesian interpretation of test results. *Rationale*: Visual support may assist in the application of Bayes Theorem [Gigerenzer and Hoffrage, 1999, Sloman et al., 2003] and may mitigate any potential cognitive heuristics and biases. Heuristics and biases have been shown to effect some probability estimates, for example, insensitivity to prior probabilities [Kahneman et al., 1982].

Cognitive heuristics and biases should be kept in mind as user constraints. Chapman and Chapman [1982] aptly point out that "test results are what you think they are". While their work was with psychiatrists' and clinical psychologists' interpretations of Rorschach and Draw-a-Person tests, it has been found in general that people tend to find correlations between things that have strong mental associations, even when the correlations do not exist [Kahneman et al., 1982]. In estimating post-test probabilities one very relevant cognitive heuristic is an insensitivity to prior probabilities (base rate neglect) and so Bayes Theorem is not applied. Any support provided for this sub-task will hopefully weaken these potential constraints.

### Visualizations

In order to address S1.1 and S1.2, we paired a calendar driven query of test results with a time-based graph of selected results. With this visual interface one can see the test result history, revealing when tests were conducted and reviewing test variability over the selected time periods. Figure 8.4, shows temperature measurements, graphing individual measurements with error bars, as well as daily min, max, and means and the average trend.

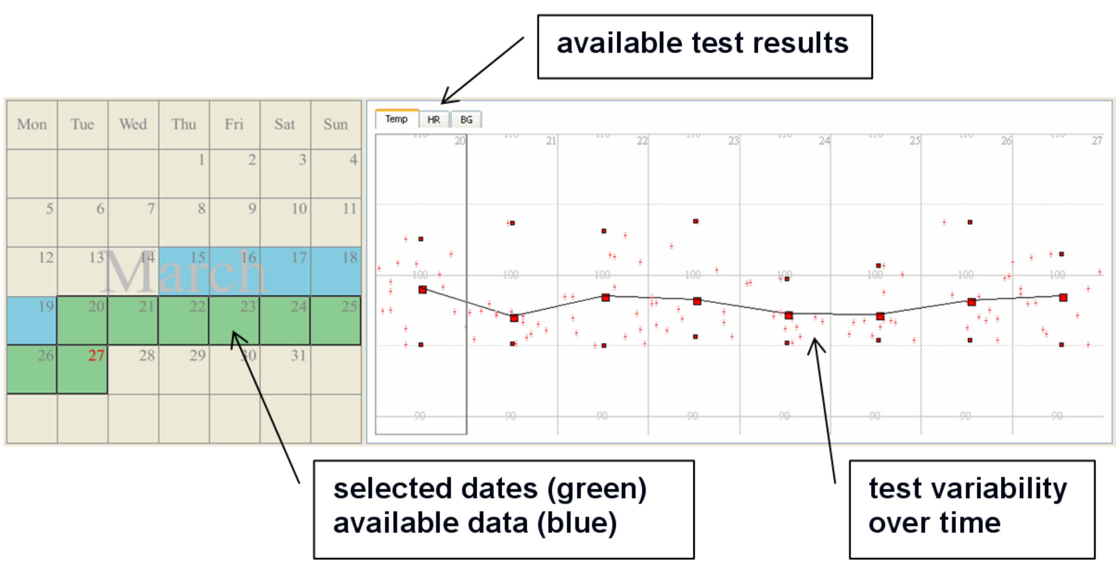



Figure 8.4: Test result variability and precision.

Another step toward addressing S1.1 is provided by allowing the use of a probability range to specify the base rate. Thus “low” or other coarse encodings of pretest probabilities could be defined, but quantitative numbers were required to compute the post-test probabilities. Allowing the default base rate to be selected from a list of published medical studies may also help meet this requirement. S1.4 was satisfied by immediately updating the decision tree visualization shown in Figure 8.5. While the TDS system hid the responses to the Wells scoring questions (Y or N), we address S1.5 by making the actuarial scoring answers visible and immediately updating any recommendations based on the score.

Requirement S1.1 was only partially addressed by letting probability ranges be entered as base rates. Thus “low” or other pretest probabilities could be defined, but quantitative numbers were required to compute the post-test probabilities. Allowing the default probability ranges to be selected from a list of published medical studies may also help meet this requirement.

Specific probabilities for a given test can be manipulated with the probability slider shown in Figure 8.6. It contains a pretest probability slider along with mappings showing connections to the lower derived post-test probabilities slider. These contour-like mappings show the compression, shifting, and expansion effects on probability and are colour-coded to compare multiple outcomes and their respective likelihood ratios. In Figure 8.6 the green mappings show the selected negative test outcome, while the positive mappings are still visible but de-emphasized with opacity. The mapping lines can be read to see the strength of the D-dimer is in ruling out PE rather than ruling in PE, as the downward shifts are more conclusive. Pink bars on the sliders encode with their length the pre- and post-test probability range for a particular study population profile or the base rate range the physician assigned for a patient.

For any given test we provide an interactive visualization of the effect of various outcomes on probability controlled with the probability slider as shown in Figure 8.7. The top portion of the figure contains the slider for exploring pretest probabilities previously

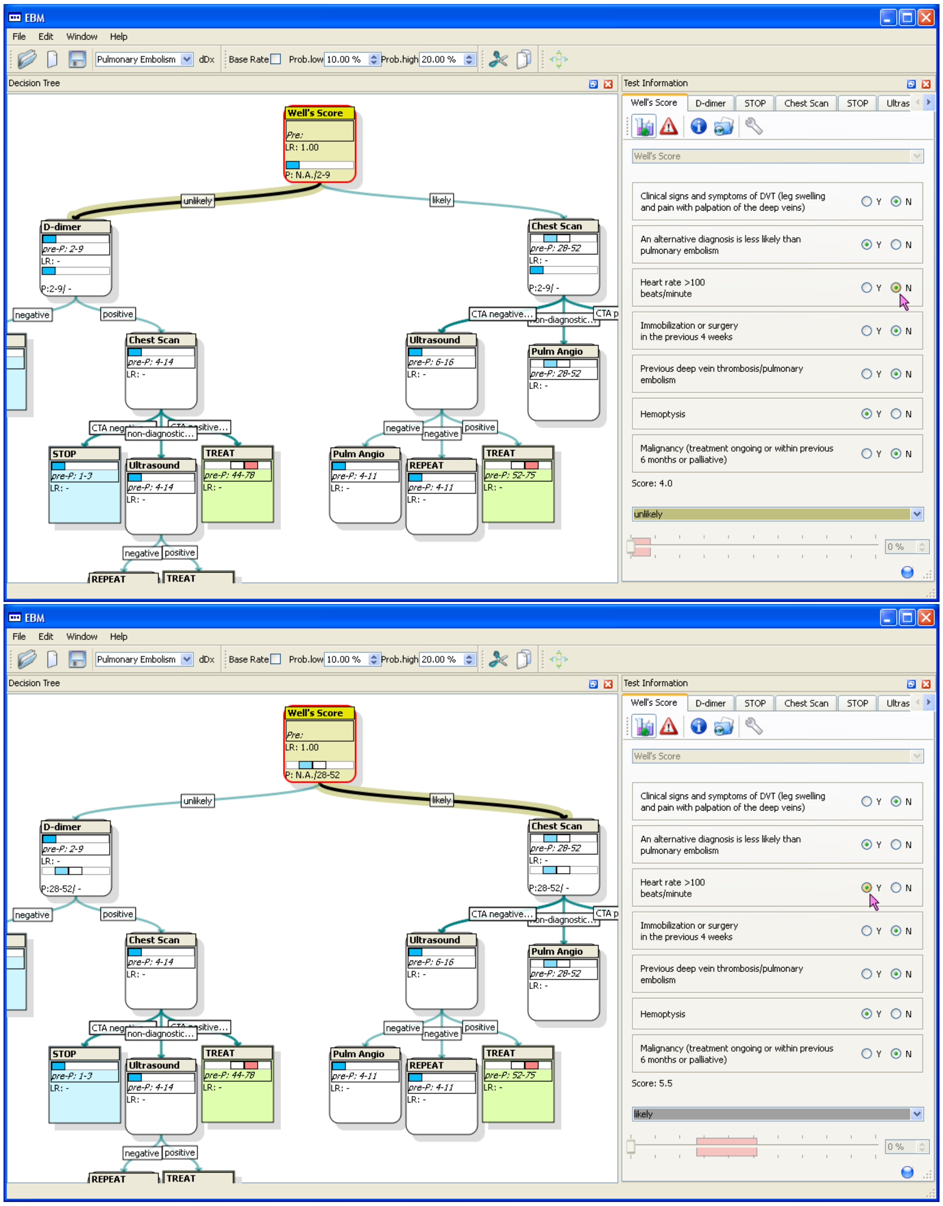


Figure 8.5: Decision Tree and Wells Scoring screens. Wells scoring screen shows answers, context of decision tree, and allows sensitivity of individual questions. Top to bottom shows changes based on changing the answer to Question #3 from No to Yes.

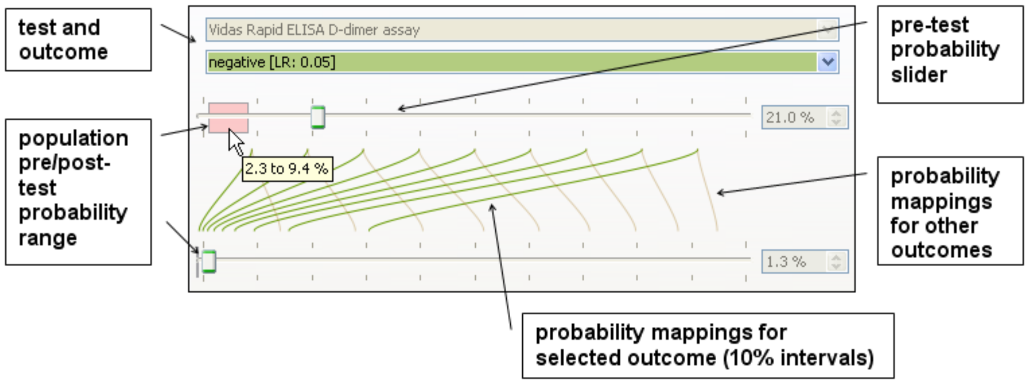



Figure 8.6: Pre/post-test probability slider showing probability response for negative D-dimer test.

described. The lower portion of Figure 8.7 shows one of three selectable representations for visualizing pre and post-test probability. The three representations are: a graph of the probability function for a likelihood ratio, natural frequency nested-set (tree) for a simulated population, and a natural frequency 2x2 table. These are shown adjacently in Figure 8.8 for better comparison. Natural frequencies have been found in some cases to be more easily utilized than probabilities for Bayesian reasoning [Gigerenzer and Hoffrage, 1999]. Hover queries allow additional pop-up textual representations of most graphical information. This visualization mainly relates to S1.6.

While uncertainties in sensitivities and specificities (and likelihood ratios) can be used in modifying probability distributions [Winkler and Smith, 2004], we chose to visualize only the confidence intervals to keep the complexity lower. We provide a visualization of likelihood uncertainties (bottom half of Figure 8.7) based on printed publication formats (e.g., Habbema et al. [2002], Roy et al. [2005]), adding the interactivity that is key to the intuitive understanding of the effects of likelihood ratio uncertainty on post-test probabilities.

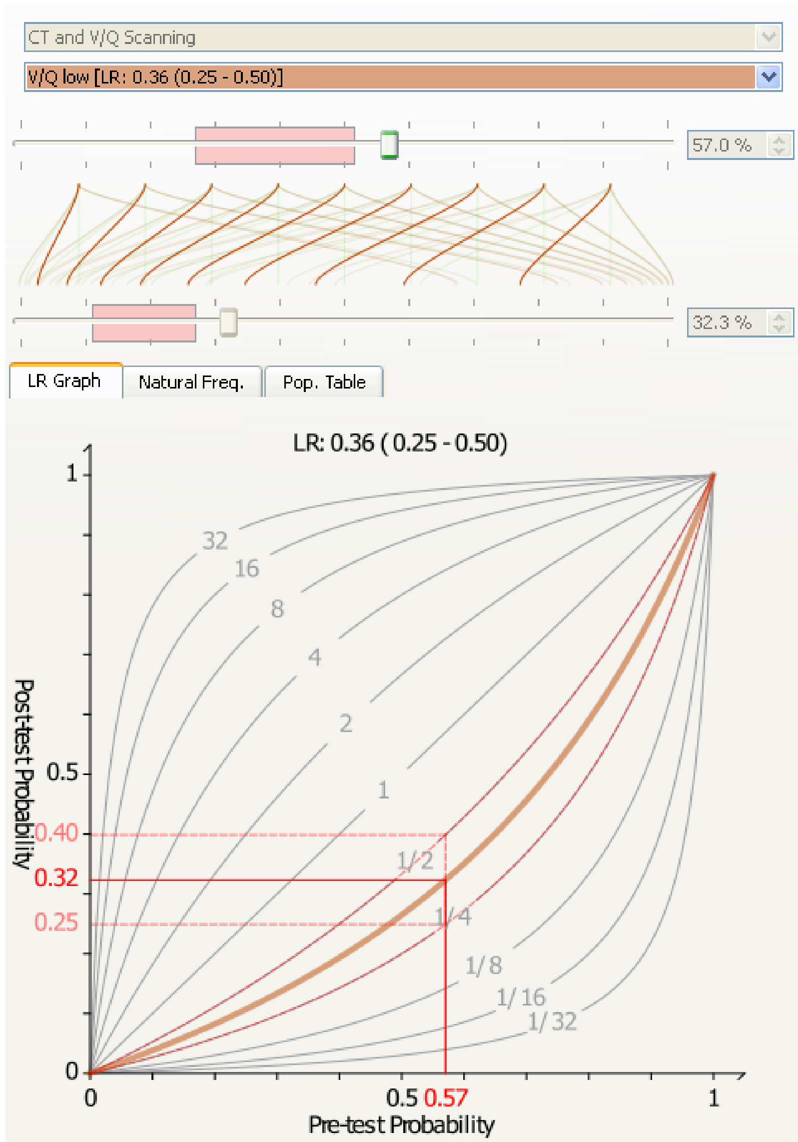



Figure 8.7: Two representations of the post-test probability function with the lower showing the effects of likelihood ratio uncertainty.

### Evaluation

The patient data visualization presented the potential to easily see previous test results and to review test variability over time. This is shown in Figure 8.4, providing a graph of

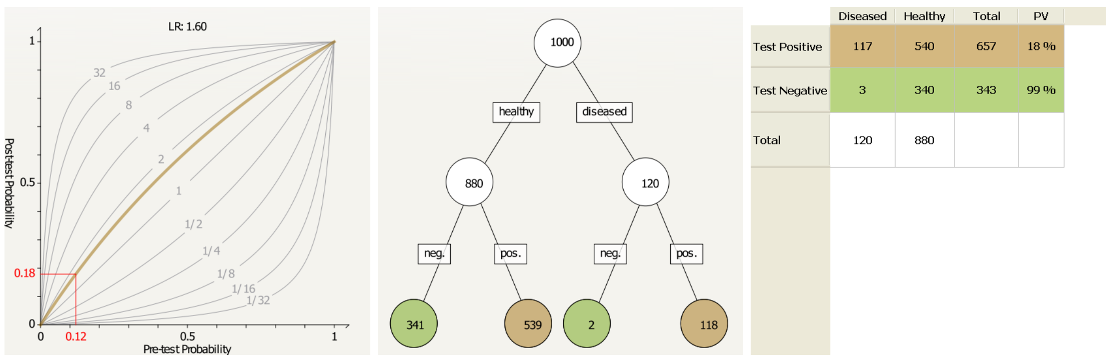



Figure 8.8: Representational options to aid the interpretation of diagnostic test results, from left to right: post-test probability function, natural frequency nested-set, and natural frequency table. Positive results are brown and negative green.

temperature measurements. “This test result variability and precision visualization would be useful” received the responses: $\frac{0}{SD}\frac{1}{D}\frac{4}{U}\frac{2}{A}\frac{2}{SA}$.

For understanding test results multiple visualizations were presented. Three questions were specifically targeted at the probability slider visualization: “Seeing the relationship between pre and post probabilities is useful”: $\frac{0}{SD}\frac{0}{D}\frac{0}{U}\frac{6}{A}\frac{3}{SA}$; “This visualization showing pre and post probabilities is comprehensible”: $\frac{0}{SD}\frac{0}{D}\frac{2}{U}\frac{6}{A}\frac{1}{SA}$; and “This visualization would assist my interpretation of the test results”: $\frac{0}{SD}\frac{0}{D}\frac{0}{U}\frac{7}{A}\frac{2}{SA}$. One comment for this visual was that,

> *Would be great to be able to access this for other types of tests too (e.g. ferritin for Dx of iron deficiency ...*

(other tests than just those on the decision tree).

Three related visual representations of this information shown previously in Figure 8.8 were also rated, the summary is presented in Table 8.1. The probability graph representation also displayed the effects of uncertainty in the likelihood ratio as shown in Figure 8.7. When asked regarding the probability graph if “the uncertainty aspects (confidence intervals) would assist my interpretation of the test results” the responses were ($\frac{0}{SD}\frac{0}{D}\frac{2}{U}\frac{4}{A}\frac{3}{SA}$).

One comment aimed toward the probability graph was:

> *"Very nice way to visualize & easy to get the actual numbers (instead of just our 'gestalt' that post-test probability is low, or ...".*

One user made a design suggestion for the probability graph,

> *"would shade in area of graph where post-test probability tells the user to go to the CT (PIOPED 2) or MRI (PIOPED 3)."*

This may be interpreted as a request for explicit visual *decision boundary mapping* which we will come back to in the next chapter.

Table 8.1: Ratings of representations for understanding test results.

| **Query** | **Representation** | **SD** | **D** | **U** | **A** | **SA** |
|---|---|---|---|---|---|---|
| Is useful | natural frequency | 0 | 0 | 3 | 4 | 2 |
| | outcome table | 0 | 0 | 3 | 3 | 3 |
| | probability graph | 0 | 0 | 1 | 6 | 2 |
| Would assist interpretation of test results | natural frequency | 0 | 1 | 4 | 3 | 1 |
| | outcome table | 0 | 0 | 3 | 3 | 3 |
| | probability graph | 0 | 0 | 1 | 5 | 3 |

### 8.4.3 Sub-task 2: Inferring Candidate Diagnoses

Weighing the initial evidence to form a list of possible causes is the core of this sub-task. A potential set of requirements from the design implications are:

S2.1 Facilitate forming associative sets of candidate disease. *Rationale*: At least seven different online resources were utilized, with uses as frequent as hourly; *availability* heuristics (less than optimal cognitive associative processes [Kahneman et al., 1982]) may influence the process.

S2.2 Provide tools for the management of evidence resources. *Rationale*: Cognitive load will be high and so we can assist with any ad hoc foraging for information. The same rationale as R2.1 is also applicable.

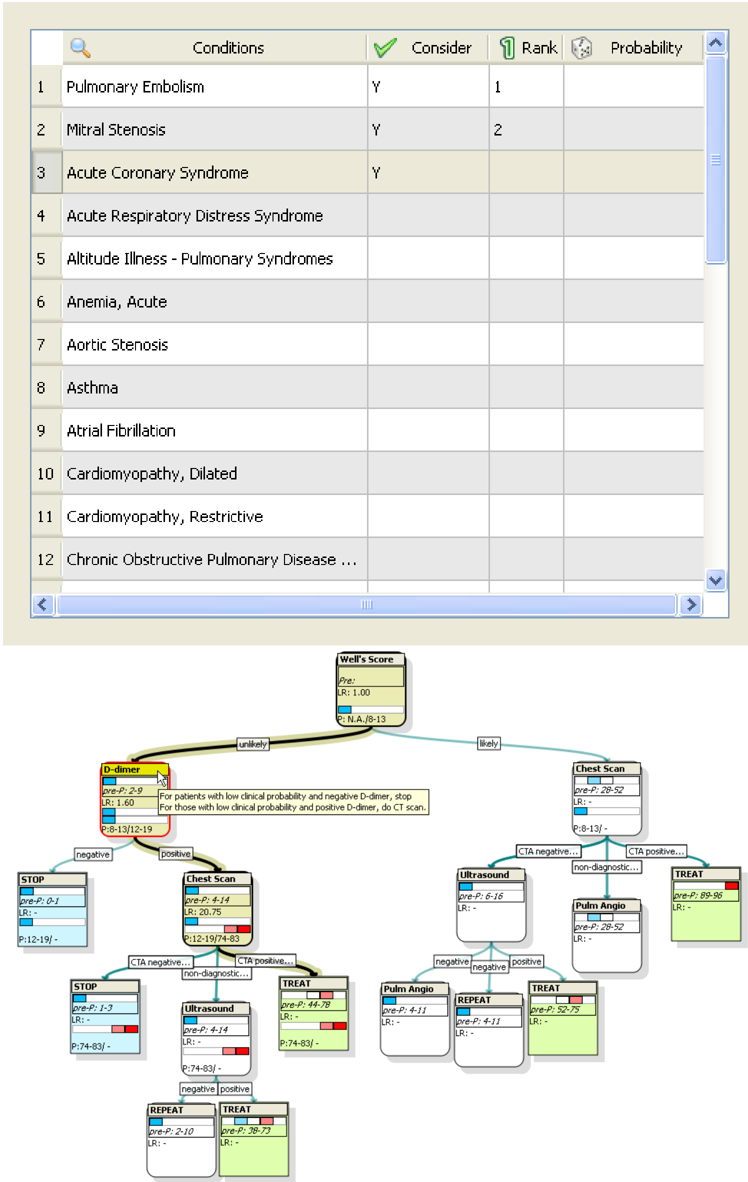
### Visualizations

Simply improved management of on-line resources could be of use for this task. Linking of symptoms to candidates with strengths could aid the formation of top consideration lists. When multiple candidates proceed to the diagnostic strategy stage managing the ranking of various hypotheses could also benefit from external support. Reducing cognitive load by offloading resource management for this task should aid the reduction of uncertainty by allowing more candidates and potential missing data to be considered. To support this we provided a table format display of possible conditions in the differential diagnosis shown in Figure 8.9. Any subset can be flagged in three categories of varying precision and then sorted based on these. Most of these considerations are applicable to the next sub-task of forming diagnostic strategies.

The list of differential diagnoses could each be linked to a diagnostic decision-tree. These could include strategies such as the tree for PE, or be references to recommended sources such as Black et al.'s [1999] "Diagnostic Strategies for Common Medical Problems". For our prototype implementation a diagnostic tree was only created for PE.

### Evaluation

Questions were based on the visualization shown in Figure 8.9 for differential diagnosis. To the question: this cognitive support for differential diagnosis would be useful, responses were: $\frac{0}{SD}\frac{0}{D}\frac{3}{U}\frac{5}{A}\frac{1}{SA}$. Figure 8.10 shows that probabilities were the least chosen form for ordering candidate conditions, which agrees with the observational study in which coarse granularity representations were only used for reporting probabilities. Comments included that this would be more geared toward medical students and included worries about the automation bias,



| | Conditions | Consider | Rank | Probability |
|---|---|---|---|---|
| 1 | Pulmonary Embolism | Y | 1 | |
| 2 | Mitral Stenosis | Y | 2 | |
| 3 | Acute Coronary Syndrome | Y | | |
| 4 | Acute Respiratory Distress Syndrome | | | |
| 5 | Altitude Illness - Pulmonary Syndromes | | | |
| 6 | Anemia, Acute | | | |
| 7 | Aortic Stenosis | | | |
| 8 | Asthma | | | |
| 9 | Atrial Fibrillation | | | |
| 10 | Cardiomyopathy, Dilated | | | |
| 11 | Cardiomyopathy, Restrictive | | | |
| 12 | Chronic Obstructive Pulmonary Disease ... | | | |



Figure 8.9: Top image: Table of possible candidate conditions in the differential diagnosis with multiple prioritizing options. Bottom image: Diagnostic tree related to single candidate condition of Pulmonary Embolism.

> *"... I'd be concerned that (esp. for junior users like residents) that this would replace clinical judgment e.g. if system doesn't say 'consider coronary syndrom' they will rule it out..."*

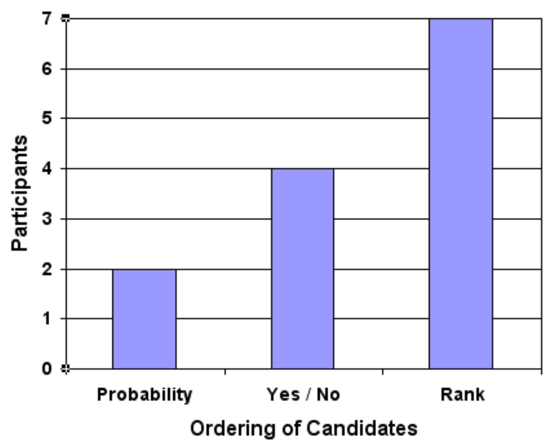



Figure 8.10: Responses to question: I would enter prioritizing information on conditions based on: (check all that apply).

### 8.4.4 Sub-task 3: Diagnostic Planning and Sub-task 4: Decision Making

The discussion of the results for the diagnostic planning and decision making sub-tasks has been combined to simplify the presentation, as both often relate to the same visualizations. The diagnostic planning sub-task involves prioritizing the potential diseases and forming an optimal ordering of tests to rule-in and rule-out candidates. One set of requirements from the design implications could be:

S3.1 Visualize the decision tree and provide direct access to evidence supporting it. *Rationale*: Weak confidence in system recommendations was reported.

S3.2 Visualize the post-test probabilities at various points in any decision tree. *Rationale*: The context of previous test results should be made obvious (see also S1.2).

S3.3 Provide access to repository for resources relevant to diagnosing condition. *Rationale*: Cognitive load will be high and ad hoc foraging for information may not provide all the relevant sources.

Requirements for the decision making sub-task may include:

S4.1 Allow the physician to seek and utilize support at any sub-task (i.e. provide visual evidence that the physician selectively uses for support). *Rationale*: Unwanted support may distract the physician and need to be circumnavigated.

S4.2 Provide easy access to evidence supporting any recommendations. *Rationale*: As evidence is continually changing any system needs to maintain appropriate confidence.

S4.3 Show how patient compares with patient profiles of those in the studies used in forming decision tree recommendations. *Rationale*: Allows the physician to weigh the applicability of recommendations.

**Visualizations**

A visualization of a decision tree for PE from our system is shown in Figure 8.11. The decision tree interface acts as a diagnostic flow chart. The tree represents a protocol derived from study evidence and the pre- and post-test probabilities displayed in the tree nodes indicate the diagnostic certainty at that point. The links represent decisions to move on to subsequent tests or diagnoses. The recommended decisions for any specified patient data and test results form a path through the tree that is illustrated by emphasized links. This tree can be considered visual cognitive support for the simulation heuristic [Kahneman et al., 1982, Klein, 1998], as the physician can directly see the Bayesian probability of a condition after future "simulated" tests.

The specification of a decision tree is contained an Extensible Markup Language (XML) file and so other conditions can be easily entered into the system. The user interface for any specified tree is created dynamically at run-time. This design allows for

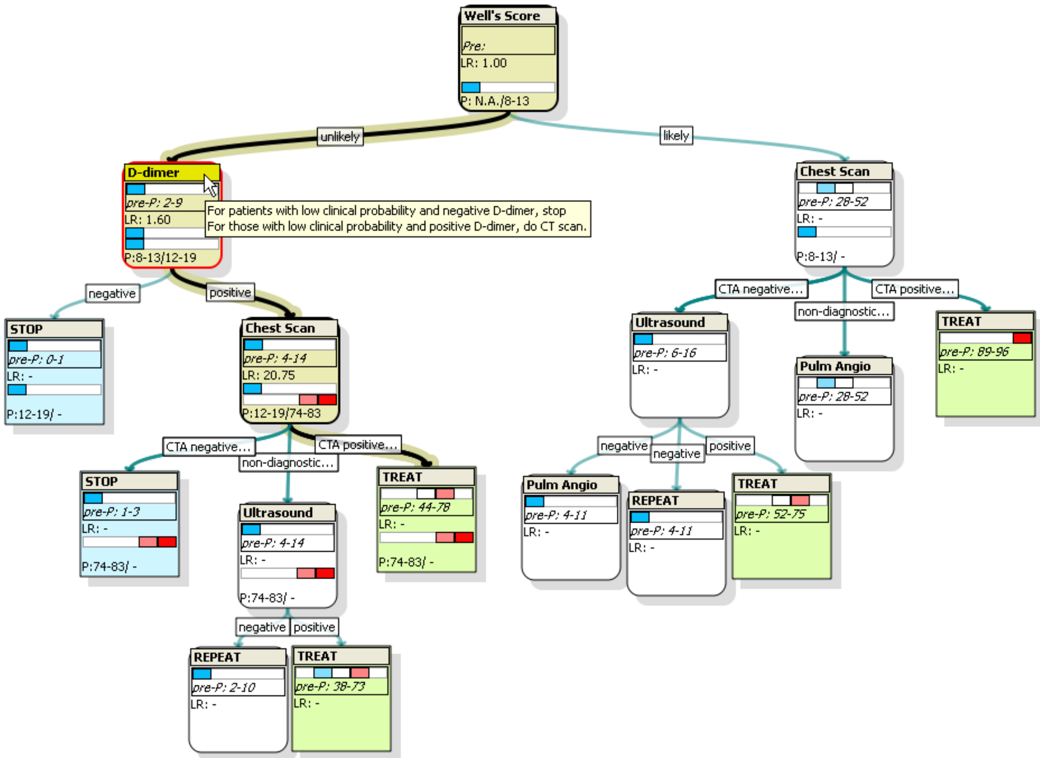



Figure 8.11: Decision tree visualization showing recommended decisions and pre- and post-test probabilities.

the continually evolving diagnostic strategies of evidence-based medicine as incremental revisions require no modification of the code, only changes in the XML file. Even within the duration of our studies we observed a change in the recommended strategy for PE diagnosis[1].

Figure 8.12 shows a single test node in the decision tree. Theoretical pretest probability is based on the base rate from a study [PIOPED investigators, 1990] and is shown in the top probability meter. A legend for reading the probability meter is shown in Figure 8.13. The

[1]The recommended test for moderate to high prior probability PE cases changed from V/Q scans to CT. It was even noted in our observational study that some participants disregarded the TDS system recommendation that was thought to be outdated.

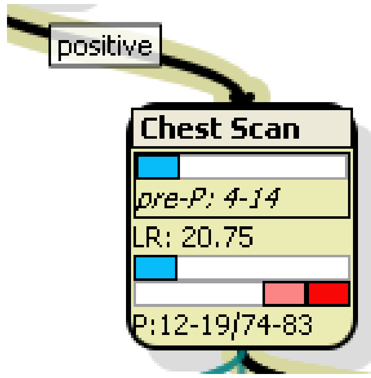



Figure 8.12: Decision tree test node. Top probability meter shows the pretest probability when reaching the specific node based on following recommended decisions and the base rate. Lower adjacent probability meters shows patient specific pre/post-test probability of PE based on selected test outcome (likelihood ratio 20.75, pretest probability is 12 to 19% and post-test 74 to 83%).

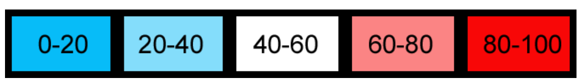



Figure 8.13: Legend indicating granularity of probability in percent for probability meter shown within test node in Figure 8.12. Any overlap of probability for each region is indicated by the fixed size colour bar.

divergent colour scheme emphasizes the two important extremes of ruling-in and ruling-out. As the base rate can be changed by the physician and tests ordered other than the recommended pathway the actual pre- and post-test probabilities for the patient are shown in the lower two probability meters. The meter's use of redundant spatial encoding (along with the colour saturation) allows the adjacent meters to accentuate large probability shifts from the test results. The likelihood ratio for the actual test outcome used to compute the post-test probability is displayed numerically. Selecting any test node by clicking on it with the mouse shows the corresponding test information in the detail view (shown in Figure 8.14).

We suggest that the diagnostic decision tree should be shown (visualized in some form) to the user at any point in which they are expected to be guided by, or conform to, its predetermined strategy. Visualizing the tree will reduce unnecessary uncertainty as to why the system makes suggestions, and increase confidence when following or disregarding recommendations. The tree should also be supported by drill-down information to justify the decision recommendations set down in the decision tree's branching structure. This should improve the lack-luster confidence users reported in the TDS system recommendations. The potential post-test probabilities are fundamental to choosing the most efficient diagnostic strategy; as if for all test outcomes the probability does not cross a decision threshold then the test may not even need to be ordered. Thus our visualization makes the potential post-test priorities for all possible pathways through the decision tree transparent to the user both with graphical probability meters and with text display (shown in Figure 8.11). The granularity of probability encoding using colour in the meter was chosen only slightly finer than three associations "low", "med", and "high" that were bound to initial decision tree branches in the existing documentation. This allows for some probability revision in the tree, and this vague encoding may have value for those who don't want to see the specific numbers. Regarding this granularity of encoding, Fox et al. [2001] has summarized this aspect of two of their earlier studies with medical diagnostic related problems and found that grossly reduced levels of probability encoding provided the same or even better user accuracy on performance of the task. This visualization also has the option to show test availability and duration uncertainties integrated within the tree.

Our visualization system was designed to be a tool used at any stage in the diagnostic process. Therefore we attempted to provide accessible visual evidence for the entire task. Rather than being forced to use the system whenever ordering a test, we envision the integration being user controlled. The support could seamlessly integrate in with the test ordering system, or be invoked from the test ordering system on demand. In this way Requirement S4.1 would be met, and users would not have to work around the system

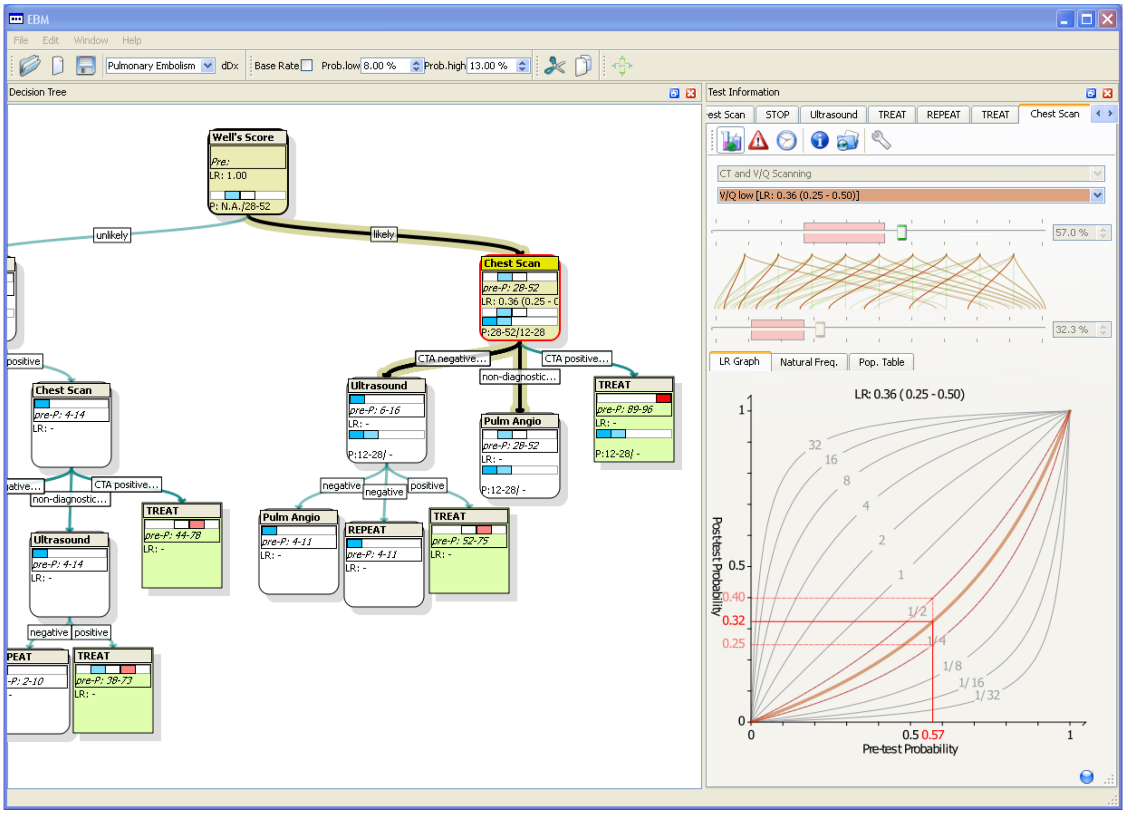


Figure 8.14: Screen shot showing one possible layout of the visualizations allowing contextual information to integrated.

support.

Our decision tree visualization as well as the natural frequency visualizations could potentially be used to aid communication with the patient as well. When appropriate patients might even have the option to utilize the system in order to better understand their diagnosis. Neufeld et al. [2008] describes using murals, animation, and interaction to more intuitively explain a node network of probabilistic relationships. However the visualization requirements of the patient were not explicitly considered at any point during system design, as we only targeted supporting the physician.

### Evaluation

Following the description of the candidate condition view the walk-through proceeded to Wells scoring assuming the physician was investigating a PE diagnosis. The visualizations shown in Figure 8.5 show the relevant aspects. For the question: Visual evidence of recommendation sensitivity to Wells scoring would be useful", responses were $\frac{0}{SD}\frac{0}{D}\frac{0}{U}\frac{4}{A}\frac{5}{SA}$. When asked to rate "The visual context of decision pathways would be useful when viewing other information (e.g. warnings, references, probability functions, ...)" the counts were: $\frac{0}{SD}\frac{0}{D}\frac{0}{U}\frac{7}{A}\frac{2}{SA}$.

Visualizing the recommended decision pathway along with the probabilities driving it was deemed important to provide confidence and transparency of the system rules. Details of this visual were shown in Figures 8.11, 8.12, and 8.13. Agreement to the statement "Pre/post-test probabilities integrated in the visual decision pathway would be useful" was $\frac{0}{SD}\frac{0}{D}\frac{0}{U}\frac{5}{A}\frac{4}{SA}$. Similarly strong responses of agreement were made to "Pre/post-test probabilities integrated in the visual decision pathway add confidence to system recommendations" ($\frac{0}{SD}\frac{0}{D}\frac{1}{U}\frac{4}{A}\frac{4}{SA}$).

As the evidence-base for diagnosis is constantly evolving, there was an integrated component relating to references for the encoded strategies and statistics (likelihood ratios, base rates, etc.). One simple visualization relating to this was a folder browser of reference material tied to any test node as shown in Figure 8.15.

For the question, "Integrated references would increase confidence in hospital decision recommendations", ratings were quite positive $\frac{0}{SD}\frac{0}{D}\frac{1}{U}\frac{5}{A}\frac{3}{SA}$. Similarly there was general agreement with the statement "This would assist me in ensuring I have read the latest evidence" ($\frac{0}{SD}\frac{1}{D}\frac{1}{U}\frac{5}{A}\frac{2}{SA}$). Some skepticism was present in one comment as, "only if you keep the evidence updated in real time (not realistic). There are always new meetings, new literature etc.".

The visualization which raised the most discussion and in general received the most lukewarm responses related to temporal uncertainty. A summary of the two forms for pro-

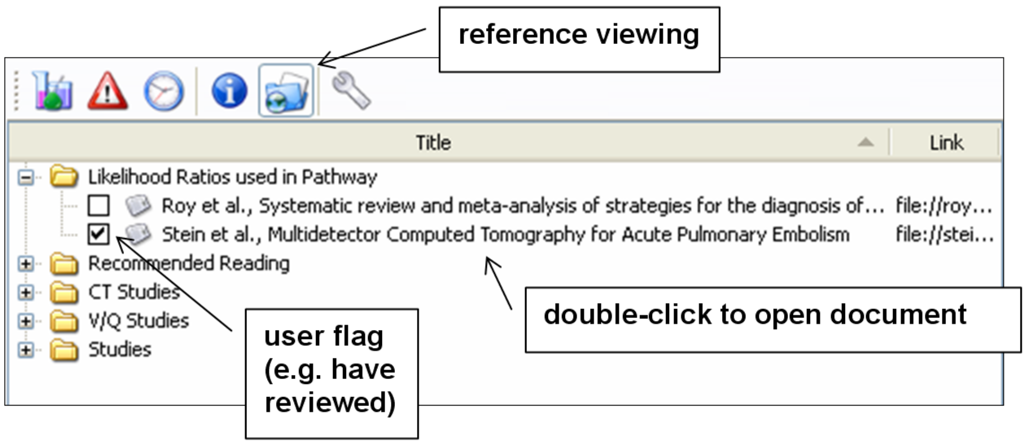



Figure 8.15: Related reference repository and search links.

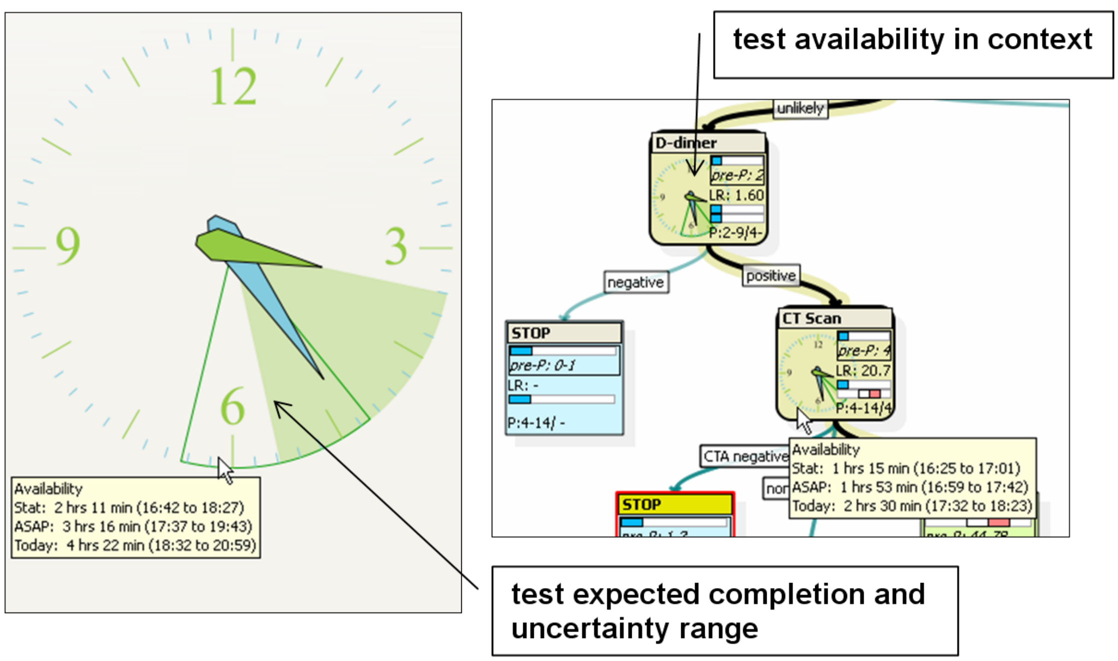



Figure 8.16: Related reference repository and search links.

viding test availability and uncertainty are shown in Figure 8.16. The discussion focused on the potential for availability awareness to lead to the use of sub-optimal tests, and thus a reduction in quality of care for the sake of more timely knowledge. Comments included

> *"a lot of competing factors need to be considered"*,

> *"my experience has been that our tests here are done quite promptly, I don't think this is useful,"*

and

> *"I'm not sure this would be clinically relevant or even desirable."*

The question "This test availability and uncertainty visualization would be useful in planning" got a very mixed rating: $\frac{0}{SD}\frac{2}{D}\frac{4}{U}\frac{2}{A}\frac{1}{SA}$.

## 8.5 General Evaluation

After the walk-through the participants completed questionnaire portions related to their personal perspectives and the system they were shown. When asked if they used computer or visual aids in making evidence-based decisions, five participants reported rarely and four sometimes (other options were never, once or twice, and always). Thus current practice appears for the majority of the time based to be based on an internal cognitive process.

This status quo leaves room for exploring additional support as shown from the responses in Table 8.2. Responses to Question P1 show interest in decision support, while the somewhat contradictory Question P2 shows (likely prudent) skepticism in this type of support. There was general agreement to the idea of providing evidence without explicit decision overriding although this was a somewhat vague question (P3). Question P4 and P5 also indicate that there is a niche to explore in increasing support for this process. It is

Table 8.2: Questions related to personal preferences and introspection.

| | Question | SD | D | U | A | SA |
|---|---|---|---|---|---|---|
| P1 | I would like more computer support for evidence-based decision making | 0 | 0 | 0 | 4 | 5 |
| P2 | I don't feel confident in computer systems advising me on decisions | 0 | 5 | 3 | 0 | 1 |
| P3 | I would like clear visual evidence and be free to make my own decisions | 0 | 1 | 1 | 5 | 2 |
| P4 | I am happy with how I manage uncertainty in making diagnostic decisions | 1 | 1 | 3 | 4 | 0 |
| P5 | I am confident with my ability to apply evidence in my decision making | 0 | 1 | 0 | 8 | 0 |

important to remember that all these participants were experienced practitioners (median experience range was 6-10 years) and so they should have had time to adaptively refine their process of EBM.

The questionnaire and responses related to an overall judgment of the demonstrated system are provided in Table 8.3. That the system would be "useful in practice", "increase my confidence", and "would use ... if it was available" all received clear agreement. The strongest agreement was given for using the system for education (Question O1).

Overall general comments from the participants were positive such as "very useful overall" with one of the strongest being,

> *"This has such terrific possibilities. This tool should be mandatory for groups that publish and distribute guidelines for therapies".*

Concern was even raised about the implications of not using the tool,

> *"It may increase the number of litigations if someone did not happen to use this tool and was wrong about decision to treat or not treat".*

We acknowledge the limited applicability of the feedback based on such a preliminary

Table 8.3: Questions related to the overall system impression†.

| | Question | SD | D | U | A | SA |
|---|---|---|---|---|---|---|
| O1 | Visual evidence like this would be useful for education and training | 0 | 0 | 0 | 3 | 5 |
| O2 | Visual evidence like this would be useful in practice | 0 | 0 | 1 | 5 | 2 |
| O3 | Visualizations of uncertainty like this set would increase my confidence in decisions | 0 | 0 | 0 | 6 | 2 |
| O4 | I would use visual evidence like this set of visualizations if it was available | 0 | 0 | 0 | 5 | 3 |

† Only 8 of the 9 participants completed these questions.

evaluation with limited practical testing. However we would say that the visual support we created appeared to address some of the issues we noted during our initial observational study.

### 8.5.1 Hardware/Availability Preferences

Regarding the participants' preference for the hardware/availability for decision support, there was a general leaning toward portability or availability as can be seen in Figure 8.17. Availability was only implicit from the generic descriptions: hand-held, any shared computer, and high end computer (which was listed as having multiple displays). Both hand-helds and shared computer terminals are quite ubiquitous in their current environment. This relates to one comment on the specific visual support availability that

> *"would be helpful if quickly accessible at point of care."*

### 8.5.2 Further Evaluation

A next step will be to go through a more formal assessment of the visualizations effects on clinical decision making. This could potentially be done in an educational setting to

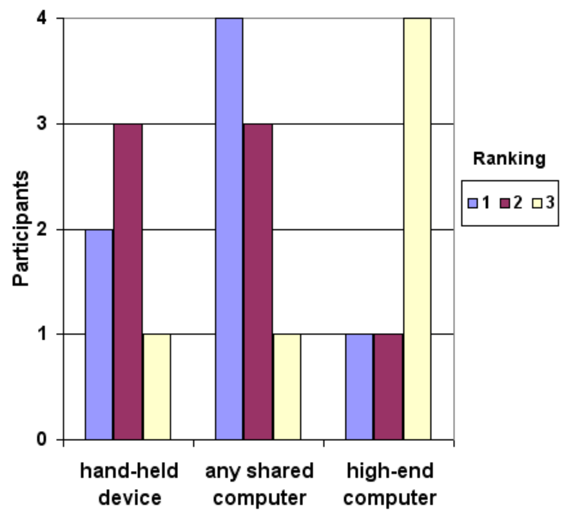



Figure 8.17: Preference as to hardware/availability for decision support.

additionally assess its value in aiding the learning of evidence-based medicine. Clinically such a tool will only be valuable (as an applied tool), if it can be shown to positively influence the diagnostic process to the point of reducing

1. diagnostic errors, and
2. adverse patients outcomes that relate to the diagnostic errors.

An uncertainty visualization may actually create new uncertainty for the physician to accommodate [Timmermans and Angell, 2001]. Therefore careful evaluation will be required to show that this additional information is providing the assistance that was intended.

### 8.5.3 Heuristic Evaluation

In a small digression, this section provides the heuristic evaluation of the medical visualization system based on the heuristics presented in Table 3.1, in the same manner as the two other domain visualizations were evaluated. The heuristics and a description of their applicability are summarized:

- **Ensure visual variable has sufficient length** – The coarse probability meter encoding used colour and value to encode five levels of uncertainty. The probability slider encodes uncertainty using the width of the pink coloured bars. Both these encodings provide sufficient length for their intended purpose.
- **Preserve data to graphic dimensionality** – For the probabilities that were visualized we would argue that any representational encoding such as width, colour, or text, do not violate this heuristic. For this type of probabilistic data, or any other dimensionless data, using regions changing in two dimensions, or volumes, could be considered a violation of this heuristic.
- **Put the most data in the least space** – The tree visualization utilizes the most space, but the space is appropriate to provide a spatial representation that may allow the abstract decision strategy to be more easily interpreted. Adding more details or simplifying this display to make it more compact may be one avenue for further refinement.
- **Provide multiple levels of detail** – This heuristic is addressed in the overview of the diagnostic strategy that is provided by the decision tree visualization, while more detailed specifics are provided in linked views. At a higher level of detail the table view of candidate conditions allows ranking of potential diagnoses.
- **Remove the extraneous (ink)** – A clean and unadorned style was chosen to depict the decision tree nodes, probability meters, and likelihood ratio graphs.
- **Consider Gestalt Laws** – In this visualization the physician is not required to look for visual patterns or read complicated encodings, therefore this heuristic may apply only to a lesser extent. Nevertheless, Gestalt readings offer a reason why a tree is a useful representation based on its exploitation of proximity and connectedness for

displaying relationships.

- **Integrate text wherever relevant –** Text is integrated into many of the displays and mouse-triggered hover queries provide additional text detail for graphic elements.
- **Don't expect a reading order from colour** – Only two base colours combined with value changes are utilized for the probability meter. While this reading order is simple enough that it should not confound the user, the positional encoding of probability increasing left to right will assist interpretation.
- **Colour perception varies with size of coloured item** – The use of colour in the design is restricted so that colour reading of varied size components is not required. This eliminates the potential misreading that this heuristic warns of.
- **Local contrast affects colour & gray perception** – The visualizations do not require decoding quantities from colour or gray levels and thus contrast effects should not have any significant impact. The probability meters are set in nodes having fixed background colours with good contrast, and with the aligned pre/post-test probability meters shown in Figure 8.12, only the same colours are vertically adjacent.
- **Consider people with colour blindness** – Value variation is combined with the colour encoding to assist those with colour deficiencies.
- **Preattentive benefits increase with field of view –** Test recommendations (decision pathways) in the tree view are encoded with both size and value so that changes may be more easily perceived.
- **Quantitative assessment requires position or size variation –** Quantitative details are provided based on positions in the graph representations, or directly in text details.

While this visualization was clearly different and more abstract than the previous two domains, the vast majority of the heuristics were still relevant. This adds further evidence that these heuristics may work well in other domains, in that they can provide a checklist for evaluating or shaping design.

## 8.6 Conclusions

Based on our observational study of the diagnosis of pulmonary embolism in an evidence-based medicine framework we created a task model. This model enabled us to systematically review the uncertainties involved and structure our analysis and development, finding both external and internal uncertainty to be prevalent throughout this task. Cognitive theory was used to suggest areas in which internal uncertainty may be problematic, most notably from heuristics and biases that may affect probability estimates based on clinical judgment.

Using our analysis of the uncertainty, we created visualizations to provide the potential for cognitive support in each sub-task of the diagnostic process. While external uncertainty has commonly been the focus of other investigations we have shown that internal uncertainty should be considered and have provided visual support for this uncertainty relating to evidence-based diagnosis. We also hope the exposing of areas of uncertainty can inform the development of future support tools for this task. How best to provide visualization support for diagnostic decisions is a large area for future research. After developing these visualizations further, the medical endpoint would be to evaluate their impact on clinical decision making, with the final measure being patient outcomes.

### 8.6.1 Generalizing Across Domains

Decisions involving uncertainty visualizations may be complicated by translation errors from vague but accurate internal representations to precise but inaccurate numerical ones. Any support to offload these types of problems must be very explicit in the translation or one risks confounding the task rather than helping it. By providing both coarse and precise representations of probability we hope that users may choose an appropriate one that they are comfortable translating between.

## 8.7 Acknowledgements

We would like to thank all the participants in our study and the Ward of the 21st Century for providing logistical support. Thanks also to Catherine Plaisant and Tim Zuk for providing feedback on prototype visualizations.

# Chapter 9

# Framework for Supporting Uncertainty Visualization

Creativity requires the courage to let go of certainties.
– Erich Fromm (1900 – 1980)

This chapter moves toward integrating and abstracting from the previous bottom-up domain investigations and the earlier top-down theory driven chapters. It provides both a categorization of cognitive uncertainty and a light-weight and readily applicable framework motivated by reducing the complexity of the cognitive tasks dealing with uncertainty. The framework provides seven directives that relate to design, evaluation, as well as the categorization of cognitive uncertainty. To illustrate the applicability of this framework, I apply each component of it to the domain specific uncertainty visualizations developed in Chapters 5 through 8.

## 9.1 Introduction

While visualizing both the data and its associated uncertainty has been accepted as beneficial for accurate interpretation, the integration of uncertainty information into an existing or new visualization is not standard practice. The practical tasks of maintaining ease of comprehension for both the data and the uncertainty are not straight forward. Hence, in building uncertainty visualizations there still exist many challenges, such as finding good representation of errors and uncertainty for 3D visualizations [Johnson, 2004], and understanding how knowledge of information uncertainty influences analysis [MacEachren et al., 2005]. As a result, even choosing an initial design may be difficult.

Frameworks are important as through utilizing continual theoretical and technological improvements the number of potential visualizations that are feasible is constantly growing; thus it is increasingly difficult to understand how to make trade-offs between them. For this reason I introduce a light-weight framework to inform the development of uncertainty visualizations. The framework describes aspects relating to the role of uncertainty in decision tasks which may provide guidance in the design of new visualizations. I illustrate the utility of this framework by describing how it applies to the three visualizations developed in the preceding chapters.

A brief review of some previous frameworks related to uncertainty visualization will be given. Often frameworks that relate to design can be used for evaluation, such as Amar and Stasko's [2005] "Knowledge precepts for design and evaluation of information visualizations". Thus evaluation should always be kept in mind as one potential use, as was discussed in Chapter 3.

## 9.2 Motivation

As task level taxonomies can be useful for the design and evaluation of visualizations [Valiati et al., 2006] looking deeper into visualization tasks at the cognitive level may also provide value. Decision making can be confounded by uncertainty, and so may deserve special attention. Decisions are often the end product of the reasoning process and therefore pushing their requirements to the forefront may help inform design. Cognitive studies have shown numerous potential weaknesses in the reasoning process when dealing with uncertainty [Gilovich et al., 2003, Kahneman et al., 1982] and so explicit considerations for this aspect may guide uncertainty visualization design [MacEachren et al., 2005, Zuk and Carpendale, 2007].

MacEachren et al. [2005] has described seven challenges for the visualization of uncertainty:

1. understanding the components of uncertainty and their relationships to domains, users, and information needs,

2. understanding how knowledge of information uncertainty influences information analysis, decision making, and decision outcomes,
3. understanding how (or whether) uncertainty visualization aids exploratory analysis,
4. developing methods for capturing and encoding analysts' or decision makers' uncertainty,
5. developing representation methods for depicting multiple kinds of uncertainty,
6. developing methods and tools for interacting with uncertainty depictions, and
7. assessing the usability and utility of uncertainty capture, representation, and interaction methods and tools.

My work described in Chapter 3 on evaluation relates to Challenge 7, in providing new forms for the assessment of uncertainty visualizations. Progress was made toward challenges 5 and 6 in the development of uncertainty visualizations provided in the three case study domains: archaeological, geophysical, and medical. These specific domains also provided insights into Challenges 1 and 2, and to a lesser degree Challenge 3. Challenge 4 is supported by the work provided in the chapter on Visualization of Uncertainty in Reasoning, but we will now turn to address this challenge in more detail. However this framework relates to most if not all of the challenges in some form.

## 9.3 Related Work

It is difficult to make generalizations as inductive processes are usually less than certain, but we remain motivated to develop knowledge of what usually, or even often, works. As noted in Chapter 2, MacEachren [1992] identified visualizing accuracy, and visualizing precision as separate tasks requiring different strategies, which begs the question what are the strategies that people use or should use. Similarly MacEachren proposed the use of colour saturation and blurring as being conducive to indicate uncertainty. While these may intuitively be more natural encodings their general superiority to other encodings remains to be proven, and as trade-offs are likely user and task dependent, this may never be proven in the general sense.

Numerous uncertainty visualizations have been proposed for different domains, data, and types of uncertainty [e.g. Grigoryan and Rheingans, 2004, Love et al., 2005, Masalonis et al., 2004, Pang et al., 1997, Rheingans and Joshi, 1999], but we still require improved understanding of what makes a good uncertainty visualization [MacEachren et al., 2005]. In the field of GIS, frameworks have been proposed to guide displaying error and uncertainty, such as Beard and Buttenfield's [1999] suggestions for mapping error analysis methods to graphical display; however, the extent to which these frameworks will generalize beyond the GIS domain is not clear. Exposing uncertainty and showing the possible effect of this uncertainty on outcomes is one of Amar and Stasko's [2005] design and evaluation precepts for information visualization. For design they stated their knowledge precepts could be used to [Amar and Stasko, 2005]:

1. generate new sub-tasks for a visualization to support or perform,
2. identify possible shortcomings in representation or data, and
3. discover possible relationships to highlight or use as the basis for a visualization.

Similarly they state the precepts could be used for a form of heuristic evaluation. However in trying to use them for evaluation these high-level goal based heuristics may be more difficult to apply than the traditional ones from usability evaluation [Zuk et al., 2006], as described in more detail in Chapter 3. In order to provide further insight and practical advice on creating uncertainty visualizations I will provide a framework, to be detailed later in Section 9.5. This framework will relate to more specific details pertaining to uncertainty visualization and is grounded more in the practical concerns found in the domain investigations of Chapters 5 to 8. Thus being lower-level than Amar and Stasko [2005] it may be more easily applicable to heuristic evaluation. On the restricted scope of uncertainty visualization it should assist in the same goals of: generating new sub-tasks, identifying shortcomings in representation or data, and describing relationships worth highlighting in a visualization. First, however, I will introduce a categorization to assist in relating the uncertainty to cognitive issues.

## 9.4 Categorization of Cognitive Uncertainty in Decision Making

Uncertainty visualizations may be utilized cognitively in a myriad of ways, but we will focus here on the aspect of decision making. As decisions are an end product of the reasoning process, I have created a categorization based on potential partitioning of decision space. The categories to be presented can be used to split the visualizations based around the decisions made using them;

1. those for which no certainty threshold can be mapped to a decision outcome,
2. those that are used for a decision based on a single threshold of certainty, and
3. those that are used with multiple thresholds or as a continuous weighing of certainty (function).

Decisions which may use a threshold (the second two types) are likely a simpler cognitive task and this forms the dichotomy of the single and multiple characterizing threshold categories to be described further. These second two categories along with example thresholds and tasks are summarized in Table 9.1. Decisions where no boundary between outcomes can be defined (the first type) are cases where no explicit thresholds are formed and may be similar to the infinite number of thresholds we included in the multiple threshold category. However, we will leave this class out of the current consideration, except for the fact that decisions of that type may benefit by decomposing aspects of them into ones that use thresholds. Similarly vague or fuzzy thresholds can also be considered as thresholds around a region of ambiguity.

One example of a single threshold may be a predetermined cut-off such as a 95% confidence interval, while a Gaussian probability distribution function (PDF) may be an

Table 9.1: Dichotomy of uncertainty thresholds in decision making.

| **characterizing thresholds** | | **example thresholds** | **example tasks** |
|---|---|---|---|
| single | single definable case | best-case, worst-case | risk scenarios, devil's advocate |
| | single decision dividing range | accept threshold, reject threshold | quality control (QC), hypothesis testing |
| multiple | definable number of ranges, overlapping ranges | levels of confidence | ordering, naming |
| | continuous function, ∞ | probability distribution function (PDF), gestalt | weighing, Bayesian reasoning |

example of a continuous function for weighing confidence. To clarify the distinction between single and multiple characterizing thresholds, it is not the discrete or continuous data extent that Pang et al. [1997] found useful for one aspect of their categorization of uncertainty visualizations, but rather a person's certainty to decision mapping being Boolean or non-Boolean in nature.

The single threshold may often be pre-determined to choose between two actions. A good example of this is the visualization for the Mariners 1-2-3 Rule [Holweg, 2000]. This rule is a tropical cyclone path forecast with 100-200-300 nautical mile margins of error added at 24-48-72 hours respectively, shown in Figure 9.1 and explained in Figure 9.2. Its Boolean value encoding of a single decision threshold is based on a data and uncertainty threshold (sustained wind speed and probability) and allows for easy interpretation of the area to avoid.

The second threshold category is more complex in that the person is required to consider and read multiple uncertainty levels from the visualization. Graphic variables that can encode more than one bit such as the plane, size, and value, will be more appropriate for encoding the thresholds. However, anything with more than one bit of capacity can be considered in this category, as for some tasks low, medium, and high levels may be

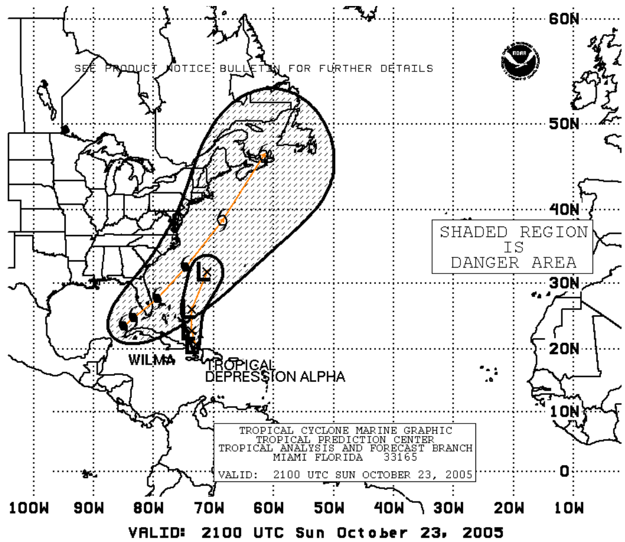



Figure 9.1: Mariners 1-2-3 Rule Chart at one point during hurricane Wilma. Image courtesy of the National Oceanic and Atmospheric Administration National Weather Service.

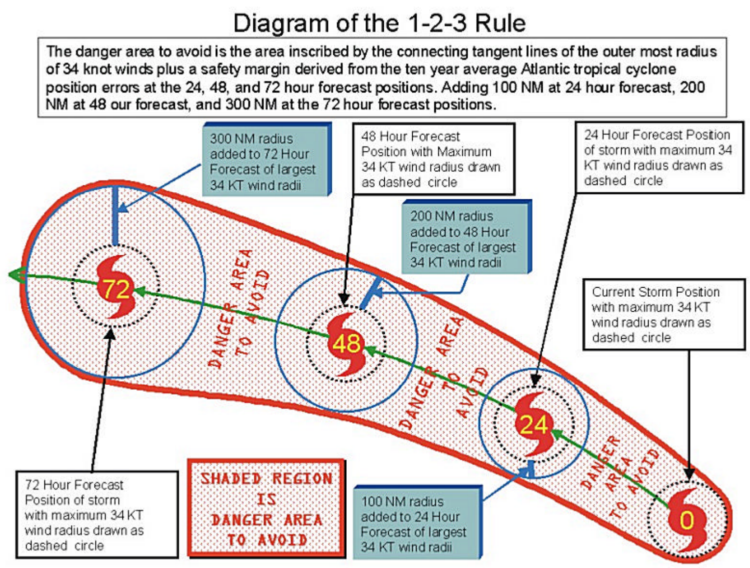



Figure 9.2: Diagram explaining the construction of the Mariners 1-2-3 Rule graphic. Image courtesy of the National Oceanic and Atmospheric Administration National Weather Service.

sufficient. When the number of ordinal or nominal categories gets too large one should provide extraction for any details with on demand-like mechanisms. The use of more representational levels is of course appropriate to more naturally match a continuous variable, or when relationships are important (i.e. relative values) and not magnitudes.

When the number of thresholds is low or in the single threshold category one should consider providing a direct visualization of the decision regions in what I will term *decision boundary mapping*. This is providing the user the flexibility to select an encoding or representation that moves from reading certainty to reading thresholds or decision boundaries, and thus visualize outcomes. Uncertainty visualizations may facilitate this by allowing user definable functions that map data and uncertainty to new derived decision related attributes that can be displayed as either discrete outcomes, or a continuous function to be thresholded. This is related to the suggestion for integrated criterion and decision spaces in spatial decision making by Jankowski et al. [2001]. This can allow one's personal decision boundaries to be transparent to others. Exploring these functions and thresholds should be interactive as often they can not be determined a priori. This notion may be useful for decisions that incorporate more than components of certainty as the weighing process may involve many factors that can not be intuitively combined (e.g. the Mariners 1-2-3 Rule).

## 9.5 Directives for Supporting the Visualization of Uncertainty

In presenting advice in the form of seven considerations and recommendations, which I will term directives[†], I try to bring the focus to specific aspects of the reasoning process. The framework I propose can be used to inform the design of a visualization, but is potentially also applicable for heuristic evaluation of uncertainty visualizations [Zuk et al.,

[†]The term directive was chosen over precept to provide equal emphasis of their potential use after development in heuristic evaluation or design review. They are not called the Prime Directives both to avoid a pun and any allusion to Star Trek that would be lost in generational or lingual translation.

2006], and therefore ties into Challenge 7. As visualizations are often developed in an iterative manner this dual capacity is especially useful as it can help both to judge what has already been done, as well as suggesting possible new avenues of design.

The directives were developed through a process similar to that of thematic analysis [Boyatzis, 1998], in that they are based on inductive aspects from the issues and solutions found in the domain investigations (Chapters 5 to 8), as well as from the gathered theory discussed in the work on uncertainty in reasoning (Chapter 4), evaluation (Chapter 3), and existing theory from the literature (Chapter 2). While this is a generalized usage of the term thematic analysis, it is nonetheless descriptive of the process. Boyatzis [1998] provides four stages in learning thematic analysis:

1. sensing themes (recognizing the codable moment),
2. doing it reliably (recognizing the codable moment and consistently encoding it),
3. developing codes, and
4. interpreting the information and themes in the context of a theory or contextual framework (contributing to the development of knowledge),

and the directives may be considered an artifact of the last stage of the learning process, as coding of themes can occur at various levels of abstraction [Charmaz, 2006]. The themes were generated from the material summarized in the earlier chapters and numerous discussions with the domain experts, and thus the directives can be thought of as themes. The directives are listed in Table 9.2 and will be described in detail in the section that follows. We do not suggest these seven provide an exhaustive basis for design or evaluation but consider them all to be valid points worth reviewing or applying to a problem or design. We will now describe the seven directives that relate to both the categorization of cognitive uncertainty and uncertainty visualization in general.

Table 9.2: Directives to support uncertainty visualization.

| | **Directive** |
|---|---|
| 1 | Provide support for cognitive task simplification. |
| 2 | Support emphasis and de-emphasis of uncertainty information. |
| 3 | Support viewing of uncertainty as metadata and separately as data. |
| 4 | Allow the user to select realizations of interest. |
| 5 | Mitigate cognitive heuristics and biases with reasoning support. |
| 6 | Provide interaction to assist knowledge creation. |
| 7 | Assess the implications of incorrectly interpreting the uncertainty. |

### 9.5.1 Provide support for cognitive task simplification

Simplification is important to allow the reduction of overwhelming information or improve efficiency. In vector field related examples, Wong et al. [2000] provide a simplification of 3D vector fields based on the aspect of interest (vorticity thresholds in their example) and Telea and van Wijk [1999] provide summary glyphs based on similarity clustering. Simplification of decision spaces (options) with multiple criterion has been suggested by Jankowski et al. [2001] via grouping options based on the Pareto-dominance principle[1]. Uncertainty has the potential to add complexity and so simplifying the related tasks may be important.

One design recommendation tied to simplification is to allow the user to reduce the ratio of the number of graphic encoding levels to decision thresholds (and the associated options or actions) to 1:1. In other words use only the number of encoding levels necessary, and when useful allow the user to reduce the levels down to provide a single threshold, thus moving the cognitive task into the first type of our decision threshold categorization. This design recommendation parallels the reduction used in "focusing", where subsets of data are interactively highlighted in order to provide a customized and simplified reading of a

[1]Pareto-dominance can be used to form a group of non-dominated, and likely preferred options. Non-dominated options can not be surpassed by other options on any evaluation criterion without reducing a different criterion [Jankowski et al., 2001].

graphic [MacEachren et al., 1998b]. The single threshold category does not necessitate only two levels of encoding for the decision boundary, it may suffice that the user can easily perceive the threshold boundary (e.g. one of the levels).

As noted earlier, a single decision threshold may be based on a more complex function of multiple types of uncertainty (e.g. precision, confidence) but provides a simpler (and possibly spatial) delineation of only two decision options (regions). Spatial encoding is natural for a set or region and is supported by gestalt of connectedness and proximity. Of course when tasks and decisions can not be clearly defined one should not reduce the representational precision to preclude potential tasks. A superior solution is allowing the mapping function and number of encoding levels (thresholds) to be controllable. This is often available through such interactions as selection and editing of colourmaps thereby enabling user customized visual queries.

The single threshold scheme allows for reduced cognitive load on the user. As the user is likely still considering the data in addition to the uncertainty meta-data, any reduction in load may be greatly appreciated. It is important to remember the extra attribute of uncertainty brings along its own context, along with the original data and its surrounding context. This leads to the design recommendation that when reducing the number of levels for a specific decision-based encoding, additional levels can still be available in a complementary or redundant representation. For colourmaps an example would be overlaid iso-value contours while the colourmap utilizes a single colour beyond the threshold.

### 9.5.2 Support emphasis and de-emphasis of uncertainty information

For decision making characterized by the single threshold, either side of the threshold may be important. For the cognitive task of rejecting data from consideration, high uncertainty may be the criteria, and so high uncertainty should draw attention, vis à vis for accepting tasks the high uncertainty should be encoded to reduce attention. Providing both variations will greatly assist the user if the task can not be determined a priori.

When a user needs to review data to decide what needs further investigation they may internally compute an uncertainty cut-off for what is warranted by this task. To facilitate this the visualization could interactively filter or visually distinguish the data with uncertainty beyond the threshold. However for this purpose it may sometimes even be superior to provide no uncertainty visualization as its primary value is only in data selection. In this case interactivity would be strongly advised and would implicitly reveal the derivative of the uncertainty via the changes in selection.

### 9.5.3 Support viewing of uncertainty as metadata and separately as data

While uncertainty is often regarded as metadata it can also be considered as data in its own right. This simplified mindset may be used to decouple the relationship between data and its uncertainty to allow one to assess the uncertainty on its own. This is useful to allow quality control inspections of the uncertainty (and its derivation) or provide for forming a mental image of the uncertainty separate from the data to provide reference for interpreting any merged visualization of data and uncertainty. The integration of uncertainty may overwhelm the reading of other important aspects of the data and so the option to see it separate, or minimally encoded with a perceptually separable variable, versus an integral one[2], can be of assistance [MacEachren et al., 1998a]. This contrasts the holistic goal of Wittenbrink et al.'s [1996] *verity* visualization criterion (that the uncertainty is holistically integrated without "overloading"), but providing both allows the user to reduce the chance of being overwhelmed if only viewing either the data alone or combined with uncertainty.

When the uncertainty is considered as data in its own right then standard task or operation categorizations may be useful to consider (e.g. locate, identify, distinguish, cluster, ... [Wehrend and Lewis, 1990]). Thus data with uncertainty can be considered both as two data sets standing alone, as well as a multi-variate dataset (uncertainty as metadata). Howard and MacEachren [1996] have suggested for geographic reliability (uncertainty)

[2]See Ware's [2004] textbook for a review of separable and integral variables.

visualization the option to separate these three aspects and also provide some spatial and general operational tasks related to the three. A taxonomy of tasks relating to multi-dimensional visualization has been provided by Valiati et al. [2006], and is appropriate as uncertainty can be treated as additional dimensions.

### 9.5.4 Allow the user to select realizations of interest

Realizations were defined earlier as specific potential outcomes from a set of probabilistic outcomes. With many types of uncertainty one needs to selectively consider potential realizations. Therefore one question to ask when creating an uncertainty visualization is: "Does the visualization provide for the explicit or implicit reading of possible realizations?" This directive of our framework focuses on allowing illustrations or filtering of specific realizations by the user. Implicit reading would entail directly showing some form of the boundary of possibilities within which the user must entertain possibilities themselves, and explicit would be directly showing a probability distribution function, or animation over possible realizations. This is directly related to Amar and Stasko's [2005] precept to show the effect of uncertainty on possible outcomes, but turns it around to advise letting the user select outcomes to see how certain they are.

With some representations this will most easily be implemented with animation over possible realizations, but it can also be provided as user driven queries of possible realizations. The animation concept has been described by Ehlschlaeger et al. [1997] for use in understanding possible realizations and uncertainty of a surface model. They describe some problems and solutions to allow smoothly interpolating between "key frames" created from specific stochastic realizations. With the query of possible realizations, the result could be either Boolean or a level of certainty based on the size of the set of realizations that satisfies the query. Returning to the surface model example, one could use a probe to query a specific surface height at one grid location and for the query result all stochastic surface realizations within a tolerance of that height could be shown together using transparency (local patches or the entire surfaces). Some novice chess players utilize the benefits of this type of realization "query" as they may temporarily move a chess piece to a location (without letting go of the piece and therefore committing) to simplify visualizing further possibilities.

### 9.5.5 Mitigate cognitive heuristics and biases with reasoning support

This directive relates to any form of graphical support that assists cognitive heuristics or allows offloading of cognitive computation relating to uncertainty. When people must incorporate uncertainty when making judgments there is a good chance that the cognitive tasks involved will use heuristics as opposed to formal logic or other algorithms [Kahneman et al., 1982, Klein, 1998, Heuer, 1999]. These heuristics have the potential to be prone to bias and other weaknesses [Griffin and Tversky, 2003, Tufte, 2006, Turpin and du Plooy, 2004]. It is therefore a valuable option to provide for extraction of the uncertainty data if the user wishes to proceed with some type of non heuristic assessment. Extraction, as well as making available the numerical details on demand, is part of the *"Visual Information-Seeking Mantra"* proposed by Shneiderman [1996].

Klein [1998] argues many real-world decision processes fit his recognition-primed decision model in which experience allows recognition of the best action without the need for a direct comparisons of options. In this model uncertainty visualization may assist in spotting anomalies which need further clarification, or assist in the evaluation process during mental simulation. This model can be considered a heuristic and Klein [1998] has described how redesigning a system interface based on decision requirements improved task performance. However, one should even consider the possibility of providing additional visualizations directly for the reasoning process itself [Zuk and Carpendale, 2007].

### 9.5.6 Provide interaction to assist knowledge creation

The previous factors and questions all elucidate multiple design options. Therefore interactivity is a key consideration that can support personal and task customization. This interactive control over the visualization is paramount to enable the user to explore and comprehend the data on their own terms. Howard and MacEachren [1996] discuss the design of interfaces for interaction with uncertainty visualization in GIS and found it useful to analyze the interface on the conceptual, operational, and implementation levels. At the implementation level, Zuk and Carpendale [2005] have shown that even when exploiting GPU programming for faster computation one does not necessarily give up interactive flexibility in simulation visualization, and the reduced computation times may even create further options for interactivity.

Applying Ware's [2004] visually aided problem solving process model, an interactive process creates an animation that can work at two levels:

- at the low level it can replace the *eye movement control loop* thus allowing one to fixate on a specific region's changes (for instance as you drag a slider), or
- at a higher level during the *problem-solving strategy* one can build up an overall understanding of changes in the entire view (perform a visual query over the entire visualization at each animation frame).

The interaction during the higher level process, such as manipulating a decision threshold, can also be utilized for change-guided exploration, as the resulting movement in the visualization may draw attention to areas of interest. As an added benefit the resulting animation can also implicitly reveal the derivative of the uncertainty. In all such interaction a relationship may be intuitively formed between the manipulated variable (often 1D) and the resulting effect.

In general interaction methods can better support exploration and manipulation of dense and complex information spaces; utilizing this we can work toward promoting com-

Table 9.3: Categorization of the effect of uncertainty visualization on confidence.

| **Category** | **Uncertainty Visualization's Effect** |
|---|---|
| Type I | creates over confidence beyond what is substantiated |
| Type II | creates under confidence beyond what is substantiated |
| Type III | confounds interpretation within normal time constraints |

prehension by providing appropriate interactivity that aids turning information into knowledge. Research has shown that both adults and children develop new insights through information manipulation [Chapman, 1988], and it is this deeper understanding that we wish to enable. This is especially important for uncertainty as people may have difficulty understanding probabilistic information. Interactive queries showing the response to changes in uncertainty or potential realizations may be more easily comprehended.

### 9.5.7 Assess the implications of incorrectly interpreting the uncertainty

One initially assumes that adding uncertainty visualizations will be superior to omitting them. However as this is not guaranteed one should at least minimally consider the scenarios where it provides sub-optimal results. We present in Table 9.3 a categorization of the impact of an uncertainty visualization on the user's confidence. It can be related back to the Type I and II visualization errors [MacEachren, 1992, 1995, Ch.10] (see Table 2.4), of seeing patterns that do not exist, and failure to notice patterns and relationships, respectively, but concentrates on the final impact of the uncertainty visualization on the user's confidence in interpreting or decision making. This can also be viewed as looking at the accuracy and precision of any uncertainty visualization support aimed at satisfying Amar and Stasko's [2005] rationale-based precepts (expose uncertainty, concretize relationships, and formulate cause and effect). The Type III effect (confounding) was stated as a concern of military analysts in Watkins [2000] study.

For consideration of this directive it may be useful to compare the costs of these types of errors, and determine if any asymmetry or exceptions are handled appropriately. These errors can be compared against the visualization without any uncertainty visualization and its same errors, and this may form one criterion for investigating MacEachren et al.'s [2005] Challenge 4: understanding how (or whether) uncertainty visualization aids exploratory analysis. This directive may also raise some key issues that need to be described in any user documentation created for the visualization.

The next three sections will describe the guidance provided by our framework as it relates to specific implementation details from the three domain investigations in Chapters 5 through 8. As the directives were being considered during the development of these visualizations, and formalized post hoc, the style of their application intertwines the forms of both design and heuristic evaluation.

## 9.6 Framework Application: Archaeological Visualizations

This section will describe the application of the directives on the visualizations developed for archaeological purposes in Chapter 5. From this domain, example visualizations with and without the encoding of uncertainty are shown in Figure 9.3.

Figure 9.3: Example visualizations for archaeology. Left image has no uncertainty cue. Right images reveals uncertainty by the depth that the sphinxes have sunk into the sand.

### 9.6.1 Provide support for cognitive task simplification

In Table 5.1 in Chapter 5 we presented a summary of various representations for the temporal uncertainty in this domain and made special note of the number of levels of encoding possible in each. The flexibility to choose an encoding with the simplest mapping to cognitive task was available in our visualization tool ArkVis.

### 9.6.2 Support emphasis and de-emphasis of uncertainty information

Again the multiple representations available in the tool developed for this domain allowed for de-emphasis (e.g. transparency as shown in Figure 9.3) and emphasis (e.g. motion). The flexibility of using GPU programs for the encoding allows for potential custom-defined encodings to be developed like plug-ins and used with minimal effort.

### 9.6.3 Support viewing of uncertainty as metadata and separately as data

ArkVis did not provide for this particular aspect, as the uncertainty information could not be viewed on its own. This directive therefore prescribes creating a separate view or visual encoding to depict only the temporal uncertainty component. This might reveal the variability in the dating provided from different excavations or artifact types and so could be a useful visualization.

### 9.6.4 Allow the user to select realizations of interest

Arkvis provided the user with the ability to use a slider to move through time thereby selecting temporal realizations of interest. This directive suggests that for spatial uncertainty the user would benefit from the ability to directly drag artifacts to query potential locations, which was not provided by our tool.

### 9.6.5 Mitigate cognitive heuristics and biases with reasoning support

No particular support was offered with this system. We can therefore look for potential designs that this directive might suggest. If the user is comparing between a subset of the artifacts, then it may potentially be of use to allow selection of the items and provide a graph of their relative uncertainty at the currently selected date. This offloading of the comparison to a graphical representation explicitly designed for that purpose lessens the potential for bias. These graphs could then be saved for comparison over different specific dates, or also include time as one of the axes. The graphs could then be a reference for the probabilistic ranking of hypotheses on the chronology of artifacts.

### 9.6.6 Provide interaction to assist knowledge creation

The system provided the user with the ability to create animations via interactive controls over time (with the time slider), the time window (both standard GUI components and direct sketch/gesture), and all the variations of uncertainty encoding. Using these options allows the monitoring of visual changes while moving either forward or backward in time, which could assist in comprehending temporal uncertainty.

### 9.6.7 Assess the implications of inaccurately interpreting the uncertainty

Poorly calibrated confidence in this domain might mean that alternative interpretations of the data that should be explored further are not, or an incorrect hypothesis is not challenged. This over or under confidence could lead to relationships being assumed between artifacts that did not coexist, or the failure to explore true dependencies. For example, visual encodings such as transparency may need calibration so that very small probabilities are not completely overlooked if they are not very perceptible, and so don't result in overconfidence that something did not exist. Due to the multiple variations of encoding, the ability to scale the temporal query down to a precise date, and the possibility of completely turning off the uncertainty, we don't expect the visualization to confound the user.

## 9.7 Framework Application: Geophysical Visualizations

This section will describe the application of the framework directives on the visualizations developed for rock fracture modelling that were described in Chapter 6. Figure 9.4 provides two images of each of the two styles of uncertainty visualization developed in this domain.

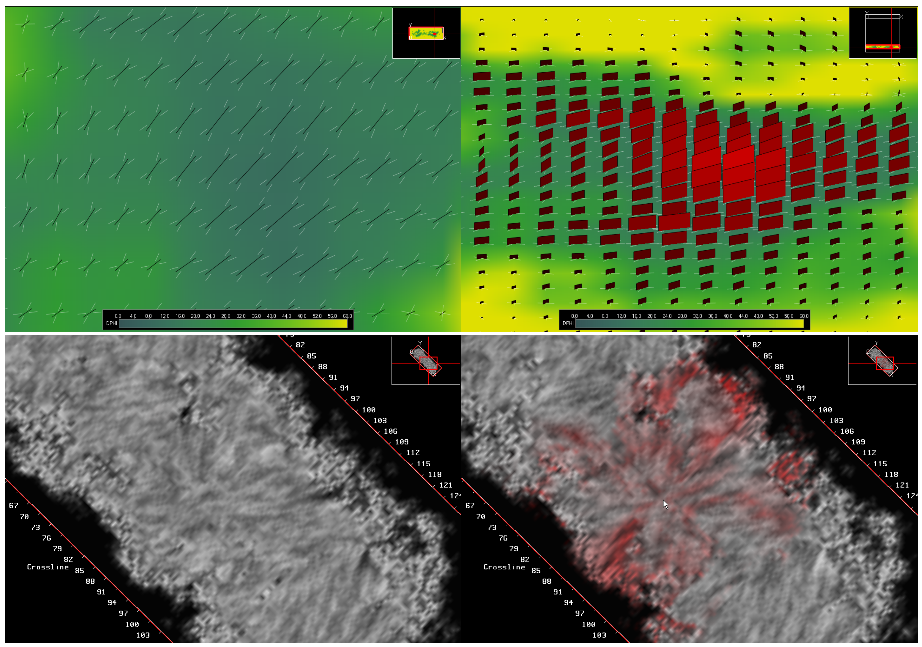



Figure 9.4: Example visualizations from seismic domain. Top images show two viewpoints of the uncertainty glyph. Bottom images show the effect of querying uncertainty in flow via red dye.

### 9.7.1 Provide support for cognitive task simplification

Both the glyph and flow visualization styles developed for this problem area provide for our two classifications of cognitive uncertainty thresholds. They both utilize a representation that allows for multi-leveled decoding, but both can be adjusted based on a user specified threshold. For the glyph visualization, this can control the presence of the glyph itself. With the flow visualization, adjustable thresholding of regions is performed by blending uncertain areas to black. Thus the user can eliminate areas from consideration; this helps the user avoid watching for patterns in regions of arbitrarily assigned directions[3]. These hard thresholds simplify the reading in that the judgment considerations used to determine the threshold can be cognitively "unloaded" after filtering to a desired confidence level, allowing other considerations to be made about the data itself.

### 9.7.2 Support emphasis and de-emphasis of uncertainty information

The first glyph form showing both magnitude and orientation (Figure 6.5) emphasizes uncertainty as it enlarges based on the uncertainty, while for the second glyph (Figure 6.6) there is less emphasis of uncertainty, and possibly even de-emphasis at low uncertainty, as the white coloured "whiskers" collapse down on each side of the black coloured orientation segment creating an edge enhancement effect. Switching between the two glyphs allows one to vary the focus between the fracture magnitude uncertainty and the orientation uncertainty. In the first glyph increased uncertainty in magnitude ($\sigma_{B_{ani}}$) adds emphasis as it increases the size of the glyph, similarly with very small magnitude uncertainty ($\sigma_{B_{ani}}$) the uncertainty components may be difficult to perceive even with large orientation uncertainty ($\sigma_{\Phi_{iso}}$). Using the second glyph that only encodes uncertainty in orientation ($\sigma_{\Phi_{iso}}$) is useful for considering fracture orientation in isolation, due to the possible overriding correlation with magnitude ($B_{ani}$) in the first glyph.

As opposed to the glyphs in which uncertainty is emphasized, the flow visualization naturally de-emphasizes uncertainty as the streamlines are precise with no diffusion. However, based on data and uncertainty parameters, colour can be injected like a dye as described by Botchen et al. [2005]. We use colour more specifically to provide visual feed-

[3]This is currently based on a lower threshold of magnitude ($B_{ani}$) as it is correlated to orientation, in that orientation ($\Phi_{iso}$) can not be defined at low levels of anisotropy.

back based on the user's interaction and testing of a specific realization. In Figure 9.5 you can see red dye being injected to indicate the angular uncertainty of flow that is redirected toward a user controlled attractor (sink). This process was illustrated in diagram form in relation to the glyph previously in Figure 6.8.

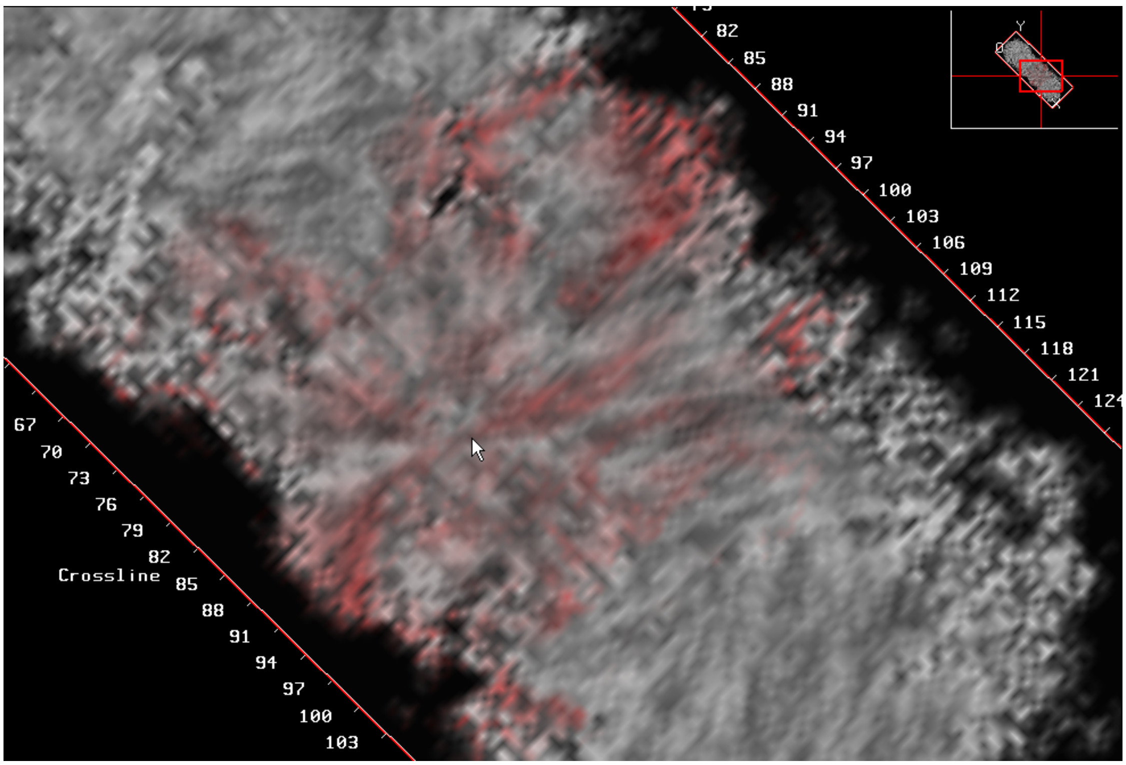



Figure 9.5: Flow visualization showing user directional and orientation query. Red dye injection is based on the difference between user requested flow orientation and the most likely direction.

### 9.7.3 Support viewing of uncertainty as metadata and separately as data

All parameters including uncertainty can be viewed as standard colour mapped slice planes through the volume, as illustrated in Figure 9.6. This colourmapped slicing allows the uncertainty to be viewed as data. The second glyph representation also provides a form of this in that it allows a simpler reading of the orientation uncertainty on its own as the magnitude ($B_{ani}$) can be replaced with a unit length vector. All the other visualizations described utilize the uncertainty as metadata.

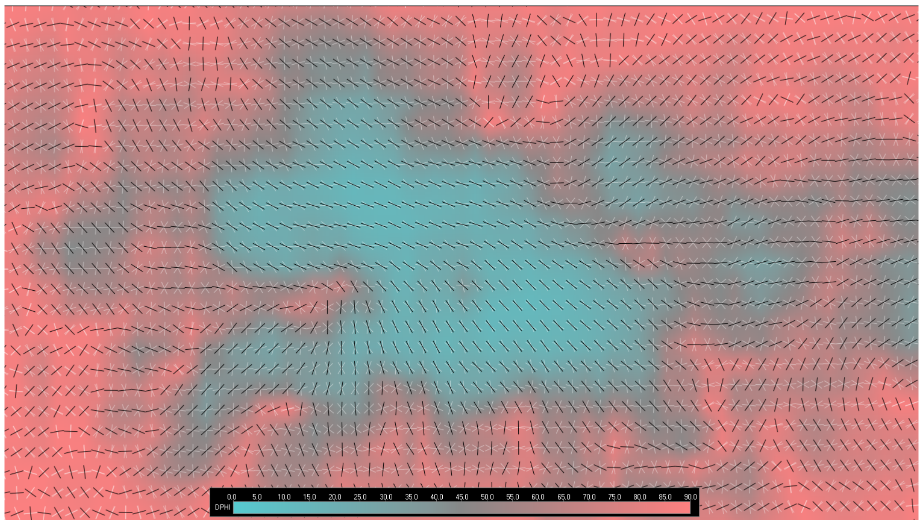



Figure 9.6: Use of an uncertainty glyph and colour mapped slice to allow inspection of uncertainty as data rather than metadata. A quality control inspection can look for correlated patterns created in a field of orientation uncertainty glyphs, while colour redundantly shows orientation uncertainty ($\sigma_{\Phi_{iso}}$)

.

### 9.7.4 Allow the user to select realizations of interest

The glyph by its design allows the user to imagine realizations within the $\pm$ one standard deviation boundary guides or "whiskers". While not as explicit as a complete outline (as Wittenbrink et al.'s [1996] glyph) it reduces the data-ink ratio while still providing constraints on the possibilities. The glyph is more natural for decoding vector or angular realizations than for example, colour coding, based on it being an iconic representation. Thus there is a direct perceptual resemblance of the representation to that which it stands for [O'Sullivan et al., 1994], versus a symbolic reading of a representation.

With the flow visualization we allow the user to interactively reverse local flow by moving a sink or source probe over the field with the cursor. Within a user defined distance of the sink cursor, any vector that points away from the cursor is automatically flipped and vice versa for the source. In the application domain this may correspond with reality in that an oil field well may either pump in fluids or be used for extraction. Similarly the user can also explore explicit realizations with the cursor probe as flow vectors are reoriented directly toward the cursor if this new vector lies within their orientation uncertainty (as defined by $\sigma_{\Phi_{iso}}$). The results of this interaction are shown in Figure 9.5.

### 9.7.5 Mitigate cognitive heuristics and biases with reasoning support

Basic extraction is supported by textual feedback of the parameter values at the current cursor position. Standard artifacts can be created by saving the screen to image file as well as the visualization configuration to an XML file. User studies could potentially be performed to determine potential heuristics being used by the interpreters.

This directive suggests possible features that might be added. Such as, if critical points in possible flow fields were automatically detected, they could be labeled showing the classification confidence. It may be of value to integrate this or the interpreter's confidence directly into the visualization as a decision aid.

### 9.7.6 Provide interaction to assist knowledge creation

Interactive visual queries can be performed by controlling the thresholds for showing the glyphs and flow. The interaction is required to provide the simpler views that are customized to answer a single threshold question. These and other functions were controlled using GUI based parameter manipulation, and the real-time cursor probe. Combined they allow various explorations of the data and uncertainty by manipulating the visualization to focus on the component of interest.

User interaction has been the enabling aspect of many of the features described in the previous subsections. Due to the large amount of extra information available when adding in the uncertainty (minimally doubling the raw data) interactive controls were provided to make it possible to visualize only the components that were deemed relevant to the current task. This allows the extra information to be managed based on user preferences. For the same reason interactive slice planes were used to reduce the complexity of visualizing the 3D volume of data. For flow visualization, de Leeuw and van Wijk [1993] provided an interactively placed data rich probe (glyph), which is another example where a large amount of detail is handled interactively in order to avoid overwhelming the viewer.

### 9.7.7 Assess the implications of inaccurately interpreting the uncertainty

For the flow visualization we should consider the possibility of Type I errors (overconfidence) in that the use of flow may overly suggest actual fluid flow. The fracture density does not directly assess actual fluid flow as many other factors come in to play. Thus, there is a danger one may get the false impression that the visualization suggests what actual fluid flow will be. This may be especially important for cases when an interpreter who knows the limitations is showing the visualization to a stake-holder (e.g. management, client). Similarly, our flow redirection was along a single vector, while other arbitrary non-linear paths are potential routes for the flow. These simplifications must be understood or the interpreter may get a false impression.

## 9.8 Framework Application: Evidence-based Medicine Visualizations

This section describes the application of the framework on the final domain of medical diagnosis which was described in Chapters 7 and 8. A screenshot of the visualization developed in this area is provided in Figure 9.7.

### 9.8.1 Provide support for cognitive task simplification

The system of visualizations for evidence-based medicine provided for cognitive task simplification through the use of multiple representations tailored for specific tasks. At the

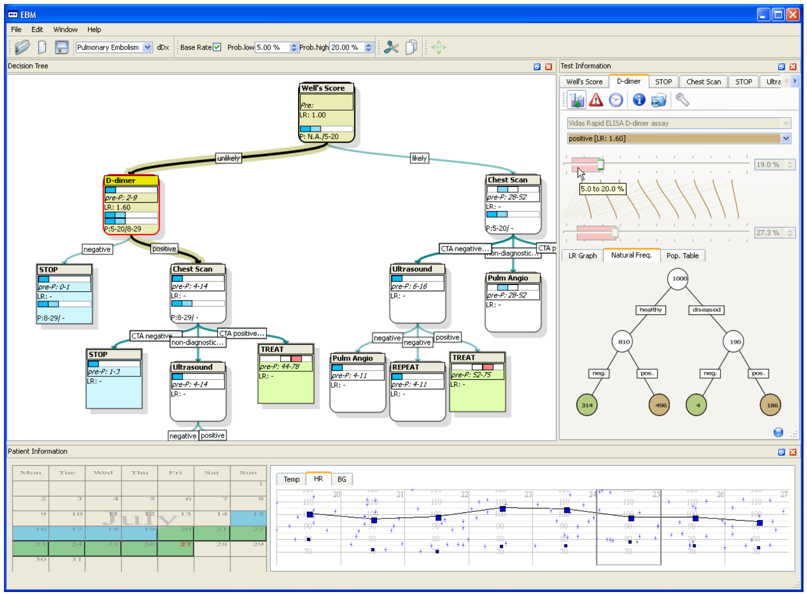


Figure 9.7: Example visualizations in the system created for diagnostic support in evidence-based medicine. Top left view provides diagnostic strategy visualization. Top right view illustrates the probabilistic evidence of a specific test. Bottom view shows calendar and graph of calendar selected test results.

tree level view the probabilities were reduced to five colour coded ranges (20% each), but text numerical values were provided as well. As some measure of patient utility should influence strategies and decisions [Weinstein and Fineberg, 1980], it is impossible to reduce the encoding levels to actions to 1:1 a priori. Other views provided the details for the calculation of the post-test probabilities at any node in the decision tree, along with uncertainty in these probabilities when there was likelihood ratio uncertainty. Thresholds were implicit in the branching structure of the tree, and visualized by emphasizing the recommended branch to follow.

### 9.8.2 Support emphasis and de-emphasis of uncertainty information

In this domain the main data is probabilities and so this directive is more applicable to the uncertainty in the probabilities, even though the probabilities are certainties as well. I will therefore term uncertainty in the probabilities second order uncertainty. This second order uncertainty was maintained as a range and so large uncertainty could be considered emphasized with the probability meter as it would be longer and more coloured regions would be visible. Some of the additional views emphasized the second order uncertainty (e.g. post-test probability graph with likelihood confidence interval, Figure 8.7), while others did not show it at all if it would overly complicate the representation.

### 9.8.3 Support viewing of uncertainty as metadata and separately as data

This directive suggests a possible additional visualization of showing only the magnitude of the probability range. This would simplify viewing of the effect of the various tests on precision of the probability estimate ($2^{nd}$ order uncertainty). It might also suggest a redesign of the probability meter to also show this aspect more clearly.

### 9.8.4 Allow the user to select realizations of interest

The decision tree exploration was made completely under the users' control. Unlike an expert system dictating the pathway to the "optimal" outcome, this design provided visual evidence rather than dictating decisions. This freedom to select realizations of interest removed the potential scenario found in the contextual interviews where "tricks" to circumnavigate the system support needed to be shared among users.

### 9.8.5 Mitigate cognitive heuristics and biases with reasoning support

This directive was a fundamental motive for the design of the entire system as its main purpose was to provide reasoning support. The main visualizations were developed to allow the offloaded application of Bayesian reasoning to be appropriately integrated into the physician's decision support process. The visualization system also provided an additional reference view of observational/measured data and their uncertainty. This is important as uncertainty interferes with judging typicality and this may lead the decision maker to gather more data [Klein, 1998], which in this scenario might include further use of our decision support tool.

### 9.8.6 Provide interaction to assist knowledge creation

The probability sliders and the corresponding representations for post-test probabilities allowed this type of exploration of the likelihood ratios. Interactive selection of single probability was important when showing uncertainty in likelihood ratios as the interpretation and comprehension of the three post-test probability functions† are difficult when trying to look at more than a single pretest probability value. Additionally the interactive control over the base-rate probability allowed changes in the pre/post-test probabilities of the entire tree to be observed, and perhaps reveal the design motivation for the decision tree structure itself.

Interaction was also part of the drill-down process to explore the details and evidence-base of support for specific decision tree nodes (tests) and branches. The layout and interaction supported was designed to allow specific details to be viewed and manipulated while keeping the context in view.

### 9.8.7 Assess the implications of inaccurately interpreting the uncertainty

These implications are important in this domain as they relate to safety issues and patient morbidity. Overconfidence may lead to the treatment of false positives, or the sending home of false negatives without treatment. Under confidence may delay treatment which may have serious consequences, waste scarce resources, or prolong unnecessary stress by keeping people in the hospital longer than required.

This tangentially relates to the test availability visualization that raised some of the more heated discussion in the cognitive walk-through evaluation. The temporal uncer-

†One is the standard post-test probability function of pretest probability, and the other two represent the curves along the 95% confidence interval of the likelihood ratio.

tainty visualization was thought to potentially lead to the use of inferior tests based on availability (thus weaker confidence), and this was deemed inappropriate for a policy of providing the best possible care. Thus it was suggested that the temporal uncertainty visualization could lead indirectly to under confidence.

## 9.9 Conclusions

In the main portion of this chapter I provided a framework, in the form of directives, that is both descriptive and prescriptive of the general processes encompassing uncertainty visualization. One additional consideration singled-out was the dichotomy of decisions using single uncertainty thresholds (worst-case, best-case, accept threshold, reject threshold) versus those using multiple thresholds or weighing of evidence (levels of confidence). Considering direct facilitation of both categories of cognitive decisions along with all the framework directives can guide the creation of more effective visualizations.

In discussing the relevance and application of this framework to the visualizations of uncertainty in multiple domains, I have shown its potential general applicability for design and evaluation. Considering and utilizing these directives should help to reach two goals:

1. to provide the interactive flexibility to aid in the performance of unpredicted cognitive tasks, and
2. to assist mapping the uncertainty visualization to match the user's cognitive needs, thereby reducing their cognitive load.

Allowing interactive control of the described aspects of the visualization is key to providing the mappings necessary to simplify arbitrary tasks as much as possible. The second goal, mapping the uncertainty to match cognitive requirements, is important as the additional information added in an uncertainty visualization will raise the user's cognitive load and so may be an obstacle to understanding [van Bruggen et al., 2003] rather than increasing comprehension and confidence. Using our directives to assist in reaching these goals will move us toward meeting MacEachren et al.'s [2005] challenges for uncertainty visualization.

# Chapter 10

# Conclusions

**Bridgekeeper:** What... is your quest?
**King Arthur:** To seek the Holy Grail.
**Bridgekeeper:** What... is the air-speed velocity of an unladen swallow?
**King Arthur:** What do you mean? An African or European swallow?
– Monty Python. *Monty Python and the Holy Grail* (1975)

In this chapter I will summarize the findings from the research contained in the previous chapters. These contributions include: components relating to the evaluation of uncertainty visualizations, design recommendations for supporting cognition under uncertainty, specific visualization implementations in three different domains, and design criteria for supporting diagnostic decision making in evidence-based medicine. The previous chapter contained conclusions based on generalizations in the form of directives, but these will not be reiterated in detail again, only extrapolated further into the problem context. Continuing toward more general contributions, progress toward the high-level goals stated in Chapter 1 will be discussed. Finally the chapter ends in the same way as the dissertation began, with a commentary on methodology, future work, and a look at the big picture.

## 10.1 Overview

The beginning of this dissertation related uncertainty visualizations to both general communication and decisions (actions). Exploring these in more detail provided insights into important factors to consider in creating uncertainty visualizations. As stated in Chapter 1, the process of translating written works can be considered a useful metaphor for looking at the issues involved in uncertainty visualization. Translation has to consider low-level issues such as choosing words, akin to graphical encodings, and high-level ones such as allusions, similar to association driven heuristics and biases. As there is uncertainty in the process itself [Pang et al., 1997], portraying uncertainty accurately for effective communication is a challenging problem. The effectiveness of this process will thus influence the interpretation which is then utilized to take action or make decisions. Improving this process was the motivation for this research.

A large portion of this dissertation dealt with building on existing theory from multiple disciplines. Similarly, expert knowledge from numerous collaborators infiltrated the discussions presented in most chapters. This is appropriate for an investigation into as general a problem as uncertainty visualization, as one needs to take a broad perspective. Similarly expansive, the focus of this summary chapter will be on generalizations and so significant insights from one particular domain or aspect may not be reiterated again. This is not intended to suggest relative merit, only generalizability.

Two strategies for researching these challenging issues were utilized. One was to work from a theoretical perspective and attempt to apply and extrapolate existing theory from information visualization, human computer interaction, and cognitive psychology. The other strategy took a more practical approach of delving into specific domains and developing uncertainty visualizations sensitive to the user and task considerations. Both strategies could lead to progress with regard to the open challenges of uncertainty visualization, but it was hoped that each method have unique strengths. The contributions resulting from these strategies will be summarized in the next section.

## 10.2 Summary of Research Contributions

This section will reiterate the contributions of this dissertation in a concise manner to overview the progress that was made. The order of presentation matches the ordering in which they appear in the chapters.

### 10.2.1 Evaluation of Uncertainty Visualizations

One of MacEachren et al.'s [2005] challenges for uncertainty visualization was assessing the usability and utility of uncertainty capture, representation, and interaction methods and tools. This corresponds to the first research contribution of this dissertation, involving the assessment of representations and interaction methods. Chapter 3 provided a look at using heuristic evaluation to further analyze and understand representational design and other aspects of uncertainty visualizations. For this purpose, a set of heuristics were extracted from information visualization theory provided by Bertin [1983], Tufte [2001], and Ware [2004]. The research results provided in Chapter 3 illustrated that this discount evaluation method, not previously utilized to any appreciable extent for visualization assessment, could provide insights into the strengths and weaknesses of existing uncertainty visualizations, and improve the understanding of particular design trade-offs[1].

Exploring the potential of heuristic evaluation further, higher level heuristics extracted from Shneiderman [1996] and Amar and Stasko [2004] were also considered and discussed for the evaluation of visualizations in general. We proposed a methodology for utilizing hierarchical levels of heuristics, and discussed ways of organizing and refining such heuristic sets. Our strategy of applying the three sets of heuristics revealed issues such as heuristic overlap, which posed the question of finding the minimal spanning set of heuristics, and experience level of the evaluator, which was a more significant factor with the high-level heuristics. These results demonstrated the value of another evaluation methodology which can be used at various points the iterative design process that is common for visualization development[2].

[1]This work was published in Zuk and Carpendale [2006].

[2]The results of this investigation were published in Zuk et al. [2006].

### 10.2.2 Relation of Uncertainty Visualization to Cognitive Reasoning Issues

In Chapter 4 we turned to the human factors side of things and explored the role of uncertainty in the reasoning process. From cognitive theory we coalesced important aspects of reasoning under uncertainty and tried to understand them from the perspective of uncertainty visualization in terms of constraints and requirements. A major component of this involved reviewing and applying the existing research into cognitive heuristics and biases [Kahneman et al., 1982]. This theory proved to be valuable in offering insights during the domain investigations. Providing support for heuristics and mitigating biases may be a way for visualization to amplify cognition. Card et al.'s [1999] six basic ways to amplify cognition using information visualization, were through:

1. increasing cognitive resources (expanding working memory),
2. reducing search,
3. enhancing the recognition of patterns,
4. supporting the easy perceptual inference of relationships,
5. perceptual monitoring, and
6. providing a manipulable medium.

Thus, to this list we add a seventh way to amplify cognition through mitigating potential heuristics and biases. One technique, for example, may be through providing introspection on the reasoning process. Other potential techniques were suggested in Chapter 4 and 9.

This research was intended to further the understanding how knowledge and ignorance of uncertainty affects analysis and decision making. Ignorance was one of the two top-level types of Watkins's [2000] in depth taxonomy of uncertainty, and should be considered separately to knowledge to ensure it receives proper attention. Based on the cognitive effects of uncertainty we presented a pipeline for visualizing reasoning uncertainty, to be used in parallel with the Pang et al.'s [1997] uncertainty visualization pipeline, and recommended its use in supporting decisions and interpretation. Explicit visual support for introspection was suggested as a key benefit of this design. Both self-monitoring and offloading of cognitive computation from this proposed approach may help address the limitations of reasoning under uncertainty[3].

### 10.2.3 Domain Investigations of Uncertainty Visualization

The specific uncertainty visualizations developed during the domain investigations are contributions in their own right, but they also served the dual purpose of providing the experience and knowledge for abstracting grounded theory. The variety of issues and tasks involved in each chosen domain area were significantly different, thus allowing each to contribute to a more general understanding of uncertainty visualization.

#### Temporal Uncertainty in Archeology

Issues relevant to uncertainty in understanding and presenting archaeological site data and reconstructions were presented in Chapter 5. An example archaeological site spanning multiple time periods is shown in Figure 10.1. One result of this research was the creation of a visualization system for archaeological site data. The visualization provided a virtual world in which the user could navigate freely (walk, fly, etc.) or jump to specified landmarks, and at any point, animate in time using a time slider or the novel time window control. The system provided for various encodings to visualize temporal uncertainty in different datasets existing at a common spatial location (dig site). New uncertainty encodings were introduced in addition to existing techniques, and the system offered the ability to use both predefined and dynamically loaded uncertainty cues based on GPU programs.

The virtual presentation of the archaeological data is a common medium for communicating findings, both to other archaeologists or researchers and to the general public. The communication goals and the audience therefore play a role in the selection of graphical encoding methods. More intuitive encodings (e.g. iconic, in the semiotic sense [O'Sullivan

[3] This work was published at the 2007 Smart Graphics Conference [Zuk and Carpendale, 2007].

Figure 10.1: Archaeological site with various structures from different time periods.

et al., 1994]) may be more applicable to museum environments, while the need for more accurate reading may prevail in research communities. Relating specific domain tasks to potential uncertainty visualizations was one contribution of this work[4].

**Model Uncertainty in Geophysics**

The seismic domain has many aspects involving uncertainty and a glimpse into some of the larger context is provided in Figure 10.2. Our visualizations developed for the seismic domain, described in Chapter 6, provided a look at the uncertainty in modelling rock properties related to fractures. The problem was to understand a 3D volume of two model attributes along with their uncertainty. In the visualizations that were developed, uncertainty could be explored in multiple ways based around a glyph or flow representation, both providing an encoding of the uncertainty in magnitude and in orientation of the rock fractures. Thus, one research contribution was providing a visual tool to assist the geo-

[4]This was presented to the archaeological, cultural heritage, and visualization communities at VAST 2005 [Zuk et al., 2005].

Figure 10.2: Illustrations of the larger seismic domain context. Left image: traditional paper-based visualizations of earth models. Right image: potential time varying earth surface and specialized trucks for acquiring seismic data. [©2007 CGGVeritas]

physicist in understanding their modelling results[5].

These two novel interactive representations were compared and contrasted in order to understand the benefits of each. Selecting one of the two representations or combining them allowed the uncertainty to be visualized in a manner tailored to suit specific task requirements (e.g. quality control versus client presentation)[6]. This contributed to understanding the role of interactivity. These and other findings were used in grounding the framework directives, which will be summarized separately later.

**Diagnostic Uncertainty in Evidence-based Medicine**

Chapters 7 and 8 described a more grounded investigation into the uncertainty in the task of evidence-based medical decision making. With this approach we began with an observational study to better grasp how existing software support integrates with the physician's task. Figure 10.3 illustrates the domain setting and where the software support can potentially be used. Analyzing the observational and contextual interview components we developed design implications for support, and in addition we created a structured organi-

[5]With a focus on the geophysical aspects, results have also been presented in Downton et al. [2007a,b].
[6]This work will be published in Visualization and Data Analysis 2008 [Zuk et al., 2008]

Figure 10.3: Hospital ward with mobile computer terminals. [©2007 C. Tang]

zation of the uncertainty involved. The design implications as well as a task performance model stand on their own as contributions to inform other designs for supporting this task domain.

Based on the aforementioned study analysis we began a participatory design process with two domain experts to prototype the visualizations for decision support. Using this methodology we developed a visualization that took on a more passive role to provide visual evidence to support the diagnostic process. Rather than making the "black box" style decision recommendations that were observed in the existing system, we developed multiple representations for exploring and understanding the different components of uncertainty and their role in the diagnostic strategy. Exposing the evidence (motivation) for hospital recommended decision trees and making the uncertainty transparent to the physician was intended to provide cognitive support more compatible with the users needs.

Multiple new interactive uncertainty representations were developed, with the most noteworthy being the decision-tree interface for understanding the probabilistic nature of the recommended strategy. Of Card et al.'s [1999] six ways, and our additional 7th way, to amplify cognition using information visualization, only one way was not directly utilized (method 5). The visualizations were qualitatively evaluated and overall were well received, supporting the design implications we derived from the initial study in Chapter 7. The mainly qualitative evaluation results provided in Chapter 8 exposed other motivations and critiqued the visual representations. This provides guidance for further refinement of the visualizations, and perhaps for uncertainty visualization in general [7].

### 10.2.4 Framework for Supporting Uncertainty Visualization

Chapter 9's contribution involved generalizing insights and practical knowledge on uncertainty visualization from the problem domains, as well as integrating issues from the earlier cognitive and evaluative centered chapters.

#### Cognitive Uncertainty Categorization

Probing into the interpretation of uncertainty visualizations, I proposed a categorization based around the use of thresholds in a decision process. The ability to see a clearly defined component in an otherwise ambiguous or probabilistic set of data provides a clear reference frame, which is valuable for tasks, such as, making comparisons. This may relate to Gigerenzer and Hoffrage's [1999] finding that natural frequencies (akin to specific realizations) may be more easily utilized than probabilities for Bayesian reasoning.

Thus, the aspect of decision thresholds was chosen as a potential avenue to impose some structure on what is often a hidden or informal reasoning process. Applying this categorization provided a way of grouping cognitive tasks tied to reaching decisions, and potentially organizing cognitive support. As one example, it provided the concept of *decision boundary mapping*, which may be considered a type of cognitive affordance, and benefit us in the same way that physical affordances help us determine possible actions.

[7]Aspects of this work have been submitted for publication, and we expect to submit the other components.

**Directives for Uncertainty Visualization**

In one of the larger specific research contributions, a set of directives were put forth in Chapter 9 in order to guide the development and evaluation of uncertainty visualizations. The directives were:

1. Provide support for cognitive task simplification,
2. Support emphasis and de-emphasis of uncertainty information,
3. Support viewing of uncertainty as metadata and separately as data,
4. Allow the user to select realizations of interest,
5. Mitigate cognitive heuristics and biases with reasoning support,
6. Provide interaction to assist knowledge creation, and
7. Assess the implications of incorrectly interpreting the uncertainty.

They are quite general and capture issues spanning both data uncertainty and reasoning uncertainty. The directives also integrate reasoning into design aspects of representation and interaction.

These directives were then applied to the three domain visualizations that were developed. This exemplified how they were useful in describing existing functionality as well as prescribing potential new functionality in each area. Their largest impact may come from their ability to bring cognitive issues into a design space in a light-weight manner. How these directives may inform visualizations in other domains will be an area of future research.

## 10.3 Progress Toward Goals

As Chapter 9 already noted the relation of the directives to MacEachren et al.'s [2005] challenges (goals), in this section I will describe progress toward the high level goals put forward in Chapter 1.

### 10.3.1 General Advice on Uncertainty Visualization

In Chapter 1 I stated a main goal was to look for commonality in all the types of uncertainty visualization. To reach this goal both existing uncertainty visualizations were evaluated (Chapter 3) and new uncertainty visualizations were developed in disparate domain areas (Chapters 5 to 8). The variety of users, needs, and tasks, that were researched strengthen the potential generalizability in any findings. Thus, by selecting diverse problem areas the commonality we discover may form the basis of general theory.

The research results regarding heuristic evaluation of uncertainty visualizations form the basis of a generalization. The guidance from the design and evaluation principles allow all uncertainty visualizations to be measured on common ground. Similarly, general understanding requires a set of common principles. Aspects of the heuristic evaluation and the creation of a more formal set of information visualization theory, represent a generalization, which was carried over into the archaeological domain investigation. In that domain a specific focus on understanding the options for encoding uncertainty, in a manner borrowed from Bertin's [1983] graphic variables (length, order, and other characteristics) provided a methodology that would be applicable in general to the structured development of options for uncertainty encoding.

The directives provided in Chapter 9 represent the amalgamation of important issues found throughout this dissertation. Their external validity was partially demonstrated by their application in Chapter 9 to the domain fields other than where they may have been primarily motivated. Some of the directives represent high-level aspects but they are all firmly grounded in the needs of uncertainty visualization and so it is hoped they may be more easily applied in a heuristic manner than the high-level heuristics which posed problems in Chapter 3. Further research into their application will be required to better assess their generality and utility.

### 10.3.2 Understanding the Relationship between Uncertainty and Decisions

Another specific goal stated in Chapter 1, was to further the understanding of how uncertainty fits into a complete and accurate interpretation and decision model. Looking at these issues often requires the tacit knowledge of a domain expert and so this hurdle was tackled in two ways:

1. establishing collaborations with the experts in the respective fields, and
2. lessening the distance between myself and the domain by becoming active in the respective research or participating in the research communities.

This methodology along with the exploration of cognitive psychology theory in Chapter 4 formed the foundation for relating uncertainty to the user's decision making process.

A specific targeting of uncertainty in decisions was most prevalent in the research on the decision processes in evidence-based medical diagnosis, as described in Chapters 7 and 8. Contributions in that domain included developing methods for encoding uncertainty relevant to decision processes into visualizations. Future evaluations of the developed visualizations may aid the understanding how knowledge of uncertainty influences decision makers. Given some of the strong feedback during our focus group evaluation, (e.g. "This tool should be mandatory for groups that publish and distribute guidelines for therapies"), their exists some desire for the additional support our visualizations provided.

Looking back at Johnson and Sanderson's [2003] statement given in the Chapter 1, that a primary goal of effective visualization is to provide a complete and accurate visual representation, I would argue that this is only a noteworthy goal when complete and accurate are defined in terms of the user's decision needs. In other words, ***an effective visual representation is one that leads to as complete and accurate a cognitive representation as required to perform a task or make a decision***. While this is also an ambiguous definition, it may be a better benchmark. The directives and dichotomy of uncertainty thresholds in decision making, provided additional criteria for pursuing this redefined goal.

## 10.4 Final Thoughts

In conclusion, the two strategies worked quite successfully to explore uncertainty on different levels. The domain investigations even within themselves provided contributions, and built up a set of issues that helped guide the generalization process. While the diversion into the aspects of cognitive reasoning may have biased the end result toward those concerns, this is not necessarily a bad bias to have.

### 10.4.1 Commentary on Methodology

Some of you may now be thinking, "But I paid for an argument!". For those readers who expected to find the test of an alternative hypothesis versus the null hypothesis, I hope you now realize that you came to the wrong place[8]. For those who did not get the preceding allusion to Chapman et al. [1972], and thus the humour, arguments do not always follow the formula you expect.

The qualitative methods I have utilized attempted to consider and capture important insights from the big picture, rather than specific performance metrics that might compare individual implementation choices. Qualitative methodologies, such as thematic analysis [Boyatzis, 1998] offer a potential to analyze and generalize from highly varying types of data (e.g. domain specific implementations and existing theory) where quantitative methods may offer little direction. Appropriately, North [2006] has suggested qualitative methodologies may be valuable in trying to further understand the issues related to measuring insight. Controlled experiments do not lend themselves toward tools requiring significant training or lengthy exploratory processes, and for these cases qualitative evaluation methods have shown value, such as Seo and Shneiderman's [2006] evaluation of their knowledge discovery tool, the Hierarchical Clustering Explorer.

While quantitative methodologies may more faithfully provide incremental advances,

[8]That place is just down the hall in the third room on your left.

they often tend toward testing predetermined hypotheses. However, looking for a rich understanding of a problem space is more important during initial exploration, and thus qualitative methods may be more appropriate. Quantitative methods have their place at the forefront of focused evaluation, and can also be used to analyze qualitative data (e.g. to find differences in sampling groups). Small numbers may fail to provide statistical power, but as shown in our medical domain investigations, they may still provide the macro level insights needed to understand what are important design factors to consider. Quantitative methods used too early in development may even be counter productive, in the same way as prematurely optimizing code in software development can lead to poor designs.

When looking for deeper understanding or trying to make large advances, a qualitative inquiry approach may offer more hope of success. A qualitative methodology may not have statistically proven specific points in the case of this dissertation, but it has rewarded us with many results that may guide our way. Returning to Einstein for a final quote[9], "Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted." Thus, in looking for things that count, it is hoped that the contributions in this dissertation may add insight into uncertainty visualization in general, or at least be a step in that direction.

### 10.4.2 Future Work

Each of the visualizations developed in the domain areas may be further refined and evaluated. The archaeological visualization tool could be utilized for educational purposes in a public cultural heritage display, but further evaluation would be required to determine its role in a more formal setting. The seismic uncertainty visualization is available in a widely used software package available at CGGVeritas offices worldwide, therefore future evaluations may be performed to refine it further as its utilization grows. Utilizing the medical visualization for educational purposes would be a natural next step, as direct integration

[9]This was written on a sign in his office and is attributed to him.

into a clinical setting would require a more formal validation and also the development of a process for continually updating the evidence base.

The directives from Chapter 9 also provided guidance on potential new functionality that should be explored. Another domain that was explored but not developed to the point of warranting inclusion in this dissertation, was that of visualizing uncertainty in stochastic simulations of mountain pine beetle management strategies. Both a stochastic simulation output browser that I developed (which was not described in this dissertation) and the *LuMPB Key* (Landscape unit Mountain Pine Beetle Key) decision support tool [Schlesier et al., 2006], evaluated with heuristics in Chapter 3, could be evaluated with the directives. It would also be appropriate to consider evaluating other published uncertainty visualizations, or creating new ones, using these directives directly.

### 10.4.3 Conclusion

As we live in a world where uncertainty is ubiquitous, it is important that we visualize it appropriately. Exposing it in an intuitive way, and interactively exploring both data and reasoning uncertainty may enhance the viewers comprehension, confidence, and ultimately task performance. To assist the achieving of this goal this dissertation has made numerous contributions including new representations, processes, and a set of directives. In designing uncertainty visualizations we must act as translators and attempt to preserve original content and meaning, but should allow the viewer to determine the language of their own comprehension.

# Bibliography

Daniel Acevedo, Eileen Vote, David H. Laidlaw, and Martha S. Joukowsky. Archaeological data visualization in VR: analysis of lamp finds at the Great Temple of Petra, a case study. In Thomas Ertl, Ken Joy, and Amitabh Varshney, editors, *Proceedings of the Conference on Visualization 2001 (VIS-01)*, pages 493–496, Piscataway, NJ, October 21–26 2001. IEEE Computer Society.

Christopher Ahlberg, Christopher Williamson, and Ben Shneiderman. Dynamic queries for information exploration: An implementation and evaluation. In *Proceedings of ACM CHI'92 Conference on Human Factors in Computing Systems*, Graphical Interfaces for Drawing, Exploring, and Organizing, pages 619–626, 1992. URL `http://www.acm.org/pubs/articles/proceedings/chi/142750/p619-ahlberg/p619-ahlberg.pdf`.

F. P. Albright. Excavations at Mārib in yemen. In R.L. Bowen and F.P. Albright, editors, *Archaeological Discoveries in South Arabia*, volume II, pages 215–268. Publications of the American Foundation for the Study of Man, Baltimore, 1952.

Diane Lindwarm Alonso, Anne Rose, Catherine Plaisant, and Kent L. Norman. Viewing personal history records: a comparison of tabular format and graphical presentation using lifelines. *Behaviour & Information Technology*, 17(5):249–262, 1998.

R. Amar and J. Stasko. A knowledge task-based framework for design and evaluation of information visualizations. In *Proc. of the IEEE Symposium on Information Visualization*, pages 143–149, 2004.

R. Amar and J. Stasko. Knowledge precepts for design and evaluation of information visualizations. *Visualization and Computer Graphics, IEEE Transactions on*, 11(4): 432–442, 2005.

anonymous. *The Nibelungenlied*. Thomas Hatto, Trans. Penguin Classics, 1969. Original circa. 1200.

David Arnott. Cognitive biases and decision support systems development: a design science approach. *Information Systems Journal*, 16:55–78, 2006.

David Astels, Granville Miller, and Miroslav Novak. *A Practical Guide to eXtreme Programming*. Prentice Hall PTR, 2002.

Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. Guidelines for Using Multiple Views in Information Visualization. In *Proceedings of AVI '00*, pages 110–119, New York, USA, 2000. ACM Press. doi: http://doi.acm.org/10.1145/345513.345271.

C. Baracchini, A. Brogi, M. Callieri, L. Capitani, P. Cignoni, A. Fasano, C. Montani, C. Nenci, R. P. Novello, P. Pingi, F. Ponchio, and R. Scopigno. Digital reconstruction of the Arrigo VII funerary complex. In *VAST 2004: Virtual Reality, Archaeology and Cultural Heritage*, 2004.

M. K. Beard and B. Buttenfield. Detecting and evaluating errors by graphical methods. In P. Longley, M. Goodchild, D. Maguire, and D. Rhind, editors, *Geographic Information Systems: Principles and Technical Issues. Vol. 1*, pages 219–233. John Wiley, New York, 1999.

Eta S. Berner, editor. *Clinical Decision Support Systems: Theory and Practice*. Springer-Verlag, 1999.

Jacques Bertin. *Graphics and Graphic Information Processing*. William J. Berg and Paul Scott, Trans. Walter de Gruyter, Berlin, 1981. Trans. of *La Graphique et le Traitement Graphique de l'Information*. 1977.

Jacques Bertin. *Semiology of Graphics*. William J. Berg, Trans. The University of Wisconsin Press, 1983. Trans. of *Semiologie Graphique*. 1967.

Nigel Bevan, Carol Barnum, Gilbert Cockton, Jakob Nielsen, Jared Spool, and Dennis Wixon. The "magic number 5": is it enough for web testing? In *CHI '03: CHI '03 extended abstracts on Human factors in computing systems*, pages 698–699, New York, NY, USA, 2003. ACM Press. ISBN 1-58113-637-4. doi: http://doi.acm.org/10.1145/765891.765936.

Randolph G. Bias. The pluralistic usability walkthrough: Coordinated empathies. In J. Nielsen and R. Mack, editors, *Usability Inspection Methods*, chapter 3, pages 63–76. Wiley and Sons, 1994.

Edgar R. Black, Donald R. Bordley, Thomas G. Tape, and Robert J. Panzer, editors. *Diagnostic Strategies for Common Medical Problems*. Americal College of Physicians - American Society of Internal Medicine, second edition edition, 1999.

Ralf P. Botchen, Daniel Weiskopf, and Thomas Ertl. Texture-based visualization of uncertainty in flow fields. In *Visualization, 2005. VIS 05. IEEE*, pages 647–654, 2005.

Richard E. Boyatzis. *Transforming qualitative information: Thematic analysis and code development*. SAGE Publications Inc., 1998.

M. Boyle and S. Greenberg. The language of privacy: Learning from video media space analysis and design. *ACM Transactions on Computer-Human Interaction*, 12(2):328–370, June 2005.

K. Bürger, J. Schneider, P. Kondratieva, J. Krüger, and R. Westermann. Interactive visual exploration of unsteady 3d flows. In *Proceedings of Eurographics/IEEE VGTC Symposium on Visualization (EuroVis 2007, May 23–25, 2007, Norrköping, Sweden)*, Eurographics Workshop Series, Aire-la-Ville, 2007. Eurographics.

Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco, CA, USA, 1999. ISBN 1-558-60533-9.

C. Melody Carswell. Choosing specifiers: an evaluation of the basic tasks model of graphical perception. *Human Factors*, 34:536–554, 1992.

CBC Radio. Ideas: In other words. http://www.cbc.ca/ideas/, 2007. (podcast in 3 parts).

CBS Broadcasting Inc. Rewriting the science. http://www.cbsnews.com/stories/2006/03/17/ 60minutes/main1415985.shtml, 2006. visited: April, 2007.

Andrej Cedilnik and Penny Rheingans. Procedural rendering of uncertainty information. In T. Ertl, B. Hamann, and A. Varshney, editors, *Proceedings Visualization 2000*, pages 77–84. IEEE Computer Society Technical Committee on Computer Graphics, 2000.

J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey. *Graphical Methods for Data Analysis*. Wadsworth Internat. Group, 1983.

Graham Chapman, John Cleese, Terry Gilliam, Eric Idle, Terry Jones, and Michael Palin. Argument clinic. In *Monty Python's Flying Circus*, Episode 29. BBC, 1972.

Loren J. Chapman and Jean Chapman. Test results are what you think they are. In D. Kahneman, P. Slovic, and A. Tversky, editors, *Judgment under Uncertainty: Heuristics and biases*, pages 239–248. Cambridge University Press, Cambridge, 1982.

M. Chapman. *Constructive evolution: Origins and development of Piaget's thought*. Cambridge University Press., 1988.

J. R. Charbonnell, J. L. Ware, and J. W. Senders. A queuing model of visual sampling:

Experimental validatiaon. *IEEE Transactions on Man-Machine Systems*, MMS-9:82–87, 1968.

Kathy Charmaz. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. Sage Publications, 2006.

W.S. Cleveland. *The Elements of Graphing Data*. Wadsworth, 1985.

Diane Cluxton and Stephen G. Eick. DECIDE hypothesis visualization tool. In *Int. Conf. on Intelligence Analysis*, 2005.

Marvin S. Cohen, Jared T. Freeman, and Steve Wolf. Metarecognition in time-stressed decision making: Recognizing, critiquing, and correcting. *Human Factors*, 38(2):206–219, 1996.

Marvin S. Cohen, Raja Parasuraman, and Jared T. Freeman. Trust in decision aids: A model and its training implications. In *Proceedings 1998 Command and Control Research and Technology Symposium*, 1998.

B. Collins. From Ruins to Reality - The Dresden Frauenkirche. *IEEE Computer Graphics and Applications*, pages 13–15, November 1993.

Christopher Collins, Sheelagh Carpendale, and Gerald Penn. Visualizing uncertainty in lattices to support decision-making. In *Proceedings of Eurographics/IEEE VGTC Symposium on Visualization (EuroVis 2007)*. Eurographics Association, 2007.

B. Craft and P. Cairns. Beyond guidelines: what can we learn from the visual information seeking mantra? In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pages 110–118. IEEE, 2005.

John W. Creswell. *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Sage Publications, Thousand Oaks, Calif., 1998.

J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America*, 2(7):1160–69, 1985.

Trevor J. Davis and C. Peter Keller. Modelling and visualizing multiple spatial uncertainties. *Computers & Geosciences*, 23(4):397–408, 1997.

Robyn M. Dawes, David Faust, and Paul E. Meehl. Clinical versus actuarial judgment. In T. Gilovich, D. Griffin, and D. Kahneman, editors, *Heuristics and Biases: The Psychology of Intuitive Judgment*, pages 716–729. Cambridge University Press, Cambridge, 2003.

Willem C. de Leeuw and Jarke J. van Wijk. A probe for local flow field visualization. In *IEEE Visualization '93*, pages 39–45, 1993.

J. Dewey. *How We Think*. Dover Publications, 1997.

Suzana Djurcilov, Kwansik Kim, Pierre Lermusiaux, and Alex Pang. Visualizing scalar volumetric data with uncertainty. *Computers & Graphics*, 26:239–248, 2002.

Robert Dougherty and Alex Wade. Vischeck. `http://www.vischeck.com/vischeck/vischeckImage.php`, 2006. visited: March, 2006.

Jon Downton and David Gray. AVAZ parameter uncertainty estimation. In *76th Annual International Meeting, SEG, Expanded Abstracts*, pages 234–238, 2006.

Jon Downton, David Gray, and Torre Zuk. Visualizing AVAZ parameter estimates and uncertainty. In *2007 CSPG CSEG Convention, Expanded Abstracts*, pages 234–238, 2007a.

Jon Downton, David Gray, and Torre Zuk. Visualizing AVAZ parameter estimates with uncertainty due to noise. In *EAGE 69th Conference & Exhibition*, 2007b.

Umberto Eco. *Semiotics and the philosophy of language*. Indiana University Press, midland book edition, 1986.

David M. Eddy. Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, and A. Tversky, editors, *Judgment under Uncertainty: Heuristics and biases*, pages 249–267. Cambridge University Press, Cambridge, 1982.

Ward Edwards. Conservatism in human information processing. In D. Kahneman, P. Slovic, and A. Tversky, editors, *Judgment under Uncertainty: Heuristics and biases*, pages 359–369. Cambridge University Press, Cambridge, 1982.

Charles R. Ehlschlaeger, Ashton M. Shortridge, and Michael F. Goodchild. Visualizing spatial data uncertainty using animation. *Computers & Geosciences*, 23(4):387–395, 1997.

Arthur S Elstein and Alan Schwartz. Clinical problem solving and diagnostic decision making: a selective review of the cognitive research literature. In J André Knottnerus, editor, *The Evidence Base of Clinical Diagnosis*, pages 179–195. BMJ Books, 2002.

Beverley J. Evans. Dynamic display of spatial data-reliability: Does it benefit the map user? *Computers & Geosciences*, 23(4):402–422, 1997.

Andrew Fall and Joseph Fall. SELES: A Spatially Explicit Landscape Event Simulator. In *Proc. of GIS and Environmental Modeling*, pages 104–112. National Center for Geographic Information and Analysis, 1996.

Andrew Fall and Joseph Fall. A Domain-Specific Language For Models of Landscape Dynamics. *Ecological Modelling*, 137(1):1–21, February 2001.

Richard Finger and Ann M. Bisantz. Utilizing graphical formats to convey uncertainty in a decision making task. *Theoretical Issues in Ergonomics Science*, 3(1):1–25, 2002.

Baruch Fischhoff. Debiasing. In D. Kahneman, P. Slovic, and A. Tversky, editors, *Judgment under Uncertainty: Heuristics and biases*, pages 422–444. Cambridge University Press, Cambridge, 1982.

Craig W. Fisher and Bruce R. Kingma. Criticality of data quality as exemplified in two disasters. *Information and Management*, 39:109–116, 2001.

Ralph M. Ford. Critical point detection in fluid flow images using dynamical system properties. *Pattern Recognition*, 30(12):1991–2000, 1997.

Maurizio Forte and Alberto Siliotti, editors. *Virtual Archaeology: Re-Creating Ancient Worlds*. Harry N. Abrams, 1997.

John Fox, David Glasspool, and Jonathan Bury. Quantitative and qualitative approaches to reasoning under uncertainty in medical decision making. In *AIME 2001*, volume 2101 of *Lecture Notes in Computer Science*, pages 272–282. Springer, 2001.

Renée C. Fox. The evolution of medical uncertainty. *The Milbank Memorial Fund Quarterly / Health and Society*, 58(1):1–49, 1980.

Bert Freudenberg, Maic Masuch, Niklas Röber, and Thomas Strothotte. The Computer-Visualistik-Raum: veritable and inexpensive presentation of a virtual reconstruction. In *VAST '01: Virtual Reality, Archeology, and Cultural Heritage*, pages 97–102. ACM Press, 2001.

Erich Gamma, Richard Helma, Ralph Johnson, and John Vlissides, editors. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, 1994.

Amit X. Garg, Neill K. J. Adhikari, Heather McDonald, M. Patricia Rosas-Arellano, P. J. Devereaux, Joseph Beyene, Justina Sam, and R. Brian Haynes. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. *JAMA*, 293(10):1223–1238, 2005.

Martha S. Gerrity, Robert F. DeVellis, and Jo Anne Earp. Physicians' reactions to uncertainty in patient care: A new measure and new insights. *Medical Care*, 28(8):724–736, 1990.

Martha S. Gerrity, Jo Anne L. Earp, Robert F. DeVellis, and Donald W. Light. Uncertainty and professional work: Perceptions of physicians in clinical practice. *The American Journal of Sociology*, 97(4):1022–1051, 1992. New Directions in the Sociology of Medicine.

Nahum Gershon. Short note: Visualization of an imperfect world. *IEEE Computer Graphics and Applications*, 18(4):43–45, July/August 1998.

W. A. Ghali and P. M. Sargious. The evolving paradigm of evidence-based medicine. *Journal of Evaluation in Clinical Practice*, 8(2):109–112, 2002.

W. A. Ghali, R. Saitz, P. M. Sargious, and W. Y. Hershman. Evidence-based medicine and the real world: understanding the controversy. *Journal of Evaluation in Clinical Practice*, 5(2):133–138, 1999.

G. Gigerenzer and U. Hoffrage. How to improve bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4):648–704, 1995.

G. Gigerenzer and U. Hoffrage. Overcoming difficulties in bayesian reasoning: A reply to Lewis and Keren (1999) and Mellers and McGraw (1999). *Psychological Review*, 106 (2):425–430, 1999.

Gerd Gigerenzer, Jean Czerlinski, and Laura Marignon. How good are fast and frugal heuristics? In T. Gilovich, D. Griffin, and D. Kahneman, editors, *Heuristics and Biases: The Psychology of Intuitive Judgment*, pages 559–581. Cambridge University Press, Cambridge, 2003.

T. Gilovich, D. Griffin, and D. Kahneman, editors. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, Cambridge, 2003.

W. D. Glanzman. Digging deeper: the results of the first season of activities of the AFSM on the Maḥram Bilqīs, Mārib. In *Proceedings of the Seminar for Arabian Studies*, volume 28, pages 89–104, 1998.

W. D. Glanzman. Clarifying the record: the Bayt Awwām revisited. In *Proceedings of the Seminar for Arabian Studies*, volume 29, pages 73–88, 1999.

W. D. Glanzman. Some notions of sacred space at the Maḥram Bilqīs in Mārib. In *Proceedings of the Seminar for Arabian Studies*, volume 32, pages 187–201, 2002.

Janice Glasgow, N. Hari Narayanan, and B. Chandrasekaran, editors. *Diagrammatic Reasoning: Cognitive and Computational Perspectives*. AAAI Press, 1995.

Bruce Gooch, Peter-Pike J. Sloan, Amy Gooch, Peter Shirley, and Richard Riesenfeld. Interactive technical illustration. In Stephen N. Spencer, editor, *Proceedings of the Conference on the 1999 Symposium on interactive 3D Graphics*, pages 31–38, New York, 1999. ACM Press.

Al Gore. *An Inconvenient Truth: The Planetary Emergency of Global Warming and What We Can Do About It*. Rodale Books, 2006.

H. Griethe and H. Schumann. The visualization of uncertain data: methods and problems. In *Proceedings of SimVis 2006*, pages 143–156. SCS Publishing House, 2006.

D. Griffin and A. Tversky. The weighing of evidence and the determinants of confidence. In T. Gilovich, D. Griffin, and D. Kahneman, editors, *Heuristics and Biases: The Psychology of Intuitive Judgment*, pages 230–249. Cambridge University Press, Cambridge, 2003.

G. Grigoryan and P. Rheingans. Probabilistic surfaces: Point based primitives to show surface uncertainty. In *Proceedings Visualization 2002*. IEEE, 2002. URL `http://citeseer.ist.psu.edu/grigoryan02probabilistic.html`.

G. Grigoryan and P. Rheingans. Point-based probabilistic surfaces to show surface uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 10(5):564–573, 2004.

J Dik F Habbema, René Eijkemans, Pieta Krijnen, and J André Knottnerus. Analysis of data on the accuracy of diagnostic tests. In J André Knottnerus, editor, *The Evidence Base of Clinical Diagnosis*, pages 117–1143. BMJ Books, 2002.

J. Halverston. The first pictures: Perceptual foundations of paleolithic art. *Perception*, 21: 389–404, 1992.

Kenneth M. Hanson, Gregory S. Cunningham, and Robert J. McKee. Uncertainties in Bayesian geometric models. In *Proceedings of the 1999 International Conference on Image Processing (ICIP-99*, pages 41–45, Los Alamitos, CA, October 24–28 1999. IEEE.

Christopher G. Healey. On the use of perceptual cues and data mining for effective visualization of scientific datasets. In *Graphics Interface*, pages 177–184, 1998. URL `http://citeseer.ist.psu.edu/healey98use.html`.

Daryl H. Hepting. Towards a visual interface for information visualization. In *Information Visualisation, 2002. Proceedings. Sixth International Conference on*, pages 295–302. IEEE, 2002.

Richards J. Heuer. *Psychology of Intelligence Analysis*. Central Intelligence Agency, 1999. Center for the Study of Intelligence.

Eric J. Holweg. *Mariner's Guide For Hurricane Awareness In The North Atlantic Basin*. Tropical Analysis and Forecast Branch, Tropical Prediction Center, National Weather Service, National Oceanic and Atmospheric Administration, August 2000.

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

Willaim C. Howell and Sarah A. Burnett. Uncertainty measurement: A cognitive taxonomy. *Organizational Behaviour and Human Performance*, 22:45–68, 1978.

Robert J. B. Hutton. Types of "uncertainty about" and strategies for uncertainty management: An integrative approach. Technical Report 04TA1-SP1-RT1, Klein Associates Inc., 2004.

R. D. Jacobson. Representing spatial information through multimodal interfaces. In *Information Visualisation, 2002. Proceedings. Sixth International Conference on*, pages 730–734, 2002.

Piotr Jankowski, Natalia Andrienko, and Gennady Andrienko. Map-centred exploratory approach to multiple criteria spatial decision making. *International Journal of Geographical Information Science*, 15(2):101–127, 2001.

M. Jenicek. *Foundations of Evidence-Based Medicine*. The Parthenon Publishing Group, New York, 2002.

Bruno Jobard, Gordon Erlebacher, and M. Yousuff Hussaini. Lagrangian-eulerian advection of noise and dye textures for unsteady flow visualization. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):211–222, JUL/SEP 2002.

Chris Johnson. Top scientific visualization research problems. *IEEE Computer Graphics and Applications*, 24(4):13–17, July/August 2004.

Chris R. Johnson and Allen R. Sanderson. A next step: Visualizing errors and uncertainty. *IEEE Computer Graphics and Applications*, 23(5):6–10, September/October 2003. ISSN 0272-1716.

Michael J. Kahn and Amanda Prail. Formal usability inspections. In J. Nielsen and Robert L. Mack, editors, *Usability Inspection Methods*, pages 141–171. John Wiley & Sons, Inc., 1994.

D. Kahneman, P. Slovic, and A. Tversky, editors. *Judgment under Uncertainty: Heuristics and biases*. Cambridge University Press, Cambridge, 1982.

Daniel Kahneman and Dale T. Miller. Norm theory: Comparing reality to its alternatives. In T. Gilovich, D. Griffin, and D. Kahneman, editors, *Heuristics and Biases: The Psychology of Intuitive Judgment*, pages 348–366. Cambridge University Press, Cambridge, 2003.

Daniel Kahneman and Amos Tversky. Variants of uncertainty. *Cognition*, 11(2):143–157, 1982.

David Kao, Jennifer L. Dungan, and Alex Pang. Visualizing 2D probability distributions from EOS satellite image-derived data sets: A case study. In Thomas Ertl, Ken Joy, and Amitabh Varshney, editors, *Proceedings Visualization 2001*, pages 457–460. IEEE Computer Society Technical Committee on Visualization and Graphics Executive Committee, 2001.

J. Kardos, A. Moore, and G. Benwell. The visualisation of uncertainty for spatially referenced census data using hierarchical tessellations. *Transactions in GIS*, 9(1):19–34, 2005.

J. D. Kardos, A. Moore, and G. L. Benwell. Visualizing uncertainty in spatially-referenced attribute data using hierarchical spatial data structures, 2003. URL `http://www.`

geocomputation.org/2003/Papers/Kardos_Paper.pdf. visited: April 2008.

Susan S. Kirschenbaum and James E. Arruda. Effects of graphic and verbal probability information on command decision making. *Human Factors*, 36(3):406–418, 1994.

Paul A. Kirschner, Simon J. Buckingham Shun, and Chad S. Charr, editors. *Visualizating Argumentation: Software Tools for Collaboration and Educational Sense-Making*. Springer, 2003.

Gary Klein. *Sources of Power: How People Make Decisions*. The MIT Press, 1998.

J. André Knottnerus, editor. *The Evidence Base of Clinical Diagnosis*. BMJ Books, 2002.

K. Koffka. *Principles of Gestalt Psychology*. Harcourt-Brace, New York, 1935.

R. Kosara, C. G. Healey, V. Interrante, D. H. Laidlaw, and C. Ware. User studies: Why, how, and when? *IEEE Computer Graphics and Applications*, pages 20–25, July/August 2003.

Robert Kosara, Silvia Miksch, and Helwig Hauser. Semantic depth of field. In *IEEE Symposium on Information Visualization 2001 (InfoVis 2001)*, San Diego, CA, USA, 22–23 2001. URL http://citeseer.ist.psu.edu/kosara01semantic.html.

D. G. Krige. Statistical applications in mine valuation. *Journal of the Institute of Mine Surveyors of South Africa*, 12:45–84, 95–136, 1962.

D.H. Laidlaw, R.M. Kirby, C.D. Jackson, J.S. Davidson, T.S. Miller, M. da Silva, W.H. Warren, and M.J. Tarr. Comparing 2d vector field visualization methods: a user study. *Visualization and Computer Graphics, IEEE Transactions on*, 11(1):59–70, Jan/Feb 2005.

Bongshin Lee, George G. Robertson, Mary Czerwinski, and Cynthia Sims Parr. Candidtree: visualizing structural uncertainty in similar hierarchies. *Information Visualization*, 6:233–246, 2007.

M. Leitner and A. Curtis. A first step towards a framework for presenting the location of confidential point data on maps - results of an empirical perceptual study. *International Journal of Geographical Information Science*, 20(7):813–822, 2006.

Michael Leitner and Barbara P. Buttenfield. Guidelines for the display of attribute certainty. *Cartography and Geographic Information Science*, 27(1):3–14, 2000.

R. Leonhardt, J. Matzka, A.R.L. Nichols, and D.B. Dingwell. Cooling rate correction of paleointensity determination for volcanic glasses by relaxation geospeedometry. *Earth and Planetary Science Letters*, 243:282–292, 2006. Issues 1-2.

Raanan Lipshitz and Orna Strauss. Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behaviour and Human Decision Processes*, 69:149–163, 1997.

M. Lischke. 3DS import library, `http://www.lischke-online.de/3DS.php`, 2005.

Suresh Lodha, Bob Sheehan, Alex Pang, and Craig Wittenbrink. Visualizing geometric uncertainty of surface interpolants. In *Proceedings of Graphics Interface 96*, pages 238–245, 1996a.

Suresh K. Lodha, Alex Pang, Robert E. Sheehan, and Craig M. Wittenbrink. UFLOW: Visualizing uncertainty in fluid flow. In Roni Yagel and Gregory M. Nielson, editors, *IEEE Visualization '96,*, pages 249–254. IEEE, 1996b.

Suresh K. Lodha, Amin P. Charaniya, Nikolai M. Faaland, and S. Ramalingam. Visualization of spatio-temporal gps uncertainty within a gis environment. In *Proceedings of the*

*SPIE Conference on Radar Sense Technology and Data Visualization*, pages 216–227, Orlando, Florida, April 2002a.

Suresh K. Lodha, Nikolai M. Faaland, Amin P. Charaniya, Pramod K. Varshney, Kishan Mehrotra, and Chilukuri Mohan. Visualization of uncertain particle movement. In *Proceedings of the* Computer Graphics and Imaging *Conference*, pages 226–232, Kauii, Hawaii, August 2002b.

A. M. Lopes. *Accuracy in Scientific Visualization*. PhD thesis, The University of Leeds, 1999.

A. L. Love, A. Pang, and D.L. Kao. Visualizing spatial multivalue data. *IEEE Computer Graphics and Applications*, 25(3):69–79, May/June 2005.

Arko Lucieer and Menno-Jan Kraak. Interactive and visual fuzzy classification of remotely sensed imagery for exploration of uncertainty. *International Journal of Geographical Information Science*, 18(5):491–512, 2004.

Arko Lucieer, Alfred Stein, and Peter Fisher. Multivariate texture-based segmentation of remotely sensed imagery for extraction of objects and their uncertainty. *International Journal of Remote Sensing*, 26(14):2917–2936, 2005.

A. M. MacEachren. Visualizing uncertain information. *Cartographic Perspective*, 13: 10–19, 1992.

A. M. MacEachren, C. A. Brewer, and L.W. Pickle. Visualizing georeferenced data: representing reliability of health statistics. *Environment and Planning A*, 30:1547–1561, 1998a.

Alan M. MacEachren. *How Maps Work: Representation, Visualization, and Design*. The Guilford Press, 1995.

Alan. M. MacEachren, Anthony Robinson, Susan Hopper, Steven Gardner, and Robert Murray. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32(3):139–160, 2005.

A.M. MacEachren, F.P. Boscoe, D. Haug, and L.W. Pickle. Geographic visualization: designing manipulable maps for exploring temporally varying georeferenced statistics. In *IEEE Symposium on Information Visualization 1998*, pages 87–94, 156, 1998b.

Jock Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2):110–141, 1986.

Jock D. Mackinlay, Stuart K. Card, and George G. Robertson. Rapid controlled movement through a virtual 3d workspace. In *SIGGRAPH '90: Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, pages 171–176, New York, NY, USA, 1990. ACM.

James R. Mahoney, et. al. Our changing planet: The fiscal year 2003 U.S. global change research program and climate change research initiative. Technical report, Climate Change Science Program and the Subcommittee on Global Change Research, 2003.

Robert Mandel. *Perception, Decision Making and Conflict*. University Press of America, Washington, D.C., 1979.

Jennifer Mankoff, Anind K. Dey, Gary Hsieh, Julie Kientz, Scott Lederer, and Morgan Ames. Heuristic evaluation of ambient displays. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 169–176, New York, NY, USA, 2003. ACM Press. doi: http://doi.acm.org/10.1145/642611.642642.

D. Marr. *Vision*. W.H. Freeman and Company, New York, 1982.

Anthony Masalonis, Sandeep Mulgund, Lixia Song, Craig Wanke, and Stephen Zobell. Using probabilistic demand prediction for traffic flow management decision support. In

*Proceedings of the 2004 AIAA Guidance, Navigation, and Control Conference*. American Institute of Aeronautics and Astronautics, 2004.

Maic Masuch and Thomas Strothotte. Visualising ancient architecture using animating line drawings. In *IV*, pages 261–266, 1998. URL http://citeseer.ist.psu.edu/424263.html.

Margaret Matlin. *Cognition*. Holt, Rinehart and Winston, 1983.

A. Michotte. *The Perception of Causality*. Methuen, London, 1963. Translated by T. Miles and E. Miles.

N. Moray. Monitoring behavior and supervising control. In K.R. Boff, L. Kaufman, and J.P. Thomas, editors, *Handbook of Perception and Human Performance*, volume 2, pages 40–46. Wiley, 1981.

N. Moray and I. Rotenberg. Fault management in process control: Eye movements and action. *Ergonomics*, 32(11):1319–1349, 1989.

E. Morse, M. Lewis, and K. A. Olsen. Evaluating Visualizations: Using a Taxonomic Guide. *Int. J. Human-Computer Studies*, 53:637–662, 2000.

National Academy of Sciences Workshop. Toward Improved Visualization of Uncertain Information, 2005. Sponsored by the Board on Mathematical Sciences and Their Applications. http://www7.nationalacademies.org/bms/1visualization_title_page.html.

Eric Neufeld, Mikelis Bickisb, and Kevin Grantc. Extending the dimensionality of flatland with attribute view probabilistic models. In *Proc. SPIE & IS&T Conf. Electronic Imaging, Vol. 6809: Visualization and Data Analysis 2008*, page 680905, 2008. URL http://link.aip.org/link/?PSI/6809/680905/1.

Allan Newell and Herbert A. Simon. *Human Problem Solving*. Prentice-Hall Inc., New Jersey, 1972.

J. Nielsen. *Usability Engineering*. Academic Press, 1993.

J. Nielsen. Heuristic evaluation. In J. Nielsen and Robert L. Mack, editors, *Usability Inspection Methods*, pages 25–62. John Wiley & Sons, Inc., 1994.

J. Nielsen and Robert L. Mack. *Usability Inspection Methods*. John Wiley & Sons, Inc., 1994.

Donald A. Norman. *Emotional Design: Why We Love (Or Hate) Everyday Things*. Basic Books, 2003.

C. North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3):6–9, 2006.

Laura R. Novick and Miriam Bassok. Problem solving. In Keith J. Holyoak and Robert G. Morrison, editors, *The Cambridge Handbook of Thinking and Reasoning*, pages 321–349. Cambridge University Press, Cambridge, 2005.

Tim O'Sullivan, John Hartley, Danny Saunders, Martin Montgomery, and John Fiske, editors. *Key Concepts in Communication and Cultural Studies*. Routledge, London, second edition, 1994.

P. Palamidese, M. Betro, and G. Muccioli. The virtual restoration of the Visir tomb. In Gregory M. Nielson and Dan Bergeron, editors, *Proceedings of the Visualization '93 Conference*, pages 420–423, San Jose, CA, October 1993. IEEE Computer Society Press.

S. E. Palmer and I. Rock. Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin and Review*, 1(1):29–55, 1994.

A. Pang, J. Furman, and W. Nuss. Data quality issues in visualization. In *In SPIE Proceedings on Visual Data Exploration and Analysis*, volume 2178, pages 12–23, 1994. URL `http://citeseer.ist.psu.edu/pang94data.html`.

Alex T. Pang, Craig M. Wittenbrink, and Suresh K. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, 1997. ISSN 0178-2789.

Philip Peterson, F. David Fracchia, and Brian Hayden. Integrating spatial data display with virtual reconstruction. *IEEE Computer Graphics and Applications*, 15(4):40–46, July 1995.

PIOPED investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism. results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). *JAMA*, 263(20):2753–2759, 1990.

Brandon Plewe. The nature of uncertainty in historical geographic information. *Transactions in GIS*, 6(4):431–456, 2002.

Seth M. Powsner and Edward R. Tufte. Graphical summary of patient status. *Lancet*, 344 (8919):386–, 1994.

Colin Renfrew and Paul Bahn. *Archaeology: Theories Methods and Practice*. Thames & Hudson, 2nd edition, 2000.

P. Rheingans and S. Joshi. Visualization of molecules with positional uncertainty. In E. Gröller, H. Löffelmann, and W. Ribarsky, editors, *Data Visualization '99*, pages 299–306. Springer-Verlag Wien, 1999. ISBN 3-211-83344-7.

P. J. Rhodes, R. S. Laramee, R. D. Bergeron, and T. M. Sparr. Uncertainty Visualization Methods in Isosurface Rendering. In M. Chover, H. Hagen, and D. Tost, editors, *Eurographics 2003, Short Papers/Presentations*, pages 83–88, Granada, Spain, September

1-5 2003. The European Association for Computer Graphics, The Eurographics Association. URL http://www.winslam.com/rlaramee/publication/.

Maria Roussou and George Drettakis. Photorealism and non-photorealism in virtual heritage representation. In *VAST 2003: Virtual Reality, Archaeology and Cultural Heritage*, pages 51–60, 2003.

Pierre-Marie Roy, Isabelle Colombet, Pierre Durieux, Gilles Chatellier, Hervé Sors, and Guy Meyer. Systematic review and meta-analysis of strategies for the diagnosis of suspected pulmonary embolism. *BMJ*, 331:259–268, 2005.

J. E. Russo and L. D. Rosen. An eye fixation analysis of multi-alternative choice. *Memory and Cognition*, 3:267–276, 1975.

C. Rutkowski. An introduction to the human applications standard computer interface, part 1: Theory and principles. *BYTE*, 7(11):291–310, 1982.

Allen R. Sanderson and Chris R. Johnson. Display of vector fields using a reaction-diffusion model. Technical Report UUSCI-2003-002, Scientific Computing and Imaging Institute, University of Utah, June 2003.

Allen R. Sanderson, Chris R. Johnson, and Robert M. Kirby. Display of vector fields using a reaction-diffusion model. In *Visualization, 2004. IEEE*, pages 115–122, 2004.

Ferdinand de Saussure. *Course in General Linguistics*. McGraw-Hill Humanities/Social Sciences/Languages, 1965. first published 1916.

Mike Scaife and Yvonne Rogers. External cognition: how do graphical representations work? *Int. J. Human Computer Studies*, 45:185–213, 1996.

Lothar Schlesier, Josie Hughes, Andrew Fall, and M. Sheelagh T. Carpendale. The

LuMPB Key: A Multiple View Interface to Explore High Dimensional Mountain Pine Beetle Simulation Data. In *Proc. of CMV*, pages 31–41. IEEE Computer Society, 2006.

L. R. Sempé. OpenGL SkyDomes, `http://www.spheregames.com/articles.asp`, 2005.

Jinwook Seo and Ben Shneiderman. Knowledge discovery in high-dimensional data: case studies and a user survey for the rank-by-feature framework. *IEEE Transactions on Visualization and Computer Graphics*, 12(3):311–322, 2006.

Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

Amy L. Shelton and Timothy P. McNamara. Systems of spatial reference in human memory. *Cognitive Psychology*, 43:274–310, 2001.

T. Sheridan. On how often the supervisor should sample. *IEEE Transactions on Systems Man and Cybernetics*, 6:140–145, 1972.

B. Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley, Reading, MA, 1987.

Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, Washington, September 3–6 1996. IEEE Computer Society Press.

Ben Shneiderman and Catherine Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: Novel Evaluation Methods for Information Visualization (BELIV 2006)*, pages 38–43. ACM Press, 2006.

Herbert A. Simon. Rational choice and the structure of the environment. *Psychological Review*, 63(2):129–138, 1956.

Linda J. Skitka, Kathleen L. Mosier, and Mark Burdick. Does automation bias decision-making? *Int. J. Human-Computer Studies*, 51:991–1006, 1999.

Terry A. Slocum, Daniel C. Cliburn, Johannes J. Feddema, and James R. Miller. Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance. *Cartography and Geographic Information Science*, 30(4):299–317, 2003a.

Terry A. Slocum, Robert B McMaster, Fritz C. Kessler, and Hugh H. Howard. *Thematic Cartography and Geographic Visualization*. Prentice Hall, 2nd edition, 2003b.

Steven A. Sloman. Two systems of reasoning. In T. Gilovich, D. Griffin, and D. Kahneman, editors, *Heuristics and Biases: The Psychology of Intuitive Judgment*, pages 379–396. Cambridge University Press, Cambridge, 2003.

Steven A. Sloman, D. Over, Lila Slovak, and J. M. Stibel. Frequency illusions and other fallacies. *Organizational Behaviour and Human Decision Processes*, 91:296–309, 2003.

Paul Slovic, Melissa Finucane, Ellen Peters, and Donald G. MacGregor. The affect heuristic. In T. Gilovich, D. Griffin, and D. Kahneman, editors, *Heuristics and Biases: The Psychology of Intuitive Judgment*, pages 397–420. Cambridge University Press, Cambridge, 2003.

Jared Spool and Will Schroeder. Testing web sites: five users is nowhere near enough. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, pages 285–286, New York, NY, USA, 2001. ACM Press. doi: http://doi.acm.org/10.1145/634067.634236.

Wynn C. Stirling. *Satisficing Games and Decision Making*. Cambridge University Press, Cambridge, UK, 2003.

T. Strothotte and S. Schlechtweg. *Non-Photorealistic Computer Graphics: Modeling, Rendering, and Animation*. Morgan Kaufmann, San Francisco, 2002.

T. Strothotte, B. Preim, A. Raab, J. Schumann, and D. R. Forsey. How to render frames and influence people. *Computer Graphics Forum*, 13(3):455–466, 1994.

Thomas Strothotte, Maic Masuch, and Tobias Isenberg. Visualizing knowledge about virtual reconstructions of ancient architecture. In *Proceedings Computer Graphics International*, pages 36–43. The Computer Graphics Society, IEEE Computer Society, 1999a.

Thomas Strothotte, Matthias Puhle, Maic Masuch, Bert Freudenberg, Sebastian Kreiker, and Babette Ludowici. Visualizing uncertainty in virtual reconstructions. In *Proceedings of Electronic Imaging & the Visual Arts, EVA Europe '99*, Berlin, 1999b. VASARI, GFaI.

Barry N. Taylor and Chris E. Kuyatt. Guidelines for evaluating and expressing the uncertainty of nist measurement results. Technical Report NIST Technical Note 1297, 1994 Edition, National Institute of Standards and Technology, Gaithersburg, MD, 1994.

Alexandru Telea and Jarke J. van Wijk. Simplified representation of vector fields. In *IEEE Visualization '99*, pages 35–42, 1999.

James J. Thomas and Kristin A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, 2005.

J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel. A typology for visualizing uncertainty. In *Proc. SPIE & IS&T Conf. Electronic Imaging, Vol. 5669: Visualization and Data Analysis 2005*, pages 146–157, 2005.

Stefan Timmermans and Alison Angell. Evidence-based medicine, clinical uncertainty, and learning to doctor. *Journal of Health and Social Behavior*, 42(4):342–359, 2001.

Bruce Tognazzini. First principles of interaction design, 2006.

M. Tory and T. Möller. Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 10(1):72–84, 2004.

Melanie Tory and Torsten Möller. Evaluating Visualizations: Do Expert Reviews Work. *IEEE Computer Graphics and Applications*, 25(5):8–11, September/October 2005. doi: http://dx.doi.org/10.1109/MCG.2005.102.

L.A. Treinish. Task-specific visualization design. *Computer Graphics and Applications, IEEE*, 19(5):72–77, 1999.

Edward R. Tufte. *Envisioning Information*. Graphics Press, Cheshire, CT, 1990.

Edward R. Tufte. *Visual Explanations*. Graphics Press, Cheshire, CT, 1997.

Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 2nd edition, 2001.

Edward R. Tufte. *Beautiful Evidence*. Graphics Press, Cheshire, CT, 2006.

Marita Turpin and Niek du Plooy. Decision-making biases and information systems. In *Decision Support in an Uncertain and Complex World: The IFIP TC8/WG8.3 International Conference 2004*, pages 782–792, 2004.

A. Tversky and D. Kahneman. Causal schemas in judgements of uncertainty. In T. Gilovich, D. Griffin, and D. Kahneman, editors, *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, Cambridge, 2003a.

A. Tversky and D. Kahneman. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. In T. Gilovich, D. Griffin, and D. Kahneman, editors, *Heuristics and Biases: The Psychology of Intuitive Judgment*, pages 19–48. Cambridge University Press, Cambridge, 2003b.

R. Twiddy, J. Cavallo, and S. M. Shiri. Restorer: A visualization technique for handling missing data. In R. Daniel Bergeron and Arie E. Kaufman, editors, *Proceedings of the Conference on Visualization*, pages 212–216, Los Alamitos, CA, USA, October 1994. IEEE Computer Society Press.

Eliane R. A. Valiati, Marcelo S. Pimenta, and Carla M.D.S. Freitas. A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: Novel Evaluation Methods for Information Visualization (BELIV 2006)*. ACM Press, 2006.

Jan M. van Bruggen, Henny P. A. Boshuizen, and Paul A. Kirshner. A cognitive framework for cooperative problem solving with argument visualization. In Paul A. Kirshner, Simon J. Buckingham Shum, and Chad S. Carr, editors, *Visualizating Argumentation: Software Tools for Collaboration and Educational Sense-Making*, pages 25–47. Springer, http://www.visualizingargumentation.info/, 2003.

Andries van Dam, Andrew S. Forsberg, David H. Laidlaw, Joseph J. LaViola, Jr., and Rosemary M. Simpson. Immersive VR for scientific visualization: A progress report. *IEEE Computer Graphics and Applications*, 20(6):26–52, November/December 2000.

F.J.M. van der Wel, R.M. Hootsmans, and F. Ormeling. Visualization of data quality. In A.M. MacEacheran and D.R. Fraser Taylor, editors, *Visualization in Modern Cartography*. Pergamon Press, 1994.

Jarke J. van Wijk. Image based flow visualization. *ACM Transactions on Graphics (TOG)*, 21(3):745–754, 2002. doi: http://doi.acm.org/10.1145/566654.566646.

M. Vergauwen, D. Pletinckx, G. Willems, F. Verbiest, L. Van Gool, and Truus Helsen. As time flies by: Mixed image and model-based rendering of an historical landscape from helicopter images. In *VAST 2004: Virtual Reality, Archaeology and Cultural Heritage*, pages 241–250, 2004.

Colin Ware. Color sequences for univariate maps: Theory, experiments, and principles. *IEEE Computer Graphics and Applications*, 8(5):41–49, September 1988.

Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers, 2nd edition, 2004.

Evan T. Watkins. Improving the analyst and decision-maker's perspective through uncertainty visualization. Master's thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 2000.

Andrew J. Weaver. The science of climate change. *Geoscience Canada*, 30(3):91–109, 2003.

S. Wehrend and C. Lewis. A problem-oriented classification of visualization technique. In *IEEE Visualization '90*, pages 139–143. IEEE Computer Society, 1990.

Milton C. Weinstein and Harvey V. Fineberg. *Clinical decision analysis*. W. B. Saunders Company, 1980.

P. S. Wells, D. R. Anderson, M Rodger, J. S. Ginsberg, C Kearon, M Gent, , A. G. Turpie, J Bormanis, J Weitz, M Chamberlain, D Bowie, D Barnes, and J Hirsh. Derivation of a simple clinical model to categorize patients probability of pulmonary embolism: increasing the models utility with the simplired d-dimer. *Thromb Haemost*, 83:416–420, 2000.

P. S. Wells, D. R. Anderson, M Rodger, Ian Stiell, J. F. Dreyer, D. Barnes, M. Forgie, George Kovacs, John Ward, and M. J. Kovacs. Excluding pulmonary embolism at the bedside without diagnostic imaging: Management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and d-dimer. *Ann Intern Med*, 135:98–107, 2001.

C. Wharton, J. Rieman, C. Lewis, and P. Polson. The cognitive walkthrough method: A practioner's guide. In J. Nielsen and R. Mack, editors, *Usability Inspection Methods*, chapter 5, pages 105–140. Wiley and Sons, 1994.

C. D. Wickens. *Engineering Psychology and Human Performance*. Harper Collins, New York, U.S.A., 2nd edition, 1992.

J. H. Wilkinson. *Rounding Errors in Algebraic Processes*. Dover Publications Inc., 1994.

Robert L. Winkler and James E. Smith. On uncertainty in medical testing. *Medical Decision Making*, 24:654–658, 2004.

Craig M. Wittenbrink. IFS fractal interpolation for 2d and 3d visualization. In *IEEE Visualization*, pages 77–, 1995. URL `http://citeseer.ist.psu.edu/132579.html`.

Craig M. Wittenbrink, Alex T. Pang, and Suresh K. Lodha. Glyphs for visualizing uncertainty in vector fields. *IEEE Transactions on Visualization and Computer Graphics*, 2 (3):266–279, 1996.

Pak Chung Wong, Harlan Foote, Ruby Leung, Elizabeth Jurrus, Dan Adams, and Jim Thomas. Vector fields simplification – a case study of visualizing climate modeling and simulation data sets. In *IEEE Visualization '00*, pages 485–488, 2000.

Geoff Wyvill and Brian Wyvill. Illustrating uncertainty. In *ACM Siggraph 2000 Sketches & Applications*, page 214. ACM, 2000.

J. Zhang and M. Goodchild. *Uncertainty in Geographical Information*. Taylor & Francis Inc., New York, 2002.

T. Zuk and M. S. T. Carpendale. Theoretical analysis of uncertainty visualizations. In Robert F. Erbacher, Jonathan C. Roberts, Matti T. Gröhn, and Katy Börner, editors, *Proc.*

*SPIE & IS&T Conf. Electronic Imaging, Vol. 6060: Visualization and Data Analysis 2006*, page 606007. SPIE, 2006. URL http://link.aip.org/link/?PSI/6060/606007/1.

T. Zuk, M. S. T. Carpendale, and W.D. Glanzman. Visualizing temporal uncertainty in 3D virtual reconstructions. In *Proc. of the 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2005)*, pages 99–106, 2005.

Torre Zuk and Sheelagh Carpendale. Interactive simulation and visualization using the GPU. In Robert F. Erbacher, Jonathan C. Roberts, Matti T. Gröhn, and Katy Börner, editors, *Proc. SPIE & IS&T Conf. Electronic Imaging, Vol. 5669: Visualization and Data Analysis 2005*, pages 262–273. SPIE, 2005. URL http://link.aip.org/link/?PSI/5669/262/1.

Torre Zuk and Sheelagh Carpendale. Visualization of uncertainty and reasoning. In Andreas Butz, Brian Fisher, Antonio Krüger, Patrick Olivier, and Shigeru Owada, editors, *Proceedings of the 7th International Symposium on Smart Graphics (June 25-27, 2007, Kyoto, Japan)*, volume 4569 of *Lecture Notes in Computer Science*, pages 164–177, Berlin, Heidelberg, 2007. Springer-Verlag. URL http://dx.doi.org/10.1007/978-3-540-73214-3_15.

Torre Zuk, Lothar Schlesier, Petra Neumann, Mark S. Hancock, and Sheelagh Carpendale. Heuristics for information visualization evaluation. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: Novel Evaluation Methods for Information Visualization (BELIV 2006)*, pages 55–60. ACM Press, 2006.

Torre Zuk, Jon Downton, David Gray, Sheelagh Carpendale, and J.D. Liang. Exploration of uncertainty in bidirectional vector fields. In Katy Börner, Matti T. Gröhn, Jinah Park, and Jonathan C. Roberts, editors, *Proc. SPIE & IS&T Conf. Electronic Imaging, Vol.*

*6809: Visualization and Data Analysis 2008*, volume 6809, page 68090B. SPIE, 2008. URL http://link.aip.org/link/?PSI/6809/68090B/1.

# Appendix A: Ethics Approval

FACULTY OF MEDICINE UNIVERSITY OF CALGARY

2005-08-31

OFFICE OF MEDICAL BIOETHICS
Room 93, Heritage Medical Research Bldg
3330 Hospital Drive NW
Calgary, AB, Canada T2N 4N1
Telephone: (403) 220-7990
Fax: (403) 283-8524
Email: omb@ucalgary.ca

Dr. S. Carpendale
Department of Computer Science
University of Calgary
Calgary, Alberta

**Dear Dr. Carpendale:**

**RE: Visualization of Uncertainty for the Diagnosis of Pulmonary Embolism in an Evidence-Based Medicine Decision Making Framework**

**Grant ID: 18657**
**PhD Student: Zuk, T.**

The above-noted thesis proposal (Version dated July 2005), and the Consent Form (Revised Version dated August 15, 2005) have been submitted for Committee review and found to be ethically acceptable.

Please note that this approval is subject to the following conditions:
(1) access to personal identifiable health information was not requested in this submission;
(2) a copy of the informed consent form must have been given to each research subject, if required for this study;
(3) a Progress Report must be submitted by 2006-08-31, containing the following information:
  i) the number of subjects recruited;
  ii) a description of any protocol modification;
  iii) any unusual and/or severe complications, adverse events or unanticipated problems involving risks to subjects or others, withdrawal of subjects from the research, or complaints about the research;
  iv) a summary of any recent literature, finding, or other relevant information, especially information about risks associated with the research;
  v) a copy of the current informed consent form;
  vi) the expected date of termination of this project.
(4) a Final Report must be submitted at the termination of the project.

Please note that you have been named as a principal collaborator on this study because students are not permitted to serve as principal investigators. Please accept the Board's best wishes for success in your research.

Yours sincerely

Christopher J. Doig, MD, MSc, FRCPC

Chair, Conjoint Health Research Ethics Board

CJD/km

c.c. Adult Research Committee Dr. K. Barker (information) Research Services Mr. T. Zuk (PhD Student)
Office of Information and Privacy Commissioner

CREATING THE FUTURE OF HEALTH An innovative medical schoo xcelle nd le d service

**Foothills Medical Centre**
1403 29 Street NW
Calgary, Alberta, Canada T2N 2T9
website www.calgaryhealthregion.ca

calgary health region
**Foothills Medical Centre**

09 September 2005

Dr. Sheelagh Carpendale
Department of Computer Sciences
University of Calgary

Dear Dr. Carpendale:

**Re: #18657 – Visualization of Uncertainty for the Diagnosis of Pulmonary Embolism in an Evidence-Based Medicine Decision Making Framework**

Thank you for submitting an application regarding the above project for review by the Adult Research Committee of the Calgary Health Region (CHR). This will confirm that the committee has granted institutional approval for this project, **contingent on approval by the Conjoint Health Research Ethics Board**.

It is understood from your submission that your study will be entirely funded through external sources and that the CHR will be reimbursed for all research costs associated with this project if applicable. **To facilitate a smooth startup of your project, please notify affected departments in the Region well in advance of your intent to initiate this study.**

Please accept the committee's best wishes for success in your research.

Yours sincerely,

Elizabeth MacKay, MD
Acting Chair, Adult Research Committee

cc: Dr. K. Barker, Conjoint Health Research Ethics Board, Mr. T. Zuk

# Appendix B: Study Materials Related to Chapter 7

UNIVERSITY OF CALGARY

Department of Computer Science
University of Calgary
2500 University Drive
Calgary, AB, CANADA T2N 1N4

## CONSENT FORM FOR STUDY PARTICIPANTS

**Research Project Title**: Visualization of uncertainty for the diagnosis of pulmonary embolism in an evidence-based medicine decision making framework

**Investigators**: Torre Zuk, Dr. Sheelagh Carpendale (PI) and Dr. William Ghali (PI)

This consent form, a copy of which has been given to you, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, you should feel free to ask. Please take the time to read this carefully and to understand any accompanying information.

### Description of Research Project

Pulmonary embolism is a very challenging medical condition to diagnose, because its detection is typically accomplished through the use of non-invasive diagnostic tests that have imperfect sensitivity and specificity. Furthermore, test results are interpreted in concert with clinical estimates of probability of disease that clinicians implicitly or explicitly combine with test results to judge whether pulmonary embolism is present or absent. Inherent in this process is the consideration of uncertainty in final diagnostic decisions. There are currently existing computer-based tools in use in the Calgary Health Region (on the regional hospital TDS system) that are designed to assist clinicians in the difficult process of accurately diagnosing pulmonary embolism. Anecdotally, however, those tools have several limitations that limit their use in clinical settings.

In this proposed research, we plan to begin with a formal user and task analysis of the existing computer-based tool, to assess providers' views of the existing tool. This will then be followed by the iterative collaborative development of an improved computer-based diagnostic tool. The goal being to develop a new diagnostic aid that can help clinicians to better visualize the uncertainty associated with diagnostic decision-making for pulmonary embolism, and that may also help clinicians make more appropriate clinical decisions for their patients with pulmonary embolism.

### Participant's Involvement

This research will be carried out in three distinct phases. Participation in phase one trials will involve one-on-one observation and discussion of the participants' use of the TDS (or other methodology) for a simulated patient. Involvement in the second and third phase will involve written feedback and group discussion of prototypical, and the final, visualization tools, respectively. Written components will be used to determine participants' demographics such as experience level. In each phase the time commitment for each participant will be under one hour, and the participant will only be expected to take part in one phase.

Participation in this study will not put you at any risk or harm and is strictly voluntary. All information regarding your personal information and those that could identify how you performed is confidential: only the researchers involved will have access to it. Participants will receive remuneration in the form of a bookstore gift certificate.

At the conclusion of the study and its analysis, we will present our findings (participant anonymity will be maintained in all reports and publications) in a debriefing session. You will also be given opportunities to ask questions about the study and the findings.

**Participant's Consent**

Your signature on this form indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a subject. In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from the study at any time. Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation. If you have further questions concerning matters related to this research, please contact:

Torre Zuk
zuk@cpsc.ucalgary.ca
Department of Computer Science, University of Calgary

If you have any questions concerning your rights as a possible participant in this research, please contact Pat Evans, Associate Director, Internal Awards, Research Services, University of Calgary, at 220-3782.

| | **Please circle one** | **Please initial your choice** |
|---|---|---|
| I agree to participate in the activities explained above | YES \| NO | |
| I agree to be audio taped for transcription purposes only | YES \| NO | |
| I agree to let my conversation during the study be directly quoted, anonymously, in presentation of the research results | YES \| NO | |

Participants Signature Date

Investigator(s) and/or Delegate's Signature Date

Witness' Signature Date

The University of Calgary Conjoint Health Research Ethics Board has approved this research study. A copy of this consent form has been given to you to keep for your records and reference.

UNIVERSITY OF CALGARY

Department of Computer Science
University of Calgary
2500 University Drive
Calgary, AB, CANADA T2N 1N4

**Visualization of uncertainty for the diagnosis of pulmonary embolism in an evidence-based medicine decision making framework**

**Clinical Case for Diagnostic Testing**

Patient A – Pathfinder, TorreA

- 52 year old Caucasian woman
- Height: 170 cm
- Weight: 61 kg
- Heart rate 98 bpm
- Temp 37.5C
- Feeling weak and short of breath
- Tinges of blood in sputum
- No recent medical problems or chronic condition (No previous DVT/PE)

Department of Computer Science
University of Calgary
2500 University Drive
Calgary, AB, CANADA T2N 1N4

**Visualization of uncertainty for the diagnosis of pulmonary embolism in an evidence-based medicine decision making framework**

**Clinical Case for Diagnostic Testing**

Patient B – Pathfinder, TorreB

- 46 year old Caucasian male
- Height: 194 cm
- Weight: 93 kg
- Heart rate 105 bpm
- Temp 38C
- Short of breath, right leg is swollen with pain on palpation
- Pleuritic chest pain
- No recent medical problems or chronic condition (No previous DVT/PE)

UNIVERSITY OF CALGARY

Department of Computer Science
University of Calgary
2500 University Drive
Calgary, AB, CANADA T2N 1N4

**Visualization of uncertainty for the diagnosis of pulmonary embolism in an evidence-based medicine decision making framework**

**Study Questionnaire**

Please circle the number that best describes your agreement with each statement.

A ) When using computers I am comfortable exploring features or options .

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

Comments: ______________________________

B ) I am confident in the system recommendations for ordering a diagnostic test.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

Comments: ______________________________

C ) The current TDS /OSCAR system helps me practice evidence-based medicine.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

Comments: ______________________________

D ) I am confident in my application of evidence-based medicine .

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

Comments: ______________________________

E ) Decision support and test ordering should be integrated into one system.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

Comments:______________________________

Phase I: Task Analysis Torre Zuk, Dept. Computer Science, zuk@cpsc.ucalgary.ca

**Visualization of uncertainty for the diagnosis of pulmonary embolism in an evidence-based medicine decision making framework.**

Participant ID ____________ Date____________

A. Based on task and use of system

1. How would you describe your interpretation of the Wells Score question: "(PE is) as or more likely than other diagnosis?"
2. Were you equally confident about all of your answers to Well Score questions?
3. Did you think about probabilities explicitly as a number during the process?
4. Did you only want to order a test when using the system?

B. Problem domain and use of system

1. What would make you more confident in a system recommendation (e.g. Post Well Score V-Q Scan recommendation of high, medium, low)?
2. How would you report the confidence in the diagnosis (so far) to the patient?
3. Do you feel the system helps you practice evidence-based medicine?
4. Have you used the diagnostic tree display?

Phase I: Task Analysis Torre Zuk, Dept. Computer Science, zuk@cpsc.ucalgary.ca

5. Do you read any additional information provided about tests or do you have it memorized?

6. Do you use the TDS to share information for consulting others?

C. General characteristics and ideas
1. How familiar are you with evidence-based medicine (how long practicing)?

2. How many years have you used computers?

3. Have you used software related to evidence-based medicine?

4. What would like to change about the TDS?

5. What other information would you like to see to improve a new system?

6. Where would you prefer to use this system? (i.e. current stations, bedside, home…)

# Appendix C: Study Materials Related to Chapter 8

UNIVERSITY OF CALGARY

Departments of Computer Science and Medicine

**Study Protocol: Visualization of uncertainty for the diagnosis of pulmonary embolism in an evidence-based medicine decision making framework**

**Informed Consent for Participation: Approved by Research Services Office, University of Calgary**

The following questions are part of the iterative design process for visualizations relating to the research of uncertainty visualization. Your input on this process is greatly appreciated. Our plan is to use your responses to inform future versions.

Filling out this form indicates that you have understood your participation will be anonymous and that you agree to participate, with the only aspect of participation being this questionnaire. In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal professional responsibilities. You are free to not answer specific items or questions. If you have any questions or concerns about the way you've been treated as a participant, you can contact the researchers (T. Zuk, S. Carpendale, W. Ghali). If you have any questions concerning your rights as a possible participant in this research, please contact Pat Evans, Associate Director, Internal Awards, Research Services, University of Calgary, at 220-3782

**<u>Visual Evidence Feedback</u>**

| | Conditions | Consider | Rank | Probability |
|---|---|---|---|---|
| 1 | Pulmonary Embolism | Y | 1 | .2 |
| 2 | Pneumonia, Bacterial | Y | 2 | |
| 3 | Myocardial Infarction | Y | | |
| 4 | Superior Vena Cava Syndrome | Y | | |
| 5 | Acute Coronary Syndrome | | | |
| 6 | Acute Respiratory Distress Syndrome | | | |
| 7 | Altitude Illness - Pulmonary Syndromes | | | |
| 8 | Anemia, Acute | | | |

**Visual 1. Differential diagnosis considerations and prioritizing.**

**Please check the box that best describes your agreement with each statement:**

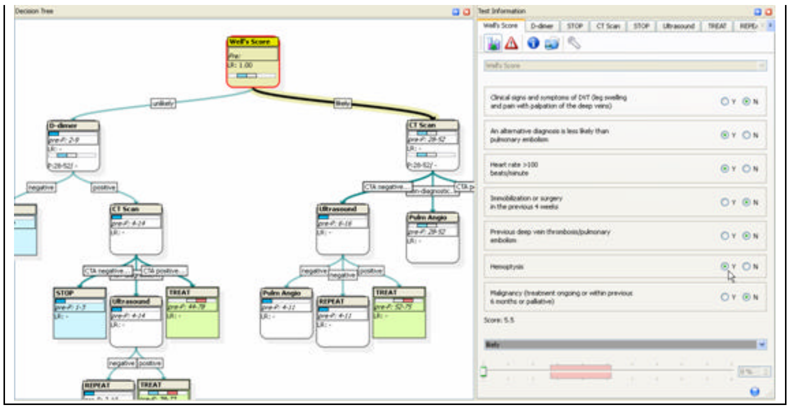
V1.A ) This cognitive support for differential diagnosis would be useful.

| **strongly disagree** | **disagree** | **undecided** | **agree** | **strongly agree** |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V1.B ) I would enter prioritizing information on conditions based on: **(check all that apply)**

| **Y/N** | **Rank** | **Probability** |
|---|---|---|
| ☐ | ☐ | ☐ |

V1.C ) Comments:




**Visual 2. Context visual for recommendation sensitivity to Wells scoring.**

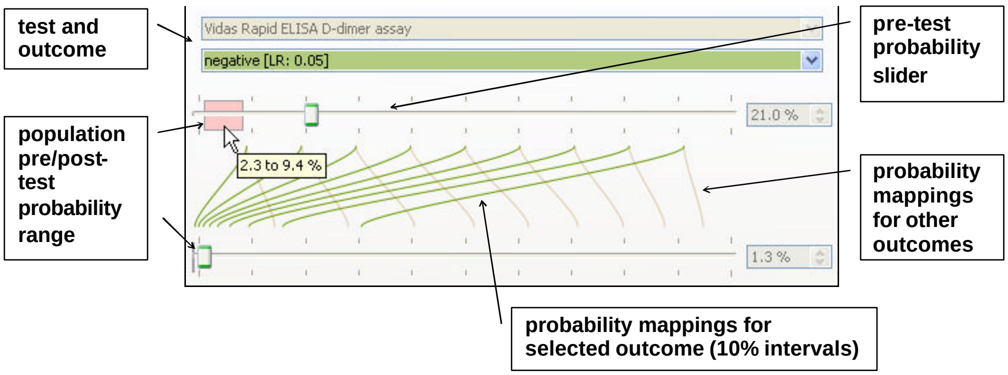
V2.A ) Visual evidence of recommendation sensitivity to Wells scoring would be useful.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V2.B ) The visual context of decision pathways would be useful when viewing other information (e.g. warnings, references, probability functions, …).

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V2.C ) Comments:




**Visual 3. Pre-post test probability slider.**

V3.A ) Seeing the relationship between pre and post probabilities is useful.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

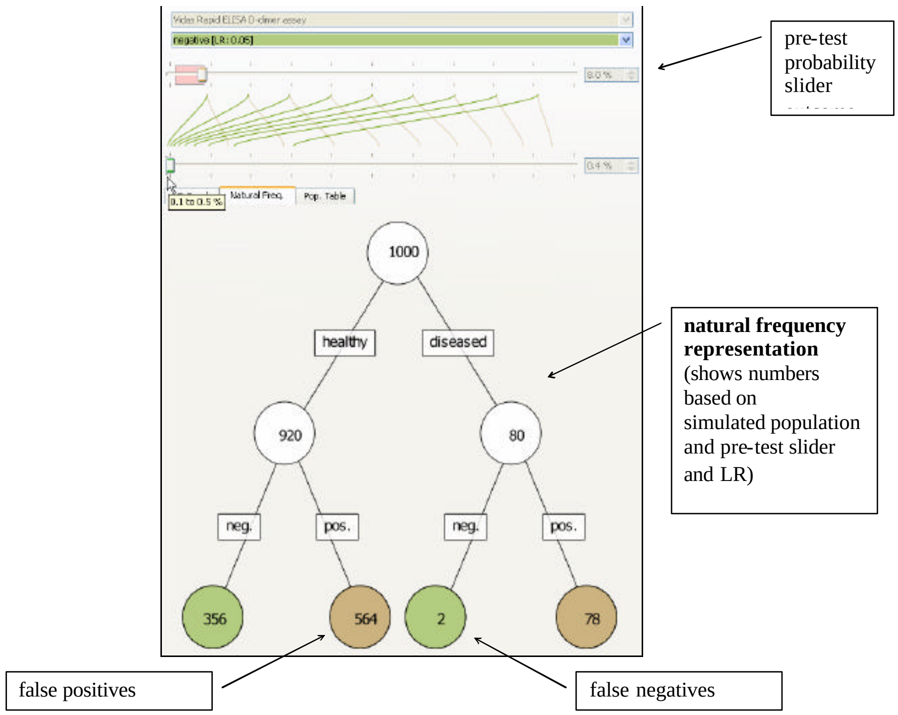
V3.B ) This visualization showing pre and post probabilities is comprehensible.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V3.C ) This visualization would assist my interpretation of the test results.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V3.D ) Comments:




**Visual 4. Natural frequency nested-set visual.**

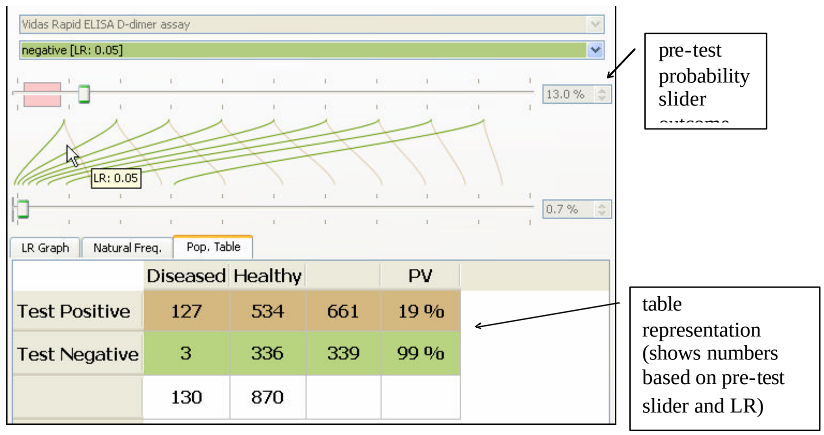
V4.A ) This natural frequency representation is useful.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V4.B ) This natural frequency representation would assist my interpretation of the test results.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V4.C ) Comments:




**Visual 5. Outcome table using pre-test slider and LR of selected outcome**

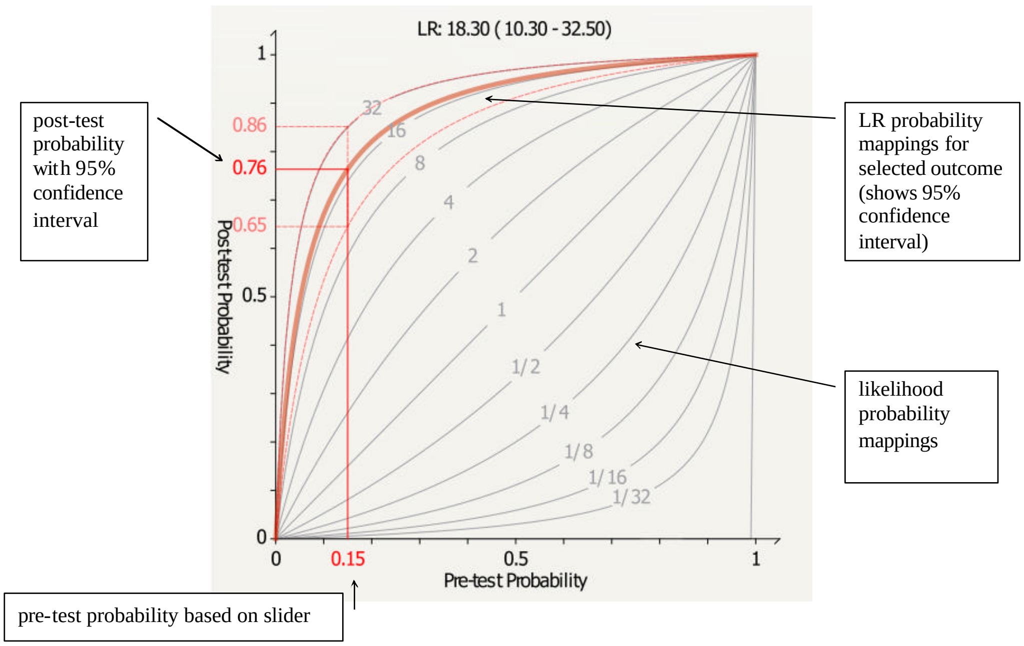
V5.A ) This outcome table would be useful.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V5.B ) This outcome table would assist my interpretation of the test results.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V5.C ) Comments:




**Visual 6. Likelihood and probability graph (V/Q Scan Negative outcome).**

V6.A ) This likelihood and probability graph would be useful.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

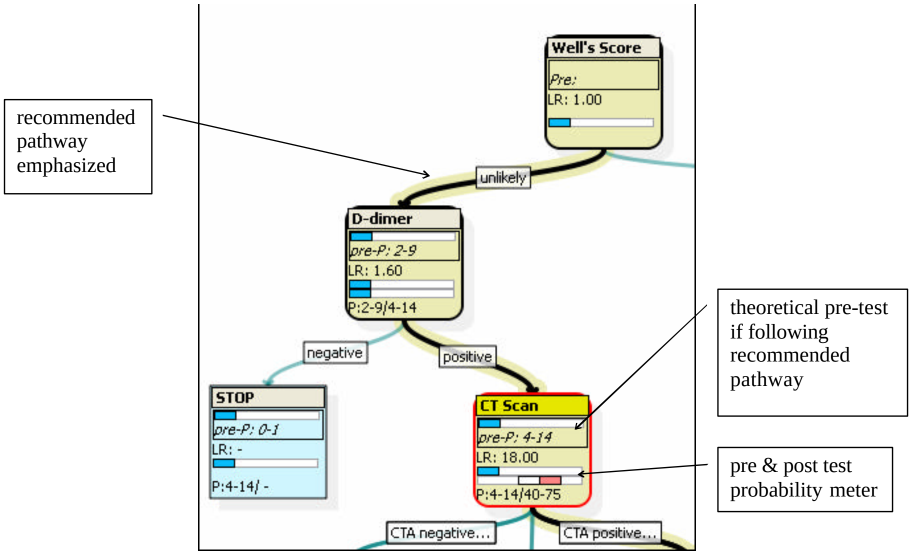
V6.B ) This likelihood and probability graph would improve my interpretation of the test results.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V6.C ) The uncertainty aspects (confidence intervals) would assist my interpretation of the test results.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V6.D ) Comments:




**Visual 7. Pre/post-test probabilities integrated in decision pathway.**

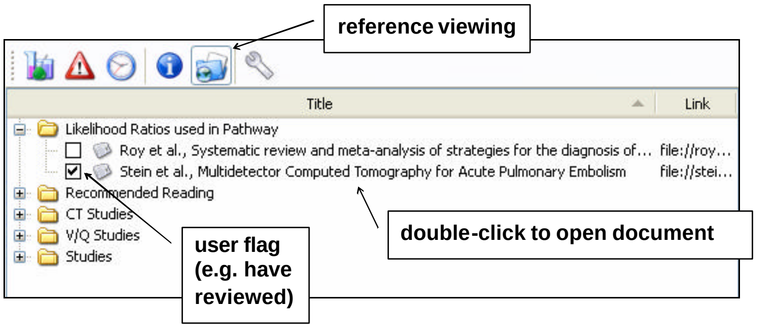
V7.A ) Pre/post-test probabilities integrated in the visual decision pathway would be useful.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V7.B ) Pre/post-test probabilities integrated in the visual decision pathway add confidence to system recommendations.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V7.C ) Comments:




**Visual 8. Group reviewed reference repository and search**

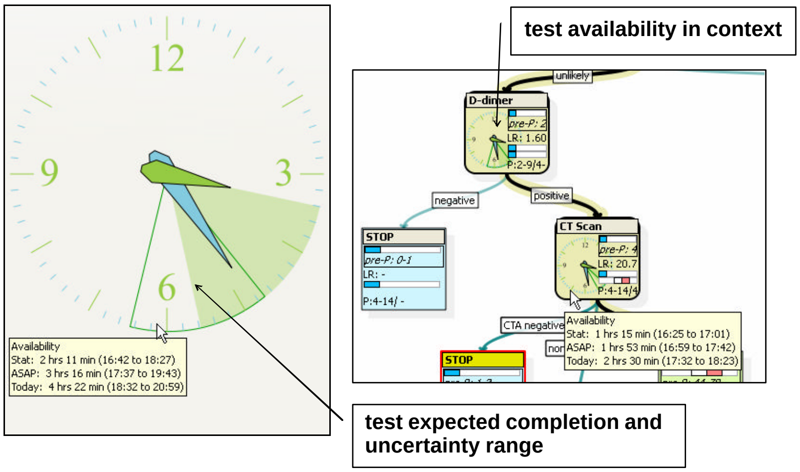
V8.A ) Integrated references would increase confidence in hospital decision recommendations.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V8.B ) This would assist me in ensuring I have read the latest evidence.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V8.C ) Comments:

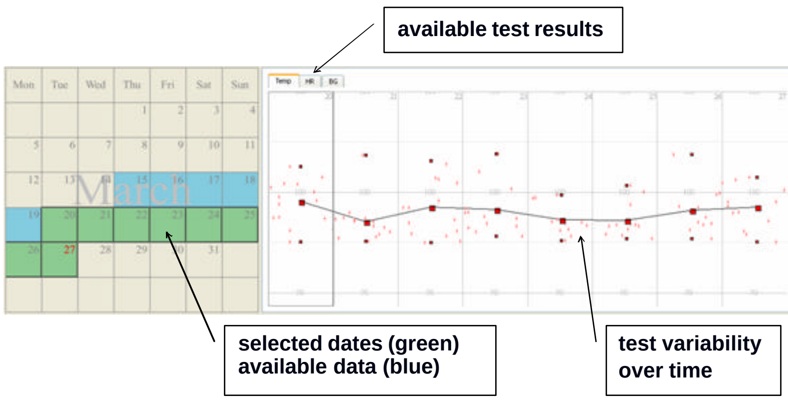



**Visual 9. Test availability uncertainty.**

V9.A ) This test availability and uncertainty visualization would be useful in planning.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V9.B ) Comments:




**Visual 10. Test result variability & precision uncertainty**.

V10.A ) This test result variability and precision visualization would be useful.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

V10.B ) Comments:

**<u>Demographic and General Questions</u>**

A ) Your sex:

| male | female |
|---|---|
| ☐ | ☐ |

B ) Your age:

| under 20 | 20-29 | 30-39 | 40-49 | 50 and over |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

C ) Years practicing evidence-based diagnosis:

| under 1 | 1-2 | 3-5 | 6-10 | 11 and more |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

D ) Do you use computer support or visual aids in making evidence-based decisions?

| never | once or twice | rarely | sometimes | always |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

**Please check the box that best describes your agreement with each statement:**

F ) I would like more computer support for evidence-based decision making.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

G ) I don't feel confident in computer systems advising me on decisions.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

H ) I would like clear visual evidence and be free to make my own decisions.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

I ) I am happy with how I manage uncertainty in making diagnostic decisions.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

J ) I am confident with my ability to apply evidence in my decision making.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

K ) My preference to get decision support would be using: ( **order from 1 to 3** )

| hand-held device | any shared computer | high-end computer (multiple displays) |
|---|---|---|
| ☐ | ☐ | ☐ |

General Comments:

**Overall System Feedback**

A ) Visual evidence like this would be useful for education and training.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

B ) Visual evidence like this would be useful in practice.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

C ) Visualizations of uncertainty like this set would increase my confidence in decisions.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

D ) I would use visual evidence like this set of visualizations if it was available.

| strongly disagree | disagree | undecided | agree | strongly agree |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

E ) Information that you feel was not present that would be useful:

F ) Other general comments:

# Appendix D: Permission for Use of Previous Publications

UNIVERSITY OF CALGARY

University of Calgary
Department of Computer Science
2500 University Drive
Calgary, AB, CANADA T2N 1N4

**Permission for the Use of**

T. Zuk and S. Carpendale, **Theoretical Analysis of Uncertainty Visualizations,** Visualization and Data Analysis 2006, edited by Robert F. Erbacher, Jonathan C. Roberts, Matti T. Gröhn, Katy Börner, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6060, 606007, 2006.

In signing this form I give Torre Zuk permission to use work from the co-authored paper listed above in his dissertation and to have this work microfilmed.

Sheelagh Carpendale

Oct 23, 2007
Date

UNIVERSITY OF CALGARY

University of Calgary
Department of Computer Science
2500 University Drive
Calgary, AB, CANADA T2N 1N4

**Permission for the Use of**

Torre Zuk, Lothar Schlesier, Petra Neumann, Mark S. Hancock, and M. Sheelagh T. Carpendale, **Heuristics for Information Visualization Evaluation**, In Proceedings of the Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV 2006), held in conjunction with the Working Conference on Advanced Visual Interfaces (AVI 2006, May 23–26, 2006, Venice, Italy), New York, NY, USA.

In signing this form I give Torre Zuk permission to use work from the co-authored paper listed above in his dissertation and to have this work microfilmed.

Lothar Schlesier
Lothar Schlesier

Oct. 4th, 2007
Date

Petra Neumann
Petra Neumann

Oct 23rd, 2007
Date

Mark Hancock

Oct 23 2007
Date

Sheelagh Carpendale

Oct 23, 2007
Date

University of Calgary
Department of Computer Science
2500 University Drive
Calgary, AB, CANADA T2N 1N4

**Permission for the Use of**

Torre Zuk and Sheelagh Carpendale. **Visualization of Uncertainty and Reasoning**, In Proceedings of the 7th International Symposium on Smart Graphics (June 25-27, 2007, Kyoto, Japan), Berlin, Heidelberg. (Andreas Butz and Brian Fisher and Antonio Krüger and Patrick Olivier and Shigeru Owada, Eds.) Springer-Verlag, pages 164-177, 2007

In signing this form I give Torre Zuk permission to use work from the co-authored paper listed above in his dissertation and to have this work microfilmed.

Sheelagh Carpendale

Oct 23, 2007
Date

UNIVERSITY OF CALGARY

University of Calgary
Department of Computer Science
2500 University Drive
Calgary, AB, CANADA T2N 1N4

**Permission for the Use of**

T. Zuk, M. S. T. Carpendale, and W. D. Glanzman. **Visualizing temporal uncertainty in 3D virtual reconstructions.** *In Proc. of the 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage* (VAST 2005), pages 99–106, 2005.

In signing this form I give Torre Zuk permission to use work from the co-authored paper listed above in his dissertation and to have this work microfilmed.

Sheelagh Carpendale — Date: Oct 23, 2007

William D. Glanzman — Date: 12 Jun 2007

University of Calgary
Department of Computer Science
2500 University Drive
Calgary, AB, CANADA T2N 1N4

**Permission for the Use of**

Torre Zuk, Jon Downton, David Gray, Sheelagh Carpendale and JD Liang. **Exploration of uncertainty in bidirectional vector fields**, Visualization and Data Analysis 2008, edited by Katy Börner, Matti T. Gröhn, Jinah Park, and Jonathan C. Roberts. Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6060, 2008, To Appear.

In signing this form I give Torre Zuk permission to use work from the co-authored paper listed above in his dissertation and to have this work microfilmed.

| | |
|---|---|
| Jon Downton | Oct 1, 2007<br>Date |
| David Gray | 2-oct-07<br>Date |
| Sheelagh Carpendale | Oct 23 2007<br>Date |
| JD Liang | Oct 1, 2007<br>Date |

UNIVERSITY OF CALGARY

University of Calgary
Department of Computer Science
2500 University Drive
Calgary, AB, CANADA T2N 1N4

**Permission for the Use of**

Torre Zuk, M. Sheelagh T. Carpendale, William Ghali, and Barry Baylis. **Unpublished works.** Submitted, or to be submitted; all resulting from research under the protocol: Visualization of Uncertainty for the Diagnosis of Pulmonary Embolism in an Evidence-Based Medicine Decision Making Framework.

In signing this form I give Torre Zuk permission to use work from the co-authored works listed above in his dissertation and to have this work microfilmed.

Sheelagh Carpendale — Oct 23, 2007 Date

William Ghali — Oct 23, 2007 Date

Barry Baylis — Oct 23, 2007. Date