HCC: CGV: Small: Modeling User Strategies for Scientific Visualization Evaluation

Caroline Ziemkiewicz (PI) and David Laidlaw (Co-PI), Brown University Department of Computer Science

Project Summary

The objective of this research is to model how individual differences influence user strategies in scientific data exploration. The researchers propose to model these strategies and study their relationship to individual cognitive differences, as well as how they affect or are affected by the use of visualization systems. The current data analysis needs of collaborators in biomedical research fields will guide this research. These users must manage and understand large amounts of complex data. Finding patterns in such data, and establishing confidence in those patterns, is a complex cognitive task that can be approached in multiple ways. As a result, designing and evaluating new visualization tools for these tasks is challenging without a more complete understanding of the strategies users employ. Preliminary observations of these users have shown evidence for at least two unique strategies that are reflected in preferences for different visualization designs.

The investigators will approach this problem from two directions: by performing a qualitative task analysis of scientific data exploration, and by building low-level cognitive models of visualization tasks as performed by different user types. Interviews with domain scientists, along with video and interaction data from recorded visualization sessions, will be collected. Using video coding and interaction analysis, these data will be analyzed for common patterns in behavior that correspond to user strategies. At the same time, the researchers will leverage existing work in individual differences in visualization to determine whether existing cognitive and psychological factors can predict which strategies a user tends to adopt. These findings will inform a set of basic cognitive models of how users gather and compare information from a scientific visualization interface. These models will take into account the user's strategy and factors of a visualization design to predict performance in terms of response time and accuracy. We will validate this model by testing its predictions against actual user performance with varying visualization designs through large-scale user studies. The investigators are well-positioned to execute this work, combining expertise in scientific visualization and cognitive aspects of visualization.

Intellectual Merits. Individual differences, along with the complex nature of visualization tasks, are a possible source of many conflicting evaluation results that prevent clear comparisons between visualization techniques. Advancing this understanding would improve the interpretation of evaluation results, deepen the cognitive theory of visualization, and provide new metrics by which to measure visualization designs. Cognitive modeling has not previously been applied to the question of individual differences in visualization, and the framing of user behavior in terms of strategy types has not been explored in depth. Finally, the proposed research will demonstrate and validate a novel methodology for evaluating visualizations.

Broader Impacts. The proposed research is highly user-centric and will involve frequent collaboration with scientists in biomedical fields. These collaborations will improve our user model through close interaction with real users, and can also be expected to directly improve the design of visualizations used by these collaborators. In addition, we will establish an online data repository which will include anonymized versions of our interaction data and open-source releases of the cognitive models we develop, and will be open to contributions from other researchers. In this way, we will both disseminate our results and create a starting point for further research in this area.

Keywords: Scientific visualization, individual differences, user modeling, task analysis

HCC: CGV: Small: Modeling User Strategies for Scientific Visualization Evaluation

Caroline Ziemkiewicz (PI) and David Laidlaw (Co-PI), Brown University Department of Computer Science

1 Overview and Objectives

The objective of this research is to model how individual differences influence user strategies in scientific data exploration. The researchers propose to model these strategies and study their relationship to individual cognitive differences, as well as how they affect or are affected by the use of visualization systems. The current data analysis needs of collaborators in the immunology, proteomics, and brain science fields will guide this research. These users manage large amounts of complex data, and this amount is growing far larger with advances in experimental techniques.

Finding patterns in such large data, and establishing confidence in those patterns, is a complex cognitive task that can be approached in multiple ways. Previous user studies by the investigator and others have suggested that when visualization tasks are more complicated than basic perception, individual differences begin to have a significant effect on user performance. As a result, designing and evaluating new visualization tools for these tasks is challenging without a more complete understanding of the strategies employed by different types of users. Preliminary observations of users in one of our target domains have shown evidence for at least two unique strategies that are reflected in preferences for different visualization designs. We propose to analyze these strategies and study their relationship to individual cognitive differences, as well as how they affect or are affected by specific visualization design choices.

Proposed approach. We plan to approach this problem from two directions: observations of user behavior, and low-level modeling of strategies based on these observations. The observations will provide qualitative empirical data about the data analysis tasks performed by biomedical researchers. In addition, we will gather data on these users' cognitive and personality differences. Task analysis methods will be used to mine this observation data for interaction patterns that indicate common approaches between users as well as how those approaches differ. In this process, we will define user strategies as variations on a single task that are common to users with a given cognitive profile. This task analysis will then be used as the basis for the modeling portion of the project. Models of user strategies for solving simple visual analysis tasks will both serve to describe the differences between types of users and help to predict and explain evaluation results.

We expect that the type of strategies employed by a user are related to the cognitive and personality factors previously found to be significant in visualization use, such as spatial ability, working memory capacity, locus of control, and openness to experience. Based on previous research, our working hypothesis is that users' cognitive profiles will fall into two groups: one which prefers more internal control over the visualization process and performs better with simpler, more interactive visualizations, and one which prefers more external control and performs better with visually complex designs such as linked multiviews. By analyzing and modeling the strategies of users that differ along these dimensions, we expect to produce better predictions of task performance on visualizations that suit these strategies to varying degrees.

Investigator expertise. The investigators are highly suited to studying and executing this vital problem. Dr. Ziemkiewicz has made key contributions to the emerging study of individual differences in visualization use, and provides a unique background in visualization theory and cognitive science. Dr. Laidlaw is a leader in producing successful scientific visualization research and applications and provides established collaborations with research groups in a variety of scientific domains, including brain research, proteomics, and biomechanics.

Research plan. We will execute this research in two primary phases. In the first phase, we will gather interaction data from scientific visualization users in a series of structured observation sessions. This data will include coded video recordings, interaction logs, user interviews, and insight evaluations. Preliminary work has shown that these types of observations can provide valuable insights into individual user strategies. In the course of gathering these data, we will also identify users' cognitive profiles by administering cognitive scales and personality surveys. In the second phase, we will use this task analysis as the basis for a set of low-level predictive models of task performance that take as input measures of a user's cognitive profile and information about a visualization interface to produce predictions of response time and accuracy on basic information retrieval tasks. We will then compare the predictions of these models against empirical data to test their validity. Taken together, the qualitative analysis of the observations and the quantitative predictions of the models will be used to develop guidelines for visualization design in the domains of our collaborators.

2 Expected Significance

Evaluating visualization systems presents unique difficulties, in large part because realistic visualization problems are difficult to capture in traditional usability testing [21]. Often, people use visualization to perform open-ended analysis. They may not know precisely what they are looking for, and they may not know how to tell when they have found it. There may be multiple ways to perform a single type of analysis, of which none is obviously superior. In real-world analytical and scientific visualization applications, these complex user tasks are not special cases; they are the core reason people use the software. These tasks are a both challenge to evaluate and one reason visualization is a critical application area.

A better understanding of these tasks and how to support them through design would improve visualization practice, advance theory and user models, and impact the many fields in which visualization is a vital part of the analysis pipeline. The task analysis portion of the proposed work will directly impact the general understanding of how real users in critical application areas perform visual analysis, and how that analysis is or is not served by current interfaces. By focusing on how individual differences impact user strategies, this work is expected to expand the theory of visualization with a greater understanding of how visualization works for different types of users.

The theories we develop will be made concrete in the modeling process, in which a user strategy is defined as a variant model of a single task that fits certain user patterns better than others. Should these models prove to accurately predict empirical data, the theories of user behavior that emerge from this work will be uniquely grounded in both realistic user observation and empirical validation. This is the high-risk portion of the proposed work, as we have not yet implemented the full strategy modeling process in practice. However, the potential rewards are high as well. A set of validated strategy models would not only contribute to the theoretical understanding of visualization, but is also expected to produce grounded visualization design guidelines that can be used in heuristic interface evaluation and system development. Such models are also a step towards automating portions of the evaluation process by formalizing principles that can produce predictions of user behavior, rather than needing to rely entirely on user studies. This could ultimately make visualization evaluation faster and more standardized.

In the short term, this project is expected to produce a large body of task analysis data, which by itself would be highly useful in understanding the types of scientific visualization applications we will study. Through our focus on individual user strategies, we expect to contribute to existing task analysis work by demonstrating how differences between users affect complex analysis tasks. A greater understanding of the role of individual differences in task performance can be used to make scientific visualization more useful and accessible for a broader range of users. Ultimately, this may help to increase the adoption rates of novel visualization systems.

3 Problem Motivation and Background

Complex data visualization systems are becoming increasingly ubiquitous in many areas of study, yet evaluations of these systems are often unreliable and inconsistent. Part of the problem is that there is a missing level of analysis in our working model of the visualization user. We can observe what the user does by recording interactions, errors, and response times, and we can ask users what they are trying to do in interviews and think-aloud protocols. However, in the complex, open-ended tasks that characterize visual data analysis, it is difficult to draw a well-supported connection between the two. How users translate their intentions into interactions remains difficult to describe and study. A major reason that this is a problem is that it obscures important individual differences between users. In interfaces that interact so closely with human reasoning, differences in cognitive style and problem-solving approach can easily be amplified. Recent visualization research has increasingly made it clear that individual differences can have a significant impact on evaluation results. Explaining how these differences influence user strategies can contribute to a more complete model of user performance under different design conditions. The proposed work fits into a long-term study of how individual differences explain and complicate visualization theory. This research project is founded both in a history of modeling the visualization user and in studying how individual differences affect a user's performance.

3.1 Models of Visualization Tasks and Users

This work builds on existing work on task analysis in the visualization domain, as well as on a history of theories of how visualization works. Both of these research directions aim to explain how a user makes sense of a visualization, although they approach the problem from different directions. Task analysis focuses on observing users in real-world situations and has tended to produce high level descriptions of their behavior. Visualization theory builds primarily on perceptual psychology and has therefore produced mainly lower level descriptions of how people read visual information. Cognitive models of visualization use fall in between these two levels of detail, but have so far proved most usable when they focus on simple perceptual models. We argue that a research plan which grounds focused models in hypotheses derived from task analysis has the potential to bridge the gap between these two branches of research. In doing so, we expect to be able to develop models that describe more complex tasks than those in existing theory while producing more falsifiable predictions than existing task analysis work.

This work adds to a body of knowledge established by previous task analysis work in the visualization domain. Springmeyer et al. [25] studied scientific data analysis in research which helped to place visualization use in the context of a larger workflow, and were among the first to argue that visualization systems should include a way to record users analytic processes. Pirolli and

Card's cognitive task analysis [20] produced the Sensemaking Loop model for intelligence analysis, which has since become influential in visual analytics theory and design. More recently, Isenberg et al.'s study [12] of collaborative analytical behavior provides a general model for how we perform a more focused task analysis. This work shows that task analysis can have a significant impact on research, but is more high-level than the analysis we intend our method to produce.

We seek a middle ground between task analysis of an entire data exploration process and the more perceptually-driven models that characterize traditional visualization theory. Visualization theory so far has produced several working models of this kind. Some of these models make use of Bertin's *Semiology of Graphics* [4], which first defined "marks" as the most basic elements of graphical encoding and defined possible mappings between types of data dimensions (nominal, ordinal, or numerical) and "retinal variables" of graphical marks (size, color, position, etc.). One of the most notable extensions of this work is the research performed by Cleveland and McGill [6] that took Bertin's basic taxonomy of visual variables and measured how finely viewers could discriminate between them in a graph-reading context. This produced a ranking of visual variables in terms of their discriminability, and concrete guidelines for which variables could be used for numerical data (e.g., length and position) and which should be reserved for nominal data (e.g., color). This provides a grounded, usable metric for one element of visualization design. Indeed, Mackinlay's APT (A Presentation Tool) [16] proposed using these mappings to automatically generate the best visual encoding for a given set of data, a system which has been implemented in the commercial visual analytics software Tableau [17].

Models that take a more cognitive perspective on the interpretation of visual information are rarer, but do exist and form an important foundation for the proposed work. Pinker's theory of graph comprehension [19] describes in detail the operations taken by a viewer of an information graphic, based on the type of visual mappings used in the graph and the particular query that a user is attempting to answer with it. Using a cognitive modeling language which is a variation on ACT-R, Lohse [15] constructed a computational model based on Pinker's theory as a test of its validity. This is an excellent foundation for a general-purpose graph comprehension model, but it has several limitations that prevent it from being immediately applicable. For one, it refers only to static graphs, while modern visualization systems use interaction as a core component. For another, Lohse's results are not strong enough to convincingly argue that this model sufficiently captures the process at work.

More successful user models have tended to focus on lower-level perceptual tasks. For example, models have been applied to the problem of choosing perceptually consistent color maps [3]. By focusing on specific visual variables, researchers have also developed mathematical models of texture discrimination [11] and symbol size discrimination [14]. Models such as these can be used to set limits on the scales used by visualization designers when employing these variables in a data mapping. A recent model of a more complex visualization process is the work on comparing zooming versus focus-and-context views by Plumlee and Ware [22]. The authors designed a mathematical model to predict how quickly and accurately a user will be able to make comparisons between detailed views in a geospatial visualization, based on whether those views are reached by a zooming operation or by generating detail windows and keeping the overall map visible. Their model proved to be successful at making predictions and explaining why focus-and-contex views are more efficient for comparison.

By modeling perceptual processes and producing design guidelines based on their predictions, this body of research shows the power of a working visualization model for improving design and evaluation. We propose to develop models on a similar level of detail to those in Plumlee and Ware's study, but the tasks we focus on will be chosen to reflect important differences in the strategies we identify in user observations. For example, if we find that some users prefer to use multiple-view visualizations rather than a series of views for a hypothesis testing task (as observed in our preliminary results, described in Section 4), the associated strategy models may predict response time on a task in which the user searches for a target that is identified by combining information from two views.

The proposed work combines the approaches of task analysis and user modeling to produce grounded, predictive theories of user cognition in scientific visualization tasks. Using task analysis to produce hypotheses for the modeling portion of the research will focus the modeling process and assist in describing tasks at a medium level of detail between simple perceptual processes and large-scale analysis procedures. However, previous research suggests that visualization tasks at this level of detail cannot be completely described without acknowledging significant variation based on individual differences.

3.2 Individual Differences in Visualization

Cognitive factors such as spatial ability, verbal ability, and working memory capacity vary substantially between individuals, and can affect reasoning in many different ways. Spatial and perceptual abilities in particular have been shown to affect how well users can perform several different tasks in a visualization system. Velez et al. [26] first showed that a number of these abilities, including spatial orientation, spatial visualization, visual memory, and perceptual speed, affect accuracy and response time on a task involving the comprehension of 3D views similar those found in scientific visualization applications.

Other work has found that this effect holds for more abstract 2D visualization tasks as well. Conati and McLaren [8] found that perceptual speed, which measures the speed at which a person can compare two figures, correlates with a user's accuracy at information retrieval tasks in one of two visualization systems: a star graph and a heatmap-like view. Users with high perceptual speed performed better with the heatmap-like view than the star graph on a comparison task, and vice versa. The authors found that this was only true for one of the tasks they studied, one in which participants were asked to compare differences in change over time between two scenarios at a global level. This was perhaps the most complex question they asked, as most of the others ask the participant to retrieve or compare a specific variable value. The proves to be a trend in studies of individual differences in visualization: frequently, significant differences are found primarily when tasks are highly complex.

Most significantly for the proposed work, some studies of individual differences have found differences not only in overall performance, but in the strategies employed by users with different cognitive profiles. For example, Cohen and Hegarty [7] found that a user's spatial ability affects the degree to which interacting with an animated visualization helps him or her perform a mental rotation task. Participants were asked to draw cross-sections of a complex 3D object. They were able to control two animated rotations of the object in order to complete the task. Participants with high spatial ability produced more accurate cross-sections and used the visualizations more, while those with low spatial ability rarely discovered the best view from which to create the cross-section. The source of this difference is that high spatial ability participants were more likely to seek out an optimal view for cross-sectioning, while the low spatial ability participants used a less

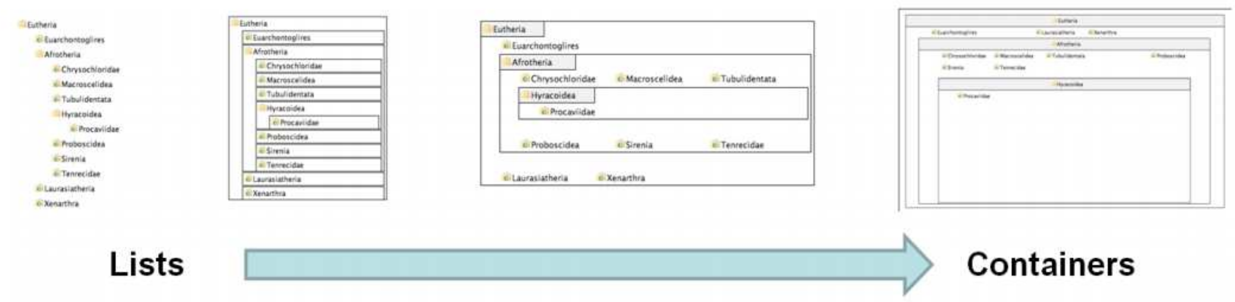

Figure 1: The four visualizations used in the study on locus of control [27]. These views were designed to vary systematically from a more list-like metaphor to a container-based metaphor. We found that participants with an internal locus of control performed progressively worse (in both accuracy and speed in solving complex analytical tasks) as the visual metaphor shifted from lists to containers. Those with an external locus of control were adept with all views, but especially with the most container-like view.

focused strategy.

Similarly, Chen and Czerwinski [5] found a relationship between spatial ability and the visual search strategies users employed in a network visualization. Participants viewed an interactive node-link visualization of a paper citation network and were given tasks to find papers on specific topics. Spatial ability was positively correlated with search task performance in general, and also predicted the use of a better navigation strategy. Low spatial ability participants were more likely to click through every node in a cluster even after determining that the cluster was irrelevant to the target topic, while participants with high spatial ability pursued a more hierarchical strategy in which they jumped from cluster to cluster until they found a likely neighborhood. The use of different strategies by users with different cognitive profiles suggests that, when user characteristics vary, there is no single way for a visualization to best support a given task. If people with varying cognitive abilities employ different strategies for the same task, a visualization designed for that task must take this into account to be effective.

In addition to cognitive factors, recent work by the principle investigator and others has shown that personality traits such as locus of control [24] can significantly affect complex task performance. Locus of control measures the extent to which a person sees herself as in control of events (*internal* locus of control), as opposed to being controlled by outside factors (*external* locus of control). Green et al. [9, 10] studied the use of visual analytics interfaces by users who differ on this dimension. The study compared two complex, dissimilar information retrieval systems, a visual analytics system and a web interface with a more list-like view. The authors found that users with a more external locus of control performed better at complex inferential tasks when using the visual analytics interface, and discovered additional correlations between neuroticism and task performance.

Building on Green et al.'s work, Ziemkiewicz et al. conducted a study [27] to identify the visual elements that resulted in the differences between the two systems. The goal of this work was to identify the specific design factors that were responsible for the reported results. To test the hypothesis that the underlying metaphor of the layout was the most significant factor, the researchers studied performance on four simple visualizations (Figure 1) that are similar in all aspects except for the overall layout style they use. The purpose of this was to isolate the significant factors in the design of the visualizations at a finer degree of detail than in previous work, which mostly studied real-world visualization tools that differed from one another in many respects.

The four views gradually shift orientation from a list view to one with a containment metaphor. Participants were first measured for locus of control and other personality factors, and then performed a series of search and inferential tasks similar to those used in Green et al. The results showed that, for inferential tasks, participants with an internal or external locus of control performed well on different visualization types, with internal participants showing increased performance as the views became more list-like. External participants showed less difference in performance overall, but were slightly more adept with the most container-like view than any others. As in Green et al., and echoing other individual difference studies, this effect was found in complex tasks but not simple search tasks.

These findings motivate our working hypothesis that user strategies in visual data analysis will tend to fall into two broad groups: those that emphasize internal processing, and those that rely more heavily on external representations. Based on the locus of control study, we expect that users who favor an internal strategy will perform best with simple visual designs that present data with minimal embellishment. Conversely, users who favor an external strategy may perform better with a visual design that gives more explicit structure to the data and has more information on the screen at one time. Preliminary results from the proposed work lend support to this working hypothesis.

4 Preliminary Results

In preliminary work on the task analysis portion of the proposed research, we have demonstrated the viability of this part of our approach and produced findings that guide our broader hypotheses. This observational study of researchers in the immunology field has revealed strategy patterns that align with our hypothesized divide between participants who rely more heavily on external representations and those who prefer a simpler visual layout and more internal processing. These results also illuminate reasons why a user may prefer one visualization strategy over another.

In this preliminary work, we have been collaborating with researchers in the CBDM Immunology Research Laboratory at Harvard Medical Center. These researchers study the genetic factors that affect immune response in mice as a model for the human immune system, work that is key to understanding autoimmune disorders such as diabetes and arthritis. As part of an ongoing effort to model this type of scientific visualization task, we observed four researchers from this lab performing typical data analyses. These researchers must manage large amounts of data from a range of experimental procedures, including gene expression data and T-cell counts in mice from varying genetic lines. In our observations, datasets ranged from 25,109 to 46,632 genes. Our users' experiments can generate up to 100 samples from 30 populations for each gene, producing tens of millions of data points that cannot be simultaneously analyzed using their existing tools. Furthermore, advanced imaging technologies may increase this number exponentially in the near future. Currently, these researchers study these data using a web-based analysis tool called GenePattern [23], which allows them to generate a variety of scatterplot graphs and select groups of genes based on filtering criteria known as signatures. Discovering the gene signature that is characteristic of a particular variation in immune system behavior is the primary goal of this type of analysis.

Our results revealed that, even when users have similar goals and are using the same software, different participants used clearly diverging strategies. We characterize the strategies we observed as within-graph and between-graph analysis. In within-graph analysis, the user maintains a small number of visual layouts and interacts heavily within each layout through selection and filtering. In between-graph analysis, the user continually generates new graphs. If selection is used, it is used to faciliate comparisons between graphs in sequence. The differences in these strategies are significant enough to suggest that they would be best supported by entirely different visualization designs.

The participants were four postdoctoral researchers in the CBDM lab. These four volunteered since they were, at the time, working on the data-analysis portion of their research. The participant group included two women and two men, all with comparable levels of experience in their field and all working on different but related experiments. The four observation sessions took place during a single day and were performed on a single workstation. The primary analysis tool used by these researchers is a web-based system for gene expression analysis called GenePattern [23], specifically the Multiplot Visualizer function (Figure 2). Multiplot was primarily used to generate scatterplots showing correlations between two gene expression variables and also "volcano plots," used to show p-values for a given analysis. One of the participants (P2) also used Tibco Spotfire S+ [1] to generate data tables.

The observation sessions yielded four hours and forty-eight minutes of video and accompanying notes. This video was coded for analysis using a coding scheme based on Springmeyer et al. [25] and the low-level visual analytic tasks identified by Amar and Stasko [2], along with quantitative data about the number and duration of graphs generated and filtering schemes applied. The results of the high-level coding pass showed a general repeated analysis pattern that was common to all participants while using GenePattern: viewing a graph, filtering, and retrieving individual values. Essentially, participants were looking for patterns in their data, identifying a possible gene signature of interest, and examining specific genes in that signature to test their hypotheses.

Although there were common high-level patterns in the four participants' analysis behavior, there were substantial differences in the details of their behavior. P2's session was somewhat anomalous: he preferred to use the scripting capabilities of the S+ system rather than perform visual analysis using GenePattern at all. This made his analysis session much shorter, since he was applying previously developed analysis scripts, and suggests that he may represent a type of user who either would not benefit from visualization or is not inclined to adopt it. Among the other three, we observed two strategies, which we refer to as within-graph and between-graph analysis.

Within-graph analysis, the strategy used by P3, was characterized by a relatively small number of primary graphs, frequent use of secondary graphs, and constant changes in highlighting schemes. We refer to this as "within-graph" because most interactions happen within a single arrangement of graphs, rather than being used to switch between graphs sequentially. Uniquely among the participants, P3 always had two scatterplots visible at once (Figure 2). Although she used multiple views, she maintained each overall layout of those views for a long period of time. The bulk of her interaction did not involve changing graphs, but rather, changing the appearance and filtering parameters of existing graphs. When asked why she used multiple views, P3 responded that it was important to see both behavior and significance at once. "If I don't have two [views], I have to go back and forth. Going back and forth, you can forget and lose time." In some ways, P3 is the ideal user for whom visualization researchers tend to design: interested in maximizing the amount of information on screen at once, focused on efficiency, and open to many multiple views. However, not every user is like P3, as P1 and P4 demonstrate.

Between-graph analysis, used by P1 and P4, was characterized by greater numbers of graphs in

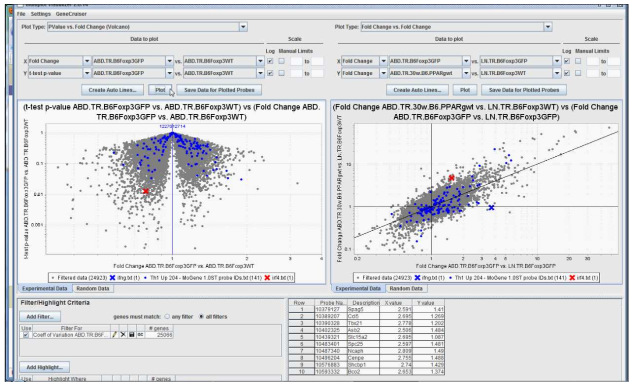

Figure 2: The GenePattern analysis environment, as used by participant P3. Users can generate one or more scatterplot graphs based on a variety of axes derived from their data. In this case, P3 is viewing a plot of the changes in gene expression between two study conditions (right) and a volcano plot of the significance of those changes (left). A blue highlight is used to identify a specific gene signature.

quick succession, with highlighting used to facilitate comparison between graphs or not at all. In this strategy, most interaction focuses on switching between graphs, rather than interacting within single graph layout. Neither participant ever had more than one scatterplot visible at a time, and both changed graphs more frequently than P3. They also changed highlighting and filtering schemes less frequently, using these highlights to track a group of genes across multiple layouts. When asked about his use of visual analysis, P1 revealed a possible explanation for viewing the graphs in sequence rather than in parallel: "I like to turn the data upside down and sideways, looking for 'realness.' If you try different plots, different views, and still see something, you can be more reassured that these genes are differentially expressed." This perspective suggests that changing graphs frequently might be a way to increase confidence in a result. While P3 worried that going back and forth would make her forget something, P1 treated that very process as a way to confirm or reject his hypotheses. Similarly, when P4 encountered an error in her data during the course of her analysis, she followed up her discovery by quickly switching between several views (contributing to her high number of graphs). If a user is concerned that a pattern seen in one view might be biased or illusory, replacing that view entirely with another could be a tactic to view the data with fresh eyes. This view of the value of interaction differs from that usually emphasized in visualization research, and merits further exploration.

Our findings add to the body of evidence that, in complex analytical environments, designing for only the "average user" is not realistic. The more complicated and abstract a goal is, the more likely it is that users will have different yet equally reasonable approaches to it. These results lend support to the claim that in applications at this level of complexity, analyzing the varying strategies within a single task allows for a more complete picture of users' needs. They also provide early support for our working hypothesis that internal and external strategies are better suited to different types of visualization design. The within-graph strategy we observed would clearly be better served by a coordinated multi-view visualization. P1 and P4's between-graph strategy would benefit from a system that allows for quicker changes between views, and possibly one which supports animated transitions to avoid change blindness. In this way, these findings demonstrate

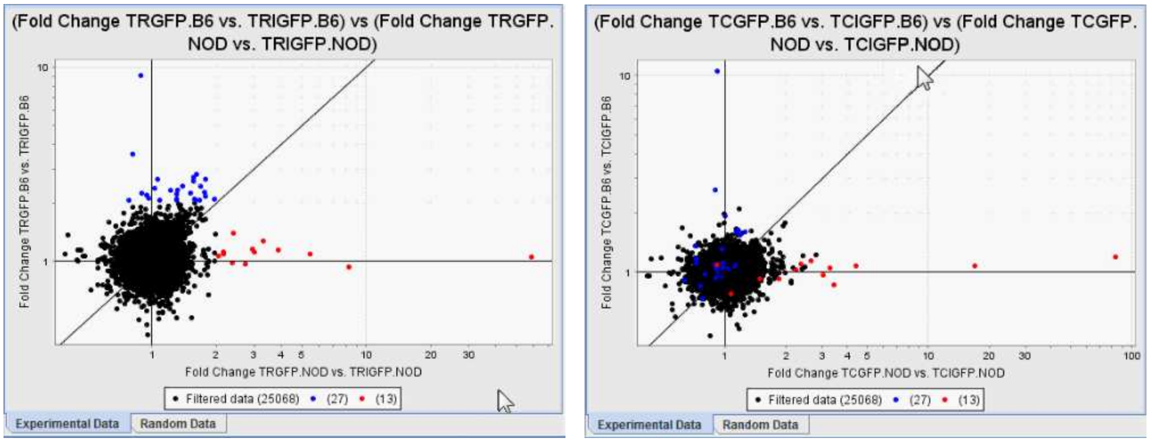

Figure 3: Two steps in P1's interaction, before and after a change in graphs. P1 used highlights such as the red and blue selections in this example to compare gene expression behavior across multiple conditions.

the viability of our observational approach in uncovering information about user strategies within a given task. Our proposed research plan builds on this preliminary work and has been shaped by the lessons learned during this research.

5 Plan of Research

Our research plan is divided into two phases, each expected to take a year. In the first phase, we will gather and analyze data on user behavior with scientific visualization systems by leveraging our existing collaborations with researchers in biomedical sciences. This phase will produce a detailed task analysis of this behavior which we will then use to direct the modeling process in the second phase of the research. This phase is expected to result in a set of focused user strategy models that make predictions about response time and accuracy based on a user's cognitive profile and variables of a visualization design. Finally, we will test these models and use them to generate design guidelines for visualizations in the domains we study.

5.1 Phase 1: User Observations and Task Analysis

A solid understanding of the analysis process of scientific users is needed to inform a general model of the kind we intend to build. Therefore, the objective of the first year of this research is to have a collection of data and knowledge on which to build our initial model. Our working hypothesis is that using a multi-view visualization system is an iterative process in which multiple data facets and data representations, represented by visualization windows, are incorporated into a user's mental model. The type of information present in these views, how they are represented and arranged, and the visual and interactive links between them all contribute to the accuracy and efficiency with which this mental model is built. We will test this hypothesis by observing the stated reasoning process of working scientists and how they use a visualization system to execute that process.

In doing so, we will leverage the existing working relationships between the Brown Visualization Research Lab and several domain-specific scientific research labs at Brown University. These projects include cell signaling proteomics with the Salomon Research Group [13] and wrist biomechanics with researchers at the Alpert Medical School [18]. The Visualization Research Lab has developed or partially developed visualization systems with each of these groups, so there are already active systems in the field. In addition, the preliminary work on the proposed research has established relationships with immunology researchers in the CBDM Laboratory at Harvard Medical School. By beginning our research in this way, we will ensure that the subsequent models are grounded in reality and reflective of actual user behavior. We expect this portion of our research to produce a rich body of knowledge that reflects general thought processes and interaction styles of varying scientific visualization users, explaining what they do with a visualization system and why. The following subsection details how we intend to build this body of knowledge.

5.1.1 Research Design

The objective of this phase is to closely observe how users employ these visualizations in pursuit of a data analysis task. The first aim of this work, therefore, will be to define these analysis tasks and how they differ between types of users. The purpose of this portion is to produce a body of knowledge that will guide the hypotheses used in model development.

- **Interviews with domain scientists.** We will conduct detailed interviews with as members of the scientific research labs with which we are collaborating. These interviews are expected to discover what the scientists intend to do with their data analysis, regardless of whether the visualization allows them to do that or not. We will ask the scientists what their analysis questions are, what problem-solving process they go through when trying to answer those questions, what obstacles they tend to encounter, and so on.

- **Observation of analysis work.** Following the format used in our preliminary research, we will schedule sessions where one of the members of our research team visits a domain scientist and observes actual data analysis with the systems currently being used. We will employ interaction logging software, screen capture software, and video recording to track the user during their performance of the analysis task. Immediately afterwards, we will present the video to the participant for a post-hoc think-aloud protocol in which they describe their thought process while viewing the video of themselves. This is intended to provide a reasonably accurate recording of the participants' thoughts during the study without disrupting their analysis process at the time.

- **Task analysis.** Based on the findings from the interviews and observations, we will perform a task analysis of the analytical work done by these domain scientists. We will first classify the tasks we observed into types if necessary, with the goal of having the fewest possible coherent task types. We will then use task analysis methods to decompose the tasks into steps and identify the subgoals of each step.

5.1.2 Expected Outcomes

This phase of the research is expected to produce the following outcomes:

- A task analysis of the analytical activity undertaken by scientific users in a variety of domains.

- A body of data describing scientific visualization analysis in practice, including measures of interaction sequences as well as qualitative information gathered from interviews, video recording, and post-experiment talk-aloud protocols.

We expect this phase of the research to take about a year, including time to design our observational methods, interview our collaborators, set up tracking equipment and interaction logging software, schedule visits with the outside research labs, and perform the task analysis.

5.2 Phase 2: Modeling and Developing Guidelines

The primary goal of the second year of the research is to take the body of data generated in Phase 1 and use it to generate specific predictions of response time and accuracy for different user types. These predictions will take the general form of the models used in Plumlee and Ware's study of zooming and multi-window visualizations [22]. Our working hypothesis is that users' cognitive profiles will fall into two groups: one which prefers more internal control over the visualization process and performs better with simpler, more interactive visualizations, and one which prefers more external control and performs better with visually complex designs such as linked multiviews. However, if this hypothesis does not appear to be a good description of our measurements and observations from Phase 1, we will formulate a revised hypothesis based on our empirical findings. This portion of the research will be iterated as necessary to produce models that make valid predictions about user behavior.

5.2.1 Research Design

The first step in this iterative process is to generate a set of initial predictions of performance differences based on our findings in Phase 1.

- **Generating an overall working hypothesis.** By modifying our initial theory based on the findings from Phase 1, we will generate a basic hypothesis about how individual differences affect analysis strategies and why. This hypothesis will act as a general guide to cognitive model development.

- **Generate specific hypotheses.** Within our overall hypothesis, we will develop specific hypotheses that are each at a simple enough level of detail to be described by a single predictive model. These hypotheses will be general predictions of user performance given parameters of a user's cognitive profile (such as external versus internal locus of control) and a type of visualization design (such as multiple-view versus single-view). We expect to generate multiple hypothesis to cover different types of tasks, based on patterns observed during the task analysis phase.

- **Build models based on hypotheses.** We will generate computational models that make specific predictions of response time or accuracy for each hypothesis generated in the previous step. These predictions will be driven by the findings from Phase 1 and by reference to relevant findings in cognitive science and perception.

- **Comparing predictions to Phase 1 findings.** Once the initial models are built, we will test whether their predictions reasonably approximate the visualization use sequences recorded in Phase 1 when the model is given similar input. If a model produces sequences that clearly contradict the empirical data, we will alter the model accordingly.

When our set of models seems plausible when compared to our initial data-gathering, we will continue to the next step of testing their predictions in a controlled setting. This will take place in a series of user of studies, each of which focuses on one of our hypotheses. The following steps will be taken for each hypothesis.

- **Experimental setup.** We will begin by designing controlled test visualizations, broadly resembling those used by our collaborators, that only differ from one another on the design parameters that act as input to the computational model. This is intended to clearly isolate these factors in the experiment. Tasks will then be designed that can be performed with these test visualizations and which represent each of the task types identified in Phase 1.
- **Generating model predictions.** These design configurations and task types will be used as input to the computational model in order to generate visualization use sequence predictions for each of the conditions that users will see in the subsequent experiment.
- **User study.** We will then perform a large-scale study of task performance under the different test conditions. In order to reach a sufficiently large number of participants, this study may include layperson participants, but will also include some of the experts from the domain research groups to test for any differences between experts and novices. The users will be trained on the visualizations and then asked to perform the tasks across each of the conditions. The task completion time, error rate, and reported usability of the systems will be recorded, as will user interaction sequences.
- **Analyzing the results.** We will analyze the data from this study with a focus on whether the response time and accuracy of the participants usually match the predictions of the model. We will also examine whether the participants' interaction sequences match the user behavior observed in Phase 1, to validate the connection between our observations and quantitative predictions. If the initial model does not appear to make valid predictions about performance, we will look into this data for an explanation of why and to determine what kind of information the model does not seem to capture.
- **Generating design guidelines.** We will use the results of this process to provide specific advice for designing a visualization for a particular strategy type or for generating interface customization options to cover multiple user strategies. These guidelines will take the general form of layout recommendations given a user type and task.

5.2.2 Expected Outcomes

If any of our initial models do not prove to be a valid predictor of user behavior at the end of this phase, this phase will be repeated for the hypothesis represented by that model as necessary and as time permits. When all iterations are complete, this phase of the research is expected to produce the following outcomes:

- A set of validated models which can be used to predict a user's performance on a specific visualization task given relevant user factors and design parameters.

- A set of cognitive theories to explain differences between user strategies.
- An empirically validated set of design guidelines for biomedical visualization applications.

We expect this process to take a year, at the end of which we will have a set of models that describe task strategies for different user types in scientific visualization applications. The goal of this phase is to turn the task analysis output of the previous phase into concrete, validated models that predict user performance.

6 Results from Prior Support

Over the last five years David Laidlaw's research group has been funded by several NIH awards or subcontracts to NIH awards, continuation funding from several NSF awards and a Keck Foundation grant, one recent NSF instrumentation grant, and a collaborative NSF award (IIS-10-16623). Most of the NIH awards and subcontracts related to developing and applying computational tools for the analysis of diffusion MRI imaging data. Laidlaw's group has published extensively in this area with collaborators from around the world. Another NIH award was for developing computational bioengineering tools to study the human carpus in collaboration with PI Joseph Crisco in the Department of orthopedics. A final NIH subcontract involves creating interactive tools for studying gene expression in the immune system. This is in collaboration with Christophe Benoist at Harvard Medical School. Laidlaw's most recent NSF funding was an MRI award to develop a new virtual reality display instrument at Brown. That project is in year two of four. Finally, he has been a co-PI on Keck and NSF awards to study animal kinematics and dynamics.

The most similar award to the one proposed is titled "GV: Small: Collaborative Research: Supporting Knowledge Discovery through a Scientific Visualization Language." This project, funded for $261,596, began on November 1, 2010, with an estimated end date of October 31, 2013. The purpose of this award is to study how best to design a scientific visualization language for creating visualizations of diffusion MRI data. As this work was very recently awarded, there have not been any publications associated with it yet. At a broad scale, however, in the last five years Dr. Laidlaw's research has led to about 25 refereed journal articles, approximately 10 refereed conference papers, two patent applications, and about 40 conference abstracts or posters. He has been recognized via a number of best poster and best panel awards at conferences and received in 2008 the prestigious IEEE VGTC Visualization Technical Achievement Award.

7 Broader Impacts

7.1 Education and Infrastructure

The budget includes funding for a graduate student research assistant, who will be an integral part of the execution of this plan. This student will learn how to conduct user experiments, formulate theories, and build cognitive models to express theories. They will also have the opportunity to conduct interdisciplinary research in collaboration with scientific experts accross domains. At the same time, this research will provide an opportunity to educate students working in brain science, proteomics, and medical research labs on cutting-edge visualization techniques. This will also serve as an opportunity for these students to examine their own analysis process and find ways to critique the analytical tools they use.

Another educational opportunity is provided by Dr. Laidlaw's graduate-level course on Interdisciplinary Scientific Visualization, in which students are asked to propose and execute a class project in collaboration with scientists at Brown University. As this research offers an ideal opportunity for students to learn how to thoughtfully collaborate with domain researchers in building and evaluating a visualization system, we will suggest portions of this research as one of the projects students can choose from.

This work will improve research infrastructure by deepening collaborative relationships between the Visualization Research Lab and other research groups at Brown and elsewhere. In addition, it will improve the analysis tools used by scientific researchers in these groups, allowing their work to proceed more efficiently and insightfully in the future. Another benefit of this close work with users is in its potential to involve underrepresented groups, particularly women, in visualization research. We strongly believe that one of the best ways to broaden participation in computer science is to demonstrate its creative potential and relevance to real-world problems. By closely collaborating with researchers outside of computer science, especially those in medical and biological fields, the proposed work has the potential of exposing female students in other disciplines to creative and practical applications of advanced computer science research.

7.2 Dissemination and Societal Benefits

Our goal in this research is to build the foundation for a broader research agenda on user modeling in scientific visualization and how individual differences affect task strategies. To that end, we will develop a public Scivis Task Modeling Data Repository to disseminate the data we collect as well as the models we develop. We will publicize this website through publications, conference talks, and social media, and will invite other researchers in similar domains to contribute their data as well. As this kind of task analysis and modeling work is both time consuming and not common in the visualization domain, we believe that creating and building this repository will be necessary to advance this area of research more quickly than has been done in the past. By collecting task data from multiple domains and research groups, we will have a better chance at finding generalizable patterns and creating a sustainable long-term research effort.

Along with deploying our improved visualization designs to our scientific collaborators, we will make a concerted effort to disseminate these results to the visualization research community at large. This will include publishing the results in appropriate research conferences and journals at regular intervals. Some of these results, such as the design guidelines from Phase 2, are expected to have an immediate impact on improving visualization development. Others will contribute to a more long-term discussion about visualization theory and research directions. In general, the proposed work has a unique potential among visualization theory research to directly impact practice as well as the broader understanding of how visualization works.

Finally, publication of design guidelines that are theoretically and empirically grounded will improve visualization designs across numerous fields. In particular, the focus on understanding individual differences in this research has the potential to make visualization more accessible to a diverse set of users. As visualization has expanded in widespread use as well as in the importance of its applications, the potential impact of foundational visualization research has expanded as well. Work that contributes to the basic scientific understanding of how a user makes sense of complex visual information can help to make tools more effective in a wide variety of fields, both scientific and otherwise.

References Cited

[1] Tibco Spotfire S+. Accessed 9/21/2011.

[2] Robert Amar, James Eagan, and John Stasko. Low-level components of analytic activity in information visualization. In *Proceedings of IEEE InfoVis*, pages 15–23, 2005.

[3] L. D. Bergman, B. E. Rogowitz, and L. A. Treinish. A rule-based tool for assisting colormap selection. In *Proceedings IEEE Visualization*, 1995.

[4] Jacques Bertin. *Semiology of Graphics*. University of Wisconsin Press, 1967.

[5] Chaomei Chen and Mary Czerwinski. Spatial ability and visual navigation: an empirical study. *New Review of Hypermedia and Multimedia*, 3(1):67–89, 1997.

[6] William S. Cleveland and Robert McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

[7] Cheryl A. Cohen and Mary Hegarty. Individual differences in use of external visualisations to perform an internal visualisation task. *Applied Cognitive Psychology*, 21:701–711, 2007.

[8] Cristina Conati and Heather Maclaren. Exploring the role of individual differences in information visualization. In *Advanced Visual Interfaces*. ACM Press, 2008.

[9] Tera M. Green and Brian Fisher. Towards the personal equation of interaction: The impact of personality factors on visual analytics interface interaction. In *IEEE Visual Analytics Science and Technology (VAST)*, 2010.

[10] Tera M. Green, Dong Hyun Jeong, and Brian Fisher. Using personality factors to predict interface learning performance. In *Hawaii International Conference on System Sciences 43*, January 2010.

[11] Danny Holten, J. J. van Wijk, and J.-B. Martens. A perceptually based spectral model for isotropic textures. *ACM Transactions on Applied Perception*, 3(4):376–411, 2006.

[12] Petra Isenberg, Anthony Tang, and Sheelagh Carpendale. An exploratory study of visual information analysis. In *ACM SIGCHI*, pages 1217–1226, 2008.

[13] Radu Jianu, Kebing Yu, Vinh Nguyen, Lulu Cao, Arthur Salomon, and David H. Laidlaw. Visual integration of quantitative proteomic data, pathways and protein interactions. *IEEE Trans. on Visualization and Computer Graphics*, September 2009.

[14] Jing Li, Jean-Bernard Martens, and Jarke J. van Wijk. A model of symbol size discrimination in scatterplots. In *Proceedings CHI*, 2010.

[15] Jerry Lohse. A cognitive model for the perception and understanding of graphs. In *Proceedings CHI*, 1991.

[16] Jock Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2):110–141, 1986.

[17] Jock Mackinlay, Pat Hanrahan, and Chris Stolte. Show Me: Automatic presentation for visual analysis. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1137–1144, 2007.

[18] G. Elisabeta Marai, Joseph Crisco, and David H. Laidlaw. A kinematics-based method for generating cartilage maps and deformations in the multi-articulating wrist joint from ct images. In *Proceedings of the IEEE International Conference of the Engineering in Medicine and Biology Society (EMBS)*, New York, NY, September 2006.

[19] Steven Pinker. A theory of graph comprehension. In Roy O. Freedle, editor, *Artificial Intelligence and the Future of Testing*, pages 73–126. Lawrence Erlbaum Associates, 1990.

[20] Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, 2005.

[21] Catherine Plaisant. The challenge of information visualization evaluation. In *Advanced Visual Interfaces*, pages 109–116, 2004.

[22] Matthew D. Plumlee and Colin Ware. Zooming versus multiple window interfaces: Cognitive costs of visual comparisons. *ACM Transactions on Computer-Human Interaction*, 13(2):179–209, 2006.

[23] M. Reich, T. Liefeld, J. Gould, J. Lerner, P. Tamayo, and J. P. Mesirov. GenePattern 2.0. *Nature Genetics*, 38(5):500–501, 2006.

[24] J. B. Rotter. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80(609), 1966.

[25] Rebecca R. Springmeyer, Meera M. Blattner, and Nelson L. Max. A characterization of the scientific data analysis process. In *Proceedings of IEEE Visualization*, 1992.

[26] Maria C. Velez, Deborah Silver, and Marilyn Tremaine. Understanding visualization through spatial ability differences. In *IEEE Visualization*, pages 511–518, 2005.

[27] Caroline Ziemkiewicz, R. Jordan Crouser, Ashley Rye Yauilla, Sara L. Su, William Ribarsky, and Remco Chang. How locus of control influences compatibility with visualization style. In *IEEE Visual Analytics Science and Technology*, 2011.

Mentoring Plan for Postdoctoral Researcher Caroline Ziemkiewicz

Dr. Caroline Ziemkiewicz, as principal investigator, will be involved in managing and coordianting the proposed research as part of the Brown group for an expected appointment of two years. Dr. Laidlaw will act in a mentoring role during this time. Our mentoring plan is designed to aid Dr. Ziemkiewicz in her development of her own long-term research and professional goals.

Interdisciplinary Collaboration. Caroline will work closely with students and faculty in our group to pursue a research agenda that involves close collaboration with scientific researchers, cognitive science, and computer science. This research will include regular collaboration with cognitive scientists and scientific domain experts. The diversity of scientific research labs involved in this work will give Caroline the chance to learn about a broad array of relevant subjects and to apply visualization research in a generalizable way independent of specific data content. The working relationships encouraged by this process will also help to establish interdisciplinary collaborations that can potentially follow Caroline as she continues her career beyond the postdoctoral level.

Publication and Training in Grant Preparation. Caroline will receive assistance from Dr. Laidlaw in the process of writing and publishing these results in reputable journals and conferences. She will be involved in preparing further proposals, as a Co-PI or a PI, to garner funding to support work on subsequent projects, and we will jointly determine how to appropriately share any resulting awards after her departure from Brown.

Career Counseling and Guidance. Dr. Laidlaw will be available to meet with Caroline weekly throughout the period of the mentorship, focusing on her evolving career and research goals. Initially, we will work to define these goals; subsequently we will evaluate whether she is on an appropriate trajectory to reach them. Progress reports at the end of each semester will be used to ensure that this trajectory is being followed and any areas requiring additional guidance. This teleological approach to mentoring naturally supports technical advising on the research itself while preserving the higher-level goals of both the mentee and mentor.

Guidance on Improving Mentoring and Teaching Skills. Caroline will be closely involved in mentoring a graduate student who will be funded under the proposed research, giving her the opportunity to learn how to guide a student towards a research goal. She will also participate in the teaching of several relevant classes taught by Dr. Laidlaw, including the Interdisciplinary Scientific Visualization course and a course titled Virtual Reality Design for Science, which is offered jointly with the Rhode Island School of Design and will provide a unique opportunity to learn how to teach in an interdisciplinary environment. She will also help to advise students performing independent studies. She will also lead group discussions in research meetings and will have the opportunity to teach her own computer science course at Brown. In addition, presenting the findings of this research at conferences will give Caroline the opportunity to practice presentation and lecturing skills and to gain exposure in the visualization research community at large.