Project Summary

CHS: Medium: Improving Visual Analysis Systems for Brain and Genomics Research Using Predictive Models of Sensemaking

Overview We propose human-computer interaction (HCI) research to advance predictive modeling of humans interacting with exploratory scientific visualization and analysis interfaces. These models will predict metrics of task performance and reasoning that occur in typical scenarios and provide a quantitative means for evaluating and ultimately improving the design of two scientific web applications. We will record usage data the applications, develop and test predictive models of human behavior from the usage data, and use the models to improve the applications in the wild. The scientific applications, BraiNet and MAGI, support visual analysis of brain connectivity and cancer genomics data, respectively. The behavioral modeling will be done at three levels or time scales: a unit-task level, where individual operations involved in unit visualization tasks will be predicted; a reasoning-phase level, where phases of sensemaking that corresponding to minutes-long, goal-oriented analysis activity will be classified from interactions; and an insight level, where the complete user interaction history in a session of tens of minutes will be used to predict the quality and quantity of insights generated.
Keywords: human-computer interaction; visual analytics; scientific visualization; human-centered computing

Intellectual Merit The intellectual merit of this project is fourfold. First, computer scientists will advance their understanding of how computers and humans interact at different levels, from perceptual and motor to cognitive reasoning. In particular, this understanding will be captured in predictive models of human behavior that can be used and extended by other researchers. Second, this improved understanding should lead to both specific improvement in our driving scientific applications as well as to generalizable design principles for similar analytic applications. Third, we anticipate that some aspects of our behavioral modeling will allow automation of the development and evaluation of effective techniques. Such automation could greatly speed the pace at which user interfaces can be evaluated and improved. Fourth, while we are targeting models to help understand interactions between users and computers, we may learn more about how humans think and behave, possibly advancing knowledge in the understanding of the psychology of human problem solving.

Broader Impact. The broader impact of this work comprises advances in genomics and brain science made possible through improvements to MAGI and BraiNet, use of our results and techniques by other HCI researchers to improve other interactive software, increased productivity of users of analytic software both scientific and otherwise, and training of students and other researchers in doing analogous interdisciplinary work. This project will provide practical experience with interdisciplinary research to graduate and undergraduate students. The anticipated results will enable scientists and analysts outside science to advance their analysis agendas more efficiently by reasoning more efficiently with the support of new knowledge-generating visual-analysis interactive software. Many other disciplines study linked networks of data, e.g., protein signaling and even crime and terrorism analysis; all have the potential to benefit. Almost any human-computer interface that involves reasoning has the potential to be improved by results from this research.

Project Description

CHS: Medium: Improving Visual Analysis Systems for Brain and Genomics Research Using Predictive Models of Sensemaking

In this project, we propose to pursue a research agenda that will integrate cognitive and task performance modeling to improve scientific visual analysis application designs and theory. To do this, we will create models that can predict user performance and analysis processes using user interactions as inputs. We will then improve visual analysis application designs using these predictions.

We will target cognitive activities that occur during the sensemaking process as analysts use scientific visualizations and investigate how interaction with these tools facilitates sensemaking and leads to improved analytical outcomes. To understand these activities, we will study and model three levels of cognitive activities with increasing complexity and time scales. The first level focuses on unit tasks, which are basic, well-defined visual analysis tasks (e.g., find the shortest path between two nodes in a graph) that are common building blocks of more complex tasks. We will create a model that takes in keystroke-level user actions and predicts the sequence of unit tasks the user has performed. The model will produce outputs that help fit models of cognitive activities that occur at the two higher levels described next – reasoning during sensemaking tasks and insight-formation. The second level focuses on phases of reasoning in more complex sensemaking tasks. We will create a model that predicts the user's short-term data analysis goal (e.g., exploratory hypothesis formation, evidence collection for a specific hypothesis) from a sequence of user actions that capture how the user manipulates views and data and the data being manipulated. The third level focuses on insight formation that results from interacting with key components of a visual analysis application. We will adopt and extend a set of established insight characteristics, such as insight depth, and create a model that predicts those characteristics from the entire sequence of user interactions in a visual analysis session. In addition, we will identify individual differences that, as suggested by previous works, may influence user behavior or performance at each level and assess how much variance they account for in our models.

Our modeling efforts will be used to improve scientific applications in two analysis domains: brain connectivity analysis and cancer genomics analysis. Initially, visual analysis applications will serve as testbeds in the wild and will be instrumented to collect user usage data. These data will be fit to the performance models which we will develop informed by previous theoretical frameworks and controlled-lab studies that we will conduct. In the latter stages of the project, these visual analysis applications will be improved using task-performance and analysis-process predictions from our models. Improvements will take place in different forms, including refined layout and interaction designs that support smoother analysis workflow and new interface components that help reduce the cognitive effort required during visual analysis.

The resulting experimental data, modeling software, and application code will be made available online. This will enable researchers to build upon the cognitive-perceptual findings we will contribute, and it will provide tools to help visualization designers improve applications using predictions of task performance and user analysis goals.

We note that there are intrinsic challenges to the proposed research, particularly because it involves gathering data about human behavior and using that data to predict future behavior. Such predictions are difficult, and that difficulty increases with the duration and complexity of the behavior. Nonetheless, as we will describe, there is some evidence that predictions can be sufficiently successful to be useful. Incorporating predictive modeling into the design of visual analysis software is an exciting research direction with the potential to impact both visualization research and discoveries in domains with a need for analytics tools. Furthermore, even if our predictions have limited success, our findings will advance general knowledge about what kinds of predictions are possible, and our experiments will help identify areas that need more research to inform future modeling work. Our experience with user studies will help us avoid some of the pitfalls and risks of experimental design that could limit the predictive power of behavioral models.

a Research Contributions

The contributions of this work will include the design and evaluation of cognitive and perceptual models for scientific visual analysis application; improved scientific visual analysis applications that support analysis in two domains, along with possible scientific findings in those domains; design guidelines for visualization layouts and components discovered after improving the scientific visual analysis applications; empirical findings on individual differences that affect exploratory visual analysis; and dissemination of data and software.

a.1 Novel models of visual analysis performance at three levels of granularity together with evaluations of their accuracy in predicting user performance. This work will provide theoretical foundations for improving visual analysis application design. At the level of unit tasks, like brushing and linking between two visualizations, we will predict meaningful sub-tasks or high-level actions using empirical interaction data collected during visualization use. At the level of reasoning and forming analysis goals for a complex dataset, we will study, create, and evaluate predictive models for intermediate reasoning phases and goals during the user's reasoning process. At the level of discovering insights about the data, we will analyze user interactions to better understand what kinds of activities lead to hypothesis generation and insight formation.

a.2 Improved scientific visual analysis software for two applications: brain science and cancer genomics. The design of two scientific visual analysis applications that are state-of-the-art will be improved using predictions of task performance and quality, and we will disseminate these applications to domain experts to support scientific analysis. We have two application prototypes in place that we have been developing with collaborators in two scientific domains: BraiNet [Guo et al., 2013] in the domain of functional brain connectivity, and MAGI [Leiserson et al., 2015] in the domain of cancer genomics. In the earlier stages of this project, these applications will serve as a testbed for evaluating predictive models of analysis and for collecting data. In the later stage, these applications will be improved using our findings and disseminated. While these applications focus on neuroscience and genomics data, the findings about improving user interfaces using predictive performance models will apply to other analytics domains with linked networks of data, like systems biology and intelligence analysis.

a.3 Design guidelines generalized from model evaluation results and improvements to the two applications. We expect to create guidelines for designing visualization layouts. An example would be a set of guidelines for optimizing grid-spacing in grid-based visualizations given different user tasks. These guidelines will be helpful for other designers who wish to improve their visual analysis software without running performance modeling software. The guidelines will provide a theoretical foundation that complements existing guidelines that are popular but anecdotal, such as those in Bertin's semiology [Bertin, 1983]. We also expect guidelines for improving interactions for UI components, and for constructing UI components that nudge users toward insight-promoting activities. For example, we expect to derive recommendations for designing components, such as notes-taking and annotation mechanisms, that nudge the analysts to explore more alternative hypotheses.

a.4 Empirical findings on how some individual differences affect exploratory analysis behaviors with scientific visual analysis applications. We will measure selected individual difference factors – like whether a person tends towards *deliberative* or *intuitive* reasoning during problem solving – that may influence visual analysis style and performance as suggested by recent work. While the primary goal of measuring individual differences is to account for variances in the proposed user studies instead of to perform a comprehensive analysis of the roles played by these factors, we expect our findings to provide additional evidence relevant to previous work on individual differences and visual analysis. Therefore, the findings may spur new hypotheses along this line of research and lead to design recommendations on user interfaces that adapt to different user styles and abilities.

a.5 Modeling tools and experimental data. The tools and data resulting from this project will be made available online so other researchers can reproduce, use, or extend our theoretical findings. The data collection tools and protocols we will develop will be designed so that other applications can integrate them as easily as possible. Source code for running predictive models and evaluation tools for visual analysis applications will be available online. Data collected from our experiments will be hosted in a public repository and formatted so that they can be reused in later analyses.

b Significance

b.1 Modeling Interactions at the Unit-task Level The proposed work will create a novel model that predicts the sequence of basic visualization tasks performed given mouse and keyboard inputs as well as performance for those tasks. This work is significant because it provides means to efficiently generate profiles that summarize how frequently each specific focused analysis task is performed with a visual representation and evaluate how well that representation supports each task. Our work will build on existing low-level modeling tools and will inform the design of sensemaking models that incorporate exploratory interactions alongside focused, task-based interactions. In this section, we discuss modeling tasks that take on the order of ten seconds, which is referred to as "unit task time" in early literature by Card and Newell [Newell, 1990]. Anderson argues that events at this time scale can impact behaviors measured at much larger time scales [Anderson, 2002], like analysis outcomes or education, which motivates the prediction-driven application improvements we plan to make at the unit-task level.

To the best of our knowledge, performance models for visualization tasks have not been explored beyond basic graphical perception processes. However, perceptual models have been successful in predicting some visual analysis behaviors, e.g., how quickly people perform a particle-advection analysis task with different streamline rendering designs [Pineo and Ware, 2008], and task completion times for basic analysis tasks with bar and line charts [Lohse, 1993]. A limitation of these earlier models is that they handle only basic graphical perception processes, and our model will handle more complex processes by incorporating theories and models which account for how specific visual design properties affect graphical perception. For example, Rosenholtz et al. [Rosenholtz et al., 2009] proposed a perceptual grouping model that computationally segments visual objects into visual groups. They demonstrated that the segmentation results are similar to human-perceived groupings on line charts.

Other predictive models for task completions on user interfaces have been created at the keystroke level by HCI researchers, but they have not yet been applied to improve visualization designs in a systematic way. Project Ernestine was one of the first applications of cognitive modeling to improve a real application workflow [Gray et al., 1993], using a modeling framework that identified a critical path of interactions that could be improved to make the workflow more efficient. One challenge in using modeling frameworks based on user goals and actions (e.g., GOMS models [John and Kieras, 1996]) is the high level of expertise required to fit data from real users to these models. A simpler predictive model, called the Keystroke-Level Model (KLM) [Card et al., 1980], assigns a time duration to each low-level interaction required by a task in order to predict expert task completions on an interface. As discussed in Sec. c, we have experience modeling low-level interactions for visual search tasks with the KLM in our Tome project, and will build on this work by incorporating perceptual models into the interaction-sequence predictions.

Developing a unit-task-level model that is specific for visual analysis applications is significant in three ways. First, with such a model, the visualization designer can simulate task performance under alternative design choices, observe how the design choices affect the perceptual and cognitive operations involved in the task completion, and make informed design decisions. Second, a model at this level can help visualization designers obtain more comprehensive performance profiles for visual representations and compare them accordingly. Our proposed work will provide tools that let developers infer interaction patterns at a scale that is impractical to collect with conventional user studies. Finally, when incorporating existing perceptual theories into the proposed task model, we will evaluate and extend those theories by applying them to new types of visual representations, which will advance our understanding of perceptual processes.

b.2 Modeling Phases of Reasoning Our proposed work will create a reasoning-level model that predicts whether the user is performing exploratory, hypothesis-forming interactions or focused interactions aimed at finding evidence for or against an existing hypothesis. The model is significant because it predicts the user's high-level analysis goal, which can be used to support the user through software tools and to understand how she pursues her analysis. The model will make predictions from empirically observable view- and data-manipulation patterns exhibited by the users during analysis tasks on the order of tens of minutes, in what Newell calls the "rational band" of behavior [Newell, 1990].

We choose to predict users' high-level analysis goals from interactions because previous work has shown that human experts can manually recover users' reasoning processes from viewing interaction logs [Dou et al., 2009; Lipford et al., 2010]. Also, researchers have proposed several theoretical frameworks to account for the reasoning and sensemaking processes during visual analysis, which can be used as the basis for our model. However, automatic reconstruction of reasoning processes from interaction logs remains an unsolved problem, and none of the existing frameworks has been tested with empirical findings. The novelty of our work lies in that it aims to bridge the theoretical frameworks and the empirical observations to inform visual-analysis application designs broadly, using semi-automated analysis methods.

Several frameworks have been proposed to describe the high level phases a user goes through when making sense of data with visual analysis applications [Pirolli and Card, 2005; Zhang et al., 2008]. While these frameworks have been used to inform visualization designs on a high level, they are rarely supported by detailed empirical user observations, since they don't specify how each of the high level reasoning phases is manifested in user actions.

On the other hand, other researchers have taken a bottom-up approach and demonstrated that we can infer the high-level reasoning process from user interaction histories. For example, Dou et al. conducted a study with WireVis where human analysts were recruited to reconstruct WireVis users' reasoning processes from their interaction data [Dou et al., 2009]. Others have demonstrated that interaction logs can be used to help analysts recall their own reasoning processes post hoc [Lipford et al., 2010]. Ethnographic studies of analysts have also been used to identify distinct interaction strategies with visual analysis applications and tasks citepCziemkiewicz-2012-AGO, kang-2011-HCV. These studies suggest that it is feasible to distill high level reasoning processes from user observations.

Developing such a reasoning model can benefit visualization research in three ways. First, to develop such a model, we will perform formative studies to study how visual analysis users manipulate views and data to achieve intermediate analysis goals during the entire reasoning process, and the user data that we plan to collect during the studies have the potential to provide empirical support for and help advance current theories about reasoning and sensemaking. Second, the ability to pinpoint specific analysis phases and goals from user interaction histories will allow visualization designers to perform more detailed evaluation of a visual analysis application based on how well each type of analysis goals is supported. Finally, inferred intermediate analysis goals can inform the design of visual analysis components that help reduce the cognitive effort required by these goals, such as by automatically keeping track of and grouping data explored in each individual hypothesis testing phase.

b.3 Modeling Insight Characteristics from User Interactions Our proposed insight modeling work is significant because it will create an insight model that connects evidence of a person's interactions with a visualization system to *insight characteristics* that quantify her "aha!" moments during analysis, such as insight depth, insight diversity, and insight quantity. Insight characteristics have been useful in comparing visual analysis tools on the basis of meaningful analysis outcomes [Saraiya et al., 2005; O'Brien et al., 2010].

We propose to predict insight characteristics from user interactions because previous work has established qualitatively how users of visual analysis applications use interactions to arrive at insights [Yi et al., 2007; Gotz and Zhou, 2009]. However, there has not been any work that quantitatively analyzes the correlations between interaction patterns and insight generation. For our novel insight model, we will study interaction patterns during visual analysis that occurs on the order of tens of minutes to an hour. In recent years, insight-based evaluation has received increasing attention, since it allows visualization designers to evaluate an application based directly on how well it promotes insights, which is the ultimate design goal of visualizations [Card et al., 1999]. However, typical insight-based evaluation methodologies [Saraiya et al., 2005, 2006] require users to report insights using a think-aloud protocol, and these utterances are manually coded and quantified, which is costly and effortful for both the study participants and expert coders.

The work is important to visualization research in two ways. First, the formative studies that we perform to lead up to the model design will provide findings that deepen our understanding of users of visual analysis applications interact with the application to arrive at insights. Second, this will be a step towards a semi-automated insight-based evaluation methodology, and will let visualization designers more easily identify interaction history segments that are indicative of interface design inefficiencies and derive interface or interaction design improvements accordingly.

b.4 Individual Differences and Analysis Performance Our analyses of individual differences during these experiments will advance understanding of how user characteristics affect task performance on multiple levels of visual analytic activities. Recent visualization studies have shown that individual differences can be predictive of the strategies and analysis outcomes that people have when exploring information representations. Brown et al. found that a person's *locus of control* – their sense of being internally in control versus controlled by external factors – was predictive of whether they preferred a details-first or overview-first approach to visual search using Google Maps [Brown et al., 2014]. This work also demonstrated that a subset of a person's early interactions with a visualization can be predictive of the final task performance. Earlier work by Ziemkiewicz et al. showed that a person's locus of control can influence task performance with different visual representations of the same data [Ziemkiewicz et al., 2011]. Other studies outside the visualization literature have identified that a person's information-seeking preferences might be predicted by a trait called *cognitive reflection* [Fernbach et al., 2013] – a measure of one's tendency toward deliberative or intuitive thinking. Cognitive reflection could be predictive of exploratory behaviors during visual analysis, but to the best of our knowledge that has not yet been studied. We plan to incorporate the three- question cognitive reflection test [Frederick, 2005] into our questionnaires and provide a novel analysis of this individual difference in the context of scientific visual-analysis tasks. Our work will build on these results by considering how individual differences affect higher-level cognitive processes like reasoning and insight generation in addition to performance on basic visualization tasks. During our experiments, we will measure several individual differences that we hypothesize will impact analysis behaviors in our applications; this will let us systematically study how these differences interact to influence analysis outcomes. In addition, our work will establish a protocol for assessing individual differences in later visualization studies.

b.5 Visual Analysis of Brain Networks and Genomic Variation Our work will be driven by and grounded in its application to two real-world visual analysis applications. The first, BraiNet, displays brain connectivity in several ways, provides links out to evidentiary publications supporting each connection, and supports annotations that facilitate analysis. The second, MAGI, displays the varying genetic mutations that are present in different types of cancer, providing a mechanism for visually identifying patterns and correlations. Figs. 1 and 2 illustrate a portion of each interface. Both applications incorporate data that is difficult or impossible to mentally capture completely, and so represent the challenge of analyzing the rapidly expanding body of scientific data that is available. We believe that human visual analysis will remain a vital tool in interpreting, analyzing, and understanding the meaning hidden in this and other such data. Through these applications, we hope to accelerate advances in the scientific domains they represent in addition to evaluating the applicability of our methods and results.

Our proposed work will establish an agenda of developing descriptive and predictive models that relate

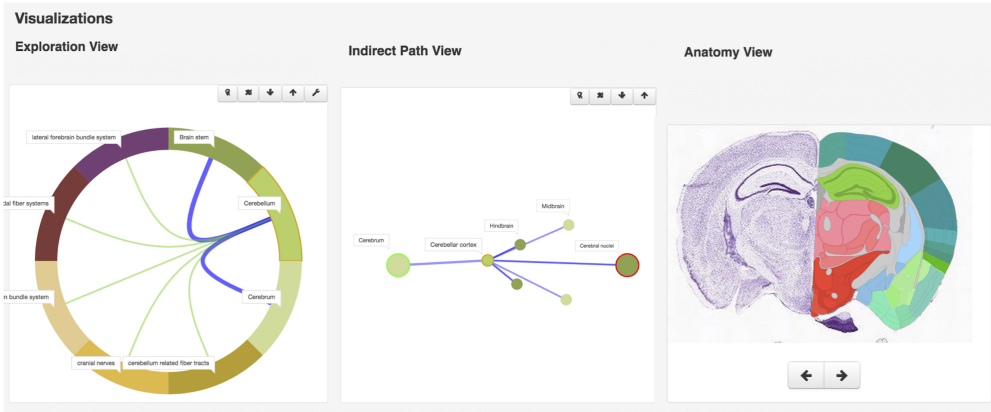

Figure 1: BraiNet, one of our driving scientific analysis applications, shows hierarchical, network, and anatomical views of connectivity in the brain. It also provides links to literature providing evidence for each connection. We are working with brain scientists to understand how combinations of these interactive views facilitate reasoning and hypothesis generation that inform experimental design for connectivity research.

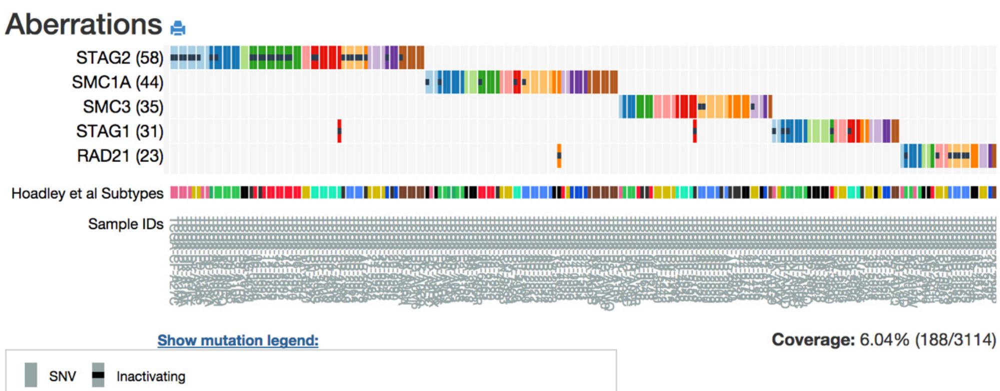

Figure 2: MAGI, our second driving scientific analysis application, displays colored cells representing how different gene mutations relate to cancer in a population. We have begun perceptual studies to choose mark size, coloring, and grouping parameters for the visualization that will help cancer genomicists identify patterns more easier.

analysis outcomes to information about human users, goals, and computer interfaces. Disseminating our models, tools, and software online will enable other researchers to integrate and extend the findings. To the best of our knowledge, we will be among the first to systematize the way cognitive theory is integrated with scientific visual analytics tools.

c Preliminary Results

Our preliminary findings demonstrate the feasibility of key components in this project. Our group is well positioned to conduct experiments applying cognitive and perceptual theories to driving problems with real visual analysis applications. We have developed initial software prototypes for the proposed scientific visual analysis applications, as well as infrastructure to support developing, maintaining, and disseminating software systems and data. Initial prototypes have been created for the two scientific applications, BraiNet [Guo et al., 2013] and MAGI, that will be used to collect data and will be improved using predictive models of visual representation quality and the user's reasoning process. These prototypes are the result of performing preliminary task analyses with collaborators in these domains.

Many of the research methods we propose in Sec. d have been used successfully in our past projects. We have tools in place for data collection, including annotating user interactions from video and monitoring keystroke- level interactions automatically [Gomez et al., 2014; Gomez and Laidlaw, 2012].

c.1 Modeling Unit Tasks and Analyzing Low-level Interactions We have completed some perceptual studies to better understand how design choices for marks in visual representations affect unit task performance. Recently, we studied how quickly and accurately participants were able to identify marks of interest in colored grids and scatter plots with different configurations of mark size, grid size, and heterogeneity of colors in groups of marks [Gramazio et al., 2014]. We performed regression analyses on the relationships between task response time and these visual variables, which will help guide creation of a predictive model of response time.

A relevant application for this work is improving the design of visual representations for cancer genomics analysis. We performed a case study with two pairs of cancer genomic researchers who analyzed datasets looking for individual targets and global trends using categorical heatmaps called *mutation matrices* that are commonly used in this domain. We created two variations with different mark sizes and hypothesized that the matrices with smaller marks would be more difficult to use, based on our earlier findings in [Gramazio et al., 2014]. Interestingly, we found that experts preferred different matrix conditions for the two different task types, which suggests that optimal visual parameters for visual search tasks might differ from those for high-level pattern detection. This finding demonstrates the need to understand how design parameters affect analysis behaviors in both low-level identification tasks, as we propose in Sec. d.1, and in tasks that require integrating multiple pieces of visual information into insights, as we propose in Sec. d.3.

Collecting and analyzing keystroke-level input is important part of understanding the motor steps for interactive visualization tasks and predicting the types of tasks performed. Toward that end, we created a semi-automated system, Tome, that analyzes discrete user interactions with visualization GUIs and predicts task completion times [Gomez and Laidlaw, 2012]. These predictions are useful in evaluating task performance for a crowd of end users and can be modified to evaluate how unimplemented GUI features affect task speed. Simulating tasks and predicting task speed is done by fitting interaction histories to a human performance model called the Keystroke-Level Model (KLM) [Card et al., 1980]. To do this, Tome interacts with the modeling tool CogTool [John et al., 2004] that simulates tasks on an interface design using a cognitive modeling architecture; this type of coupling of models with software interfaces is a specific aim of the proposed work. We have also conducted studies in which we developed a novel approach for analyzing mouse cursor data [Huang et al., 2011], inspired by eye-tracking data analysis methodologies. Those studies demonstrated that analysis of mouse cursor data can shed light on user behavior. Our experience using the KLM in the Tome project and analyzing and relating mouse cursor data to search behavior suggest that our plan to create a unit-task-level model based on keystroke-level interactions in Sec. d.1 is feasible.

c.2 Improving Visual Analysis Workflows At a higher level, we performed a preliminary GOMS analysis for the workflow of brain scientists who analyze EEG data using a combination of Matlab and a tool called EEGLab [Guo et al., 2012]. Our analysis let us identify unnecessary interactions in the existing workflow, and we used this to design a new visual analysis tool for EEG data called EEGVis aimed at reducing the cognitive load of the analysis workflow. Two expert users provided feedback, including responses on the NASA Task Load Index questionnaire, that suggests that EEGVis imposes much less cognitive load than the original workflow involving two tools, and that EEGVis encouraged more focused analysis than before. This experiment demonstrates that improving visual analysis tools using cognitive frameworks like GOMS is feasible and beneficial for domain scientists.

A goal of this work is to effect better analysis outcomes through improved human-computer interface design. We approached this recently by designing user interface components in a scientific visual analytics application that nudge users toward creating exploring alternative hypotheses and seeking more disconfirming evidence for hypotheses [Jianu and Laidlaw, 2012]. This work included a controlled study of variations on a protein signaling network visualization using a Google Maps-style interface. We found that small changes, e.g., redesigning the visual component that displays discovered evidence for hypotheses, resulted in statistically significant ($p < .05$) improvements in the amount of evidence gathered using the interface and how frequently users switched between competing hypotheses while investigating the data. This work provides a basis for understanding how interface designs can promote insights by analysts, which is one of our main goals in creating the insight-level model described in Sec. d.3.

c.3 Measuring Analysis Outcomes and Insights At the level of reasoning and insight formation, we have performed user studies in various science domains that compared how visual analytics tools support analysis workflows and outcomes. Working with researchers from Dr. Christophe Benoist's lab at Harvard, we performed a task analysis of visual exploration of gene expression data in the course of an immunobiology research project. This analysis uncovered patterns of interaction and important differences in behavior across users [Ziemkiewicz et al., 2012]. Specifically, experts used divergent strategies for interacting with multiple views of data. We will build on this finding by identifying how affordances of different user interface designs best support different workflows in the application areas of brain connectivity and cancer genomics.

In this project, measuring insight will help us evaluate how well the scientific applications we will improve and disseminate lead analysts to new discoveries. Our group has experience designing insight-based evaluations that quantify characteristics of hypothesis generation and discovery during visualization use in three previous studies. First, in the domain of genomics visualizations, we performed an insight-based evaluation comparing two interactive visualizations that showed genomic rearrangements using different spatial layouts [O'Brien et al., 2010]. Second, we performed an evaluation of both task performance and insight-promoting characteristics of interactive visualizations of spatiotemporal network data from the 2011 VAST Challenge [Gomez et al., 2014]. We found that laying out network information using some attributes of the data helped people understand the dataset better compared to a force-directed layout, even though it did not improve performance on structured tasks. These findings suggest that visual representations can shape mental models of relationships in a dataset and potentially affect insight formation; the findings apply to both proposed brain network and genomics applications, which will help scientists analyze biological network data. Finally, we found correlations between the frequency of certain types of interactions performed during exploratory analysis of data from the 2015 VAST Challenge and insight metrics like the number of facts that were identified during analysis [Guo et al., 2015]. This work will provide a foundation for studying the link between interactions with visualization systems and corresponding insight metrics in the domains of brain networks and genomics, as described in Sec. d.3.

c.4 Established Collaborations The project will benefit from existing collaborations between the PI's group and domain scientists studying cognition and perception, as well as driving problems in neuroscience and genomics analysis. The BraiNet prototype has been designed based on needs elicited from Dr. Mark Schnitzer's neuroscience group at Stanford and Dr. David Badre's cognitive neuroscience group at Brown. These scientists and members of their labs have participated in preliminary evaluations of BraiNet. The MAGI prototype has been designed in collaboration with Dr. Ben Raphael's computational biology group at Brown. The PI's group has worked with cognitive scientists Dr. Steven Sloman and Dr. Karen Schloss at Brown to study cognitive and perceptual aspects of task performance with visualizations. In addition, the PI's group has collaborated with Dr. Steven Sloman to develop the theoretical framework behind the reasoning model.

d Research Plan

We propose to develop models that predict metrics of task performance and reasoning. The models target visual analysis activities that represent typical scenarios for the brain and genomics analysis applications. These models correspond to activities that span different scales of complexity and time: a unit-task level

(**UM**), reasoning-phase level (**RM**), and insight level (**UM**). This section describes our research plan to conduct formative studies; design, fit, and evaluate the three models; and to apply the model outputs during the development of BraiNet and MAGI applications to support improved visual-analysis workflows. Milestones for each model are indicated by numbered labels (e.g, "(**UM2**)"); we propose a timeline for completing these in Sec. d.4.

d.1 Unit-task-level Model: Simulating task completion for unit visual-analysis tasks The goal of developing such a model is to computationally infer high-level user interactions and intentions from keystroke-level user actions, so that visualization researchers can conduct in-depth analyses that rely on understanding sequences of user interactions, such as those described in the reasoning model (Sec. d.2) and the insight-level model (Sec. d.3).

The model will take mouse and keyboard interaction logs as input and leverage established algorithms that summarize the data and help evaluators identify patterns corresponding to typical visual analysis tasks. We will demonstrate the utility of the model by implementing a novel analysis tool called BLADE (**B**etter **L**og **A**nalysis of **D**omain **E**xperts) that will be used to evaluate our proposed genomics application, MAGI. BLADE will allow visualization developers to conduct representative evaluations using only a small number of in-lab subjects through the aid of web-collected interaction log corpora.

Formative studies We will conduct a formative study to collect interaction data and task-performance data during unit-level tasks with MAGI (**UM1**). Before each session, researchers participating in the study will be asked to think of a series of gene sets they have looked at recently and would like to explore with MAGI. Once starting the session in-lab, each participant will be given a guiding task to generate hypotheses about the importance of each gene set and corresponding MAGI query results. Each session will conclude by participants self-annotating screencapture recordings for accurate analysis task deconstruction. These self-annotation results will then inform measuring insights for each participant.

In addition to screen captures and participant video annotations, we will also collect all user keyboard and mouse input in each session. We will also measure a number of perceptual and cognitive traits, including perceptual speed, visual memory, and locus of control, using tasks and questionnaires that have already been established to measure these traits [Goldberg et al., 2006; Ekstrom et al., 1976]. Once the data has been collected, we will analyze the captured video and action data to derive general rules that govern the series of perceptual and cognitive operations users performed to complete these tasks. We will perform statistical analysis over measured individual traits to assess their impacts on the sequence of operations they perform to complete the tasks.

Model design and fitting Data collected from the formative study will be used to inform the design and implementation of BLADE, a software prototype for analyzing mouse and keyboard level interaction logs (**UM2**). BLADE will support a number of quantitative measurements such as interaction sequence frequency using techniques like dynamic timeseries warping [Ratanamahatana and Keogh, 2004], and qualitative techniques such as heatmaps of mouse interactions similar to those used to analyze gaze in eye tracking studies [Jacob and Karn, 2003].

Model evaluation To evaluate the model, we will conduct two experiments to assess the usefulness of BLADE. The first experiment will test whether BLADE will improve insight identification in evaluation analysis by providing actionable, quantifiable analysis results. The evaluation will look at how four different researchers analyze MAGI interaction log data collected from the formative study: two will use just traditional insight-based evaluation methods, whereas the other two will also have access to BLADE. In the second experiment, we will collect a much larger interaction log corpus from in-the-wild use of MAGI. We will then examine whether BLADE will be able to identify pervasive individual and sequences of analytical tasks performed from that corpus through comparison against the interaction logs alongside ground-truth task sequences collected from the formative study (**UM3**). One potential risk in this project is that we will not attain a sufficient number of MAGI interaction logs from the formative study to perform powerful enough mining. Should this be the case, we instead propose to replicate the first evaluation study with a different group of researchers and the in-the-wild interaction logs, and then cross-validate results between the two studies.

Applying the model Results from our studies will be used to refine MAGI's layout with a novel layout manager that will suggest default layouts based on either an individual's or global trends in how frequently different visualizations are used together (**UM4**). Using this new layout manager we will then re-evaluate MAGI and compare the number of insights from the layout-intelligent MAGI against the results from our first study (**UM5**).

d.2 Reasoning Model: Predicting focused or exploratory analyst interactions

The goal of developing this model is to enable new visualization application features that reduce the effort required to complete complex, open-ended visual analysis tasks. These features will support improved workflows by factoring in the intermediate analysis goals of a user when displaying information. To this end, the model will predict a user's subgoals based on the sequence of analysis activities he performs.

At the core of the model is a two-phase framework for characterizing sensemaking behavior, which is developed based on existing sensemaking frameworks. One commonality shared by the existing frameworks (e.g., [Russell et al., 1993; Pirolli and Card, 2005; Zhang et al., 2008]) is that they all recognize two distinct search stages: the exploratory search stage, where the analyst forms and updates schemas, i.e. representations that captures the analysis goals, the features of the dataset, and relationships among subsets of data or data attributes of the problem space; and the focused search stage, where the analyst collects and organizes information to instantiate and sometimes update the schemas. Our two-phase framework states that an analyst loops through exploratory search and focused search phases to complete open-ended analysis tasks. We choose this two-phase framework for its simplicity and consistency with most of the existing sensemaking frameworks. More importantly, unlike some of the other sensemaking frameworks, this framework contains only observable states and can be empirically tested. The model will segment interaction histories into sequences that correspond to individual intermediate analysis goals and label each sequence as corresponding to one of the two phases.

We will demonstrate the utility of this model by developing and evaluating two application features that will use outputs from the model and improve visual-analysis workflows for scientists.

Formative studies We will conduct a formative study with the BraiNet prototype to collect interaction data and user-reported descriptions of participants' reasoning processes (**RM1**). We will recruit around 20 undergraduate and graduate students in cognitive science and neuroscience from Brown to participate in the study. Participants will be asked to complete open-ended sensemaking tasks, like exploring the metadata of a set of neuroscience literature to identify several promising research directions. We will instrument the prototype to capture high-level interaction data. The interaction data will include both visualization operations, such as search, filter, or select, and the target data of the operations. We will develop an interaction coding scheme using existing interaction taxonomies as a starting point and refine the scheme through open coding. In Table 1, we list the interaction taxonomies that we have surveyed and will use when developing the interaction coding scheme. We will capture self-reports of each participant's reasoning process by videotaping each session and asking the participant to watch the video afterwards and recall their analysis goals during each time interval. We will also measure the user's problem solving style using the assessment instrument developed by Treffinger et al. [Treffinger et al., 2008] and the user's locus of control. These data will be analyzed to check if individual difference plays a significant role in analysis behavior, and if so we will perform additional procedures, such as user stratification or introducing additional variables in data analysis, to control for individual difference.

Model design and fitting We will code the user-reported reasoning phases from data collected in the formative study using the two-phase sensemaking framework, which states that an analyst goes through two major phases when making sense of data: exploratory search (**ES**) and focused search (**FS**). In general,

Taxonomy	Interactions
Shneiderman (1996)	overview, zoom, filter, details-on-demand, relate, history, extract
Yi et al. (2007)	select, explore, reconfigure, encode, abstract, elaborate, filter, connect
Brehmer and Munzner (2013)	select, navigate, arrange, change, filter, aggregate, annotate, import
Schulz et al. (2013)	browse, search, elaborate, summarize, extract, abstract, gather, derive

Table 1: Interaction taxonomies that will be used to develop the coding scheme for collected interaction data.

these phases correspond to when a person is forming hypotheses or pursuing specific ones, respectively. This framework is developed based on existing literature in sensemaking [Russell et al., 1993; Pirolli and Card, 2005; Zhang et al., 2008] and creative thinking [Howard-Jones, 2002; Müller-Wienbergen et al., 2011; Pringle, 2011], and in consultation with Dr. Steven Sloman.

To fit the model, we will extract interaction sequences that temporally match either an exploratory phase or a focused search phase in the user-reported reasoning processes. We will group the extracted interaction sequences by phase and extract the following features from each interaction sequence: the transition among actions, the average similarity between data targeted by consecutive actions, and the frequency of different types of actions. Statistical analysis will be performed over these features to identify differences between sequences from the two groups. We will then devise rules accordingly for extracting sequences that correspond to each of the two phases. We will also perform statistical analysis over the user's problem-solving styles and locus of control to determine if any of the individual traits has a significant effect on the chains of actions a user performs for each of the two phases. If any of the traits shows a significant effect, we will adapt the rules to tolerate the variance induced by that trait. While fitting the model, Laidlaw's lab will also collaborate with Sloman's lab to further refine the theoretical framework given the empirical data (**RM2**).

Model evaluation We will run another study with a different group BraiNet users and collect interaction data along with user- reported reasoning processes similar as in the formative study (**RM3**). We will then use the model to predict the sequence of exploratory or focused phases that the user has gone through, and evaluate the prediction accuracy using the self-reported ground-truth data. If the model yields low prediction accuracy, we will perform additional qualitative analysis of data collected from both user studies to identify possible sources of variance that have not been captured by the model. We will then modify the predictive model, e.g., by re-classifying how logged interaction data affects the exploratory phase and the focused phase. In this case, we will run another test study to evaluate the modified model.

Applying the model We will integrate the reasoning model into BraiNet and MAGI and use the model to extract each segment of user interactions that correspond to each exploratory and focused phase during the sensemaking process. A user of BraiNet or MAGI may use the application to perform exploratory analysis and generate hypotheses, which is an example of sensemaking, and based on the two-phase sensemaking framework, the analyst will enter focused phase to gather evidence for each of the hypotheses that has emerged.

We will use the model predictions to implement an automatic evidence tracking component that helps reduce the effort that the user needs to spend to keep track of all the evidence (**RM4**). This interface component will organize the data that the user has explored in each focused phase into groups so that the user can easily review the evidence that has been gathered for each hypothesis. After the component is implemented and released with the applications, we will collect feedback from online users to evaluate how well the evidence tracking component supports hypothesis generation. We will also invite around four members from Badre's lab as a focus group and observe how they use the tool with the evidence tracking component (**RM5**). This will let us gain an in-depth understanding of whether the component is accurately capturing analysis stages and how well it supports different analysis scenarios.

Since the two-phase sensemaking framework predicts that analysts will achieve higher analysis quality by well-paced transitions between exploratory search and focused search, we will design an interface feature which embeds a focus+context view in the detailed view of BraiNet, which shows how a specific set of information relates to all available information and topics. We hypothesize that such an interface design will ease the transition from focused search into exploratory search and increase the amount of hypotheses come up by an analyst, as implied by the sensemaking framework. To evaluate the interface design, we will conduct a user study where participants are assigned into either the control condition (BraiNet without the proposed interface feature) or the treatment condition (with the proposed interface feature). We will use the model to measure the time spent in each of the two sensemaking phases and conduct statistical tests to analyze whether analysts spend more time in exploratory search and come up with more hypotheses in the treatment condition.

d.3 Insight-level Model: Analyzing user interaction histories to expand insight-based evaluation metrics

The goal of developing such a model is to reduce effort required in insight-based evaluation by connecting the dots between interaction histories and insight characteristics. To achieve this goal, the model will take a user's interaction with a visualization tool as input and predict insight-generation characteristics of the user. Visualization developers will be able to use this model to conduct evaluations of tools based on predicted insight-generation characteristics using interaction data without having to perform think-aloud insight-based evaluations, which are time and labor-intensive.

We will demonstrate the utility of the model through two applications. The first application is a visual interface that lets a visualization designer explore the connections among the raw interaction histories from BraiNet / MAGI, intermediate representations such as frequent patterns and transition diagrams, and insight metrics. This application is an example of tools that support novel evaluation paradigms based on the connections between user interaction patterns and insight metrics. The second application is an evaluation of BraiNet / MAGI based on interaction histories collected from remote users, and it exemplifies semi-automated evaluation of visualization tools enabled by predictions of insight characteristics.

Formative studies During the formative study from Sec. d.2, we will collect user-reported insights during the sensemaking tasks using a think-aloud protocol, then we will code insight characteristics from these utterances (**IM1**). Later, we will analyze these characteristics in conjunction with the corresponding interaction sequences participants performed. We will capture high-level actions, like "Filter documents by brain region", performed by users. We will code the insights into a set of characteristics based on previous insight-based evaluations and literature in creative ideation and information discovery. For example, we will adapt four characteristics identified in [Kerne and Smith, 2004]: fluency (quantity of insights), flexibility (number of different categories of insights), originality (statistical infrequency of insights), and practicality. We will administer a cognitive reflection test and a questionnaire that measures the user's locus of control.

Model design and fitting With the sequence of actions captured, we will abstract the actions using an interaction taxonomy developed using a procedure similar to what we will use for fitting the reasoning model (Sec. d.2). With the abstract actions, we will mine the interaction sequences to identify information-seeking patterns, such as representative action state transitions and action frequencies. We will then run a regression analysis to assess the correlation between each information-seeking pattern and each insight characteristic. We will test a number of hypotheses, including that the proportion of scent-finding behaviors will be positively correlated with insight flexibility. We will compare how well each abstraction reveals information-seeking patterns that are predictive of insight characteristics. We will then construct a model for predicting insight characteristics using the information- seeking patterns that can account for the most variance. We will also perform statistical analysis over individual difference factors captured earlier to assess if any of the individual difference factor has an impact on the user's insight characteristics and account for its variance in the model accordingly (**IM2**).

Model evaluation We will evaluate the model by conducting a second user study with BraiNet, recruiting around 20 senior undergraduate and graduate students studying cognitive science at Brown as participants (**IM3**). We will first predict insight characteristics using the captured interaction histories and evaluate the predictive power of the model. If the model turns out to yield low prediction accuracy, we will perform additional qualitative analysis on data collected from both user studies to identify possible sources of variance that have not been accounted for in the model and distill those into recommendations to guide future research.

Applying the model We will test two applications of the model to demonstrate its utility.

The first application will be a visual interface that facilitates data-driven analysis of user behavior and usage pattern in visualization evaluation by highlighting intermediate representation of interaction histories that are indicative of good or bad analysis quality (**IM4**). For example, if a visualization designer notices a user session with low value in a given insight metrics, she can use the visual interface to show the intermediate representations, e.g., action state transitions, that predict that insight value. She can then mouse over an action or a motif in those representations and highlight all actions or motifs of that type in the raw interaction history to perform more detailed analysis of possible causes that led to those actions or motifs. To evaluate the visual interface, we will recruit a small group (around 6) of visualization researchers to use the tool to evaluate BraiNet and MAGI given interaction logs and screen-captured videos collected during the two user studies. For comparison, another group of visualization researchers will be recruited to evaluate BraiNet and MAGI just by watching the screen-captured videos. We will compare the evaluation outcomes from the two groups to evaluate whether the visual interface makes it easier to diagnose interface design flaws and come up with interface improvements.

In the second application, we will collect interaction data from BraiNet and MAGI when they are used in the wild and apply the model to predict insight characteristics for the sessions collected (**IM5**). For sessions with low predicted insight characteristics, we will use the model to extract interaction history segments that exhibit information-seeking patterns that are negatively correlated with insight characteristics. We will then perform qualitative analysis of the extracted interaction history segments to examine if those patterns are caused by interface design decisions, and suggest interface improvements accordingly. After the improvements have been implemented, we will conduct an in-lab user study to assess the effect of the improvements (**IM6**).

d.4 Timeline Our plan is divided into four years and will be driven by multiple measurable milestones, e.g., model implementations, design reviews, and experiments. In general, we will begin by developing the scientific applications and sensemaking models, then we will fit the models using data collected from formative user studies. Finally, we will use the modeling results to improve the applications and then release these enhanced tools to domain scientists. Deliverables in the timeline are labeled based on their primary contribution: **UM** (unit-task-level model), **RM** (reasoning-level model), **IM** (insight-level model), or • (application development). The investigators responsible for each deliverable are also labeled.

Year 1 Deliverables

- • Baseline BraiNet and MAGI fully implemented and released [**Laidlaw Lab**]
- **UM1** Formative study with MAGI conducted; interaction data collected and analyzed [**Laidlaw Lab** and **Huang Lab**]
- **RM1** Formative study with BraiNet conducted; user interaction data and self-reported reasoning processes collected; reported reasoning processes coded using the sensemaking framework [**Laidlaw Lab**]
- **IM1** Formative study with BraiNet conducted; user interaction data and self-reported insights collected; insight data coded [**Laidlaw Lab**]

Year 2 Deliverables

- **UM2** BLADE implemented; interaction data collected from the lab study with MAGI analyzed [**Laidlaw Lab** and **Huang Lab**]
- **RM2** Interaction data from the formative study analyzed and distinguishing features for interaction segments from the two sensemaking phases identified; individual difference data analyzed and effects on sensemaking patterns identified; interaction segmentation rules established [**Laidlaw Lab** and **Sloman Lab**]
- **RM3** Second user study conducted with BraiNet; interaction data and self-reported reasoning processes collected; model evaluated using collected data [**Laidlaw Lab**]
- **IM2** Interaction and insight data from the formative study analyzed and correlations between interaction patterns and insights identified; individual difference data analyzed and effects on insight generation identified; regression model for insight prediction constructed [**Laidlaw Lab**]
- **IM3** Second user study conducted with BraiNet; interaction data and self-reported insight data collected; model evaluated using collected data [**Laidlaw Lab**]
- BraiNet and MAGI advertised to expert user bases, and in-the-wild interaction data is collected [**Laidlaw Lab**]

Year 3 Deliverables

- **UM3** MAGI interaction log corpus queried using BLADE to discover common analysis interaction trends [**Laidlaw Lab**]
- **UM4** MAGI Layout recommender system implemented based on log-analysis results from BLADE [**Laidlaw Lab**]
- **RM4** Evidence-tracking component and focus+context interface feature in BraiNet designed and implemented based on the model [**Laidlaw Lab**]
- **IM4** Visual interface for interaction-log analysis built and used to evaluate BraiNet and MAGI [**Laidlaw Lab**]
- **IM5** In-the-wild interaction data from BraiNet and MAGI analyzed to identify sessions with low predicted insight generation; qualitative analysis of corresponding interaction histories conducted and design improvements identified [**Laidlaw Lab**]

Year 4 Deliverables

- **UM5** MAGI with the layout recommender system evaluated in comparison with the baseline version with the default layout [**Laidlaw Lab**]
- **RM5** Evidence-tracking component and focus+context interface feature in BraiNet evaluated using both lab studies and online interaction data analysis [**Laidlaw Lab** and **Badre Lab**]
- **IM6** BraiNet and MAGI design refined based on evaluations performed using the interaction history analysis interface and insight predictions [**Laidlaw Lab**]
- Improved MAGI and BraiNet released [**Laidlaw Lab**]
- Model source code and APIs released [**Laidlaw Lab**]

e Broader Impact of the Proposed Work

The broader impact of the proposed work is that it will add to the foundation supporting future HCI research, add to the foundation supporting future cognitive and perceptual psychology research, accelerate advances made using our driving applications of brain network and cancer genome analysis, accelerate advances in other analytic domains, and help train future interdisciplinary researchers through the research itself and through courses offered by the PI (in the following Curriculum Development Activities, Sec. f).

Much of the potential for impact in HCI research is outlined in the significance section below. In psychology, the models and other findings may provide new hypotheses for how people reason and act in analytic situations.

Because our driving applications will be deployed for brain scientists and cancer genomicists to use, we hope that they will be able to leverage the applications to speed analysis of the vast amount of data in both areas and improve users' understanding of brain connectivity and of the genomics of cancer. The extent to which their efforts are accelerated is always difficult to measure precisely, but we believe that with better tools targeted at generating faster, more accurate, and more insightful analytic results, the acceleration will be significant.

This acceleration is also likely to generalize to other domains with analogous analytic needs. Numerous other areas of biology have data similar in scope and structure to cancer genomics data. Examples include the genomics of development, of the immune system, and of other pathologies; the signaling of proteins during biological processes; and the interaction of protein signaling and gene expression. Network analyses akin to the brain network analysis we will study are also prevalent in other domains, including social computing, intelligence analysis, and systems biology.

f Curriculum Development Activities

Beyond the direct training of students performing the proposed research, we will leverage the work by incorporating it into three courses. This will amplify the educational impact from a few students to several dozen. The research will add significantly to three computer science courses at Brown that link education and research. One, "Interdisciplinary Scientific Visualization" centers around designing and executing research projects by emulating the US model of research design, funding, and execution. Students identify a research problem with a collaborator from another discipline, explore potential solutions, write a "funding" proposal, peer review the set of proposals, do the research, write it up, and present it. This is all done during one 13 week semester. They get a taste of the excitement, challenge, and risk inherent to interdisciplinary research in a context where the real risk is minimal. This class will serve as a first line of outreach for our proposed work, broadening exposure from the handful of students directly involved as research assistants to a dozen or more each time it is taught. From past experience, we expect that some of these students will go on to participate actively in the proposed work or other research projects. A number of past students in this class are now in successful research careers.

A second course, "Cognition, Human-Computer Interaction, and Visual Analysis" is a research seminar focused on how humans and computers can interact effectively when performing scientific analysis. Students in the course complete semester-long research projects, which in the past have included user studies on visual search using touch screens; a design for EEG data visual analysis based on a task analysis [Guo et al., 2012]; and analyzing the effectiveness of using MTurk to perform quality control on MRI scans. These students will benefit by having access to a cutting-edge investigation in human-centered visualization happening on-site at Brown with interdisciplinary, complementary driving applications. We expect that students will have excellent opportunities to participate in larger, longer-term experiments than in the past by building off this research agenda and collaborating with the PI and research team. Results from past instances of this course comprise a significant portion of this proposal.

A third course that will benefit from this research is "Virtual Reality Design for Science." This course, jointly listed and taught at Brown and the Rhode Island School of Design, teaches design students enough science so that they can author new interactive tools for scientists. We plan to accelerate the process of evaluating these interfaces without going through the months-to-years implementation process, providing a demonstration of the acceleration our research will make possible.

g Results from Prior NSF Support

Laidlaw was a Co-PI on a collaborative award IIS-1016623, "GV: Small: Collaborative Research: Supporting Knowledge Discovery through a Scientific Visualization Language," $269K, 2010-2014. **Intellectual**

Merit: Publications include ([Gomez and Laidlaw, 2012; Gomez et al., 2012; Ziemkiewicz et al., 2012; Gomez et al., 2014] from Laidlaw's group at Brown along with several others from their collaborative groups. All are aimed at improving the scientific analysis of diffusion MRI data. **Broader Impact:** The broader impact will be in a better understanding of brains and of how they can effectively be studied as well as a better understanding of the principles underlying effective scientific visualization.

Huang has recently received NSF CISE Research Initiation Initiative Award #1464061 "CRII: CHS: Scalable Webcam Eyetracking by Learning from User Interactions" (2015-04-01 to 2017-03-31) for $175,000. **Intellectual Merit:** Substantial progress has been made towards the development of a baseline eye tracker for remote capture that uses user interactions to train the eye tracking model. **Broader Impacts:** This grant has been supporting a female graduate student, Alexandra Papoutsaki. No publications have been produced under this award yet.

Sloman and Badre have not had NSF support during the past five years.

h Summary

In summary, we propose a human-computer interaction (HCI) research agenda to advance predictive modeling of humans interacting with exploratory scientific visualization and analysis interfaces. We believe that it will advance HCI research by improving existing models, allowing for more principled evaluation of interfaces, and providing insights into the strengths and weaknesses of different levels and types of behavior modeling. Our work should improve the two driving applications: brain network analysis and cancer genome analysis. More broadly, we anticipate accelerating scientific progress in those application domains, training future HCI researchers, and providing behavioral and modeling data that other HCI researchers can build on.

Bibliography

John R. Anderson. Spanning seven orders of magnitude: a challenge for cognitive modeling. *Cognitive Science*, 26:85–112, 2002.

J. Bertin. *Semiology of Graphics*. University of Wisconsin Press, 1983.

E.T. Brown, A. Ottley, H. Zhao, Quan Lin, R. Souvenir, A. Endert, and R. Chang. Finding waldo: Learning about users from their interactions. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12): 1663–1672, Dec 2014. ISSN 1077-2626. doi: 10.1109/TVCG.2014.2346575.

Stuart K. Card, Thomas P. Moran, and Allen Newell. The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, 23(7):396–410, 1980.

Stuart K Card, Jock D Mackinlay, and Ben Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

Wenwen Dou, Dong Hyun Jeong, Felesia Stukes, William Ribarsky, Heather Richter Lipford, and Remco Chang. Recovering reasoning processes from user interactions. *IEEE Computer Graphics and Applications*, (3):52–61, 2009.

Ruth B. Ekstrom, John W. French, Harry H. Harman, and Diran Dermen. Manual from kit of factor-references cognitive tests. *Educational Testing Service (1976): Princeton, NJ.*, 1976.

Philip M. Fernbach, Steven A. Sloman, Robert St. Louis, and Julia N. Shube. Explanation fields and foes: How mechanistic detail determines understanding and preference. *Journal of Consumer Research*, 39(5): 1115–1131, 2013.

Shane Frederick. Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42, 2005.

Lewis R Goldberg, John A Johnson, Herbert W Eber, Robert Hogan, Michael C Ashton, C Robert Cloninger, and Harrison G Gough. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1):84–96, 2006.

Steven R. Gomez and David H. Laidlaw. Modeling task performance for a crowd of users from interaction histories. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI), Best Note Honorable Mention*, 2012.

Steven R. Gomez, Radu Jianu, Caroline Ziemkiewicz, Hua Guo, and David H. Laidlaw. Different strokes for different folks: Visual presentation design between disciplines. In *IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis)*, volume 18, pages 2411–2420, 2012.

Steven R. Gomez, Hua Guo, Caroline Ziemkiewicz, and David H. Laidlaw. An insight- and task-based methodology for evaluating spatiotemporal visual analytics. In *Proceedings of IEEE VAST*, 2014.

David Gotz and Michelle X Zhou. Characterizing users' visual analytic activity for insight provenance. *Information Visualization*, 8(1):42–55, 2009.

Connor Gramazio, Karen B. Schloss, and David H. Laidlaw. The relation between visualization size, grouping, and user performance. In *IEEE Transactions on Visualization and Computer Graphics*, volume 20, pages 1953–1962, 2014.

Wayne D. Gray, Bonnie E. John, and Michael E. Atwood. Project ernestine: Validating a goms analysis for predicting and explaining real-world task performance. *Human-Computer Interaction*, 8:237–309, 1993.

Hua Guo, Diem Tran, and David H. Laidlaw. Incorporating goms analysis into the design of an eeg data visual analysis tool. In *Proceedings of IEEE InfoVis (Posters)*, 2012.

Hua Guo, Arthur Yidi, Steven R. Gomez, Mark J. Schnitzer, David Badre, and David H. Laidlaw. Toward a visual interface for brain connectivity analysis. In *CHI EA '13 CHI '13 Extended Abstracts on Human Factors in Computing Systems*, pages 1761–1766, 2013.

Hua Guo, Steven R. Gomez, Caroline Ziemkiewicz, and David H. Laidlaw. A case study using visualization interaction logs and insight metrics to understand how analysts arrive at insights. VAST '15, 2015.

Paul A Howard-Jones. A dual-state model of creative cognition for supporting strategies that foster creativity in the classroom. *International journal of technology and design education*, 12(3):215–226, 2002.

Jeff Huang, Ryen W White, and Susan Dumais. No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1225–1234. ACM, 2011.

RJ Jacob and Keith S Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, 2(3):4, 2003.

Radu Jianu and David H. Laidlaw. An evaluation of how small user interface changes can improve scientists analytic strategies. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012.

Bonnie E. John and David E. Kieras. The goms family of user interface analysis techniques: Comparison and contrast. *ACM Trans. Comput.-Hum. Interact.*, 3(4):320–351, December 1996. ISSN 1073-0516. doi: 10.1145/235833.236054. URL `http://doi.acm.org/10.1145/235833.236054`.

Bonnie E. John, Konstantine Prevas, Dario D. Salvucci, and Ken Koedinger. Predictive human performance modeling made easy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 455–462, New York, NY, USA, 2004. ACM. ISBN 1-58113-702-8. doi: 10.1145/985692.985750. URL `http://doi.acm.org/10.1145/985692.985750`.

Andruid Kerne and Steven M Smith. The information discovery framework. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 357–360. ACM, 2004.

Mark D M Leiserson, Connor C Gramazio, Jason Hu, Hsin-Ta Wu, David H Laidlaw, and Benjamin J Raphael. Magi: visualization and collaborative annotation of genomic aberrations. *Nat Meth*, 12(6): 483–484, 06 2015. URL `http://dx.doi.org/10.1038/nmeth.3412`.

Heather Richter Lipford, Felesia Stukes, Wenwen Dou, Matthew E Hawkins, and Remco Chang. Helping users recall their reasoning process. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 187–194. IEEE, 2010.

Gerald Lee Lohse. A cognitive model for understanding graphical perception. *Human-Computer Interaction*, 8(4):353–388, 1993.

Felix Müller-Wienbergen, Oliver Müller, Stefan Seidel, and Jörg Becker. Leaving the beaten tracks in creative work–a design theory for systems that support convergent and divergent thinking. *Journal of the Association for Information Systems*, 12(11):714–740, 2011.

Allen Newell. *Unified Theories of Cognition*. Harvard University Press, Cambridge, MA, USA, 1990. ISBN 0-674-92099-6.

Trevor M. O'Brien, Anna M. Ritz, Benjamin J. Raphael, and David H. Laidlaw. Gremlin: An interactive visualization model for analyzing genomic rearrangements. *IEEE Trans. on Visualization and Computer Graphics (Proc. Information Visualization '10)*, 2010.

Daniel Pineo and Colin Ware. Neural modeling of flow rendering effectiveness. *ACM Trans. Appl. Percept.*, 7(3):20:1–20:15, June 2008. ISSN 1544-3558. doi: 10.1145/1773965.1773971. URL `http://doi.acm.org/10.1145/1773965.1773971`.

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*, volume 5, pages 2–4, 2005.

Andrew J Pringle. Shifting between modes of thought: a mechanism underlying creative performance? In *Proceedings of the 8th ACM conference on Creativity and cognition*, pages 467–468. ACM, 2011.

Chotirat Ann Ratanamahatana and Eamonn Keogh. Everything you know about dynamic time warping is wrong. In *Third Workshop on Mining Temporal and Sequential Data*. Citeseer, 2004.

Ruth Rosenholtz, Nathaniel R Twarog, Nadja Schinkel-Bielefeld, and Martin Wattenberg. An intuitive model of perceptual grouping for hci design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1331–1340. ACM, 2009.

Daniel M Russell, Mark J Stefik, Peter Pirolli, and Stuart K Card. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 269–276. ACM, 1993.

Purvi Saraiya, Chris North, and Karen Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):443–456, July 2005. ISSN 1077-2626. doi: 10.1109/TVCG.2005.53. URL `http://dx.doi.org/10.1109/TVCG.2005.53`.

Purvi Saraiya, Chris North, Vy Lam, and Karen A. Duca. An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1511–1522, November 2006. ISSN 1077-2626. doi: 10.1109/TVCG.2006.85. URL `http://dx.doi.org/10.1109/TVCG.2006.85`.

Donald J Treffinger, Edwin C Selby, and Scott G Isaksen. Understanding individual problem-solving style: A key to learning and applying creative problem solving. *Learning and Individual Differences*, 18(4): 390–401, 2008.

Ji Soo Yi, Youn ah Kang, John T Stasko, and Julie A Jacko. Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13 (6):1224–1231, 2007.

Pengyi Zhang, Dagobert Soergel, Judith L Klavans, and Douglas W Oard. Extending sense-making models with ideas from cognition and learning theories. *Proceedings of the American Society for Information Science and Technology*, 45(1):23–23, 2008.

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. *Proceedings of IEEE VAST*, 2011.

Caroline Ziemkiewicz, Steven R. Gomez, and David H. Laidlaw. Analysis within and between graphs: Observed user strategies in immunobiology visualization. In *ACM CHI*, 2012.