Project Description

CHS: Small: Cognitive and Perceptual Modeling for Exploratory Interactive Visual Analysis

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 reasoning phases using visual representation designs and user interactions as inputs. We will then improve visual analysis application designs using these predictions.

We will target three levels of cognitive activities with increasing complexity and time scales. On the first level, we will focus 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 a visual representation design and a set of visual analysis tasks, and predicts the time it takes for the user to complete each task. The second level involves 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 will involve insight formation that results from interacting with key components of a visual analysis application, and we will create a model that predicts insight characteristics such as insight diversity from the entire sequence of user interactions in a visual analysis session. We will capture individual difference factors that are likely to influence user behavior or performance at each level, assess their effects, and incorporate them into 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 predictions from our models. Improvements will take place in different forms, including refined visual parameters of marks, 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 risks 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 marks and layouts discovered after improving the scientific visual analysis applications; an analysis of 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 extracting values from individual visualizations, we will simulate task completions to predict user performance characteristics like response time. 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 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 choosing visual parameters for marks, like color, size, and grouping configurations. 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 layout configurations and interactions for UI components, and for constructing UI components that nudge users toward insight-promoting activities.

a.4 Quantitative evaluations of how some individual differences affect exploratory analysis behaviors with scientific visual analysis applications. We will build on recent work that suggests some measurable individual differences – like whether a person tends towards *deliberative* or *intuitive* reasoning during problem solving – are predictive of task performance in visual search tasks. Our findings might lead to design recommendations for the user interface that are specific to how a person thinks, and to 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 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 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. c).

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.

c 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.

d Significance

d.1 Modeling Interactions at the Unit-task Level This work is significant because it will create a novel model that can predict performance for a variety of basic visualization tasks given a visual representation design and task descriptions. Our experiments will validate the model on graph tasks. 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. Pineo and Ware built a predictive model of the effectiveness of flow rendering methods by simulating how visual marks excite neurons in the visual cortex [Pineo and Ware, 2008]. They found that the model could predict which streamline rendering method would let people perform a particle-advection visual-analysis task fastest. Lohse [Lohse, 1993] demonstrated that we can use UCIE (Understanding Cognitive Information Engineering), a computer program that simulates graphical perception, to predict task completion time for simple tasks with bar charts and line charts. A limitation of both the UCIE model and the neural model by Pineo and Ware 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, which computationally segments visual objects into visual groups, and they have demonstrated that the segmentation results are similar to human-perceived groupings on line charts. On a different note, multiple measures have been proposed for how "cluttered" an interface is, and Rosenholtz et al. [Rosenholtz et al., 2007] has assessed how well each measure predicts the time it takes for a subject to locate a visual object.

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 validations of using cognitive modeling to improve a real application workflow for telephone operators [Gray et al., 1993]. To do this, Gray et al. used the CPM-GOMS cognitive modeling framework and identified a critical path of interactions that could be improved to make telephone operator tasks more efficient. One challenge in using models in the GOMS family (*goals, operators, methods, selections*) is the expertise require to transcribe and 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. Unlike CPM-GOMS, KLM is based on a serial model of human information-processing [John and Kieras, 1996] which makes it easier to simulate but might not accurately predict visualization task components that are done in parallel, such as identifying target marks and characterizing high-level patterns in a dataset. Our work will mitigate this challenge by approaching visual analysis activities at three levels, rather than finding one model that fits the entire visual analysis process. As discussed in Sec. e, we have experience modeling low-level interactions for visual search tasks with KLM in our Tome project, and will build on this work by incorporating perceptual models into the time 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. The relative effectiveness of a visual representation sometimes depends on specific properties of the dataset to be visualized, such as total number of data points, number of data dimensions, or edge density for graph data. The proposed model can take in a visual representation and a number of test datasets with varying configurations, and generating task completion time predictions for each dataset with the given visual representation, which is impractical with conventional user studies. Finally, when incorporating individual graphical perception theories and models into the proposed task performance model, we will evaluate and extend those theories by applying them to visual representations that they have not been tested on before, which will advance our understanding of the corresponding perceptual processes.

d.2 Modeling Phases of Reasoning Our proposed work in modeling phases of reasoning is significant because it will create a reasoning-level model that predicts what phase the user is in during the sensemaking process. 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].

Some researchers have proposed theoretical models for the reasoning process during visual analysis, while others have reported empirical results on recovering reasoning processes from interactions. However, the empirical findings have rarely been connected to the proposed theoretical models. Also, the reconstruction of reasoning processes from interactions were entirely done by human analysts in the previous work. 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 et al., 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]. Lipford et al. took that one step further and demonstrated that interaction history logs can be used by visual analysis application to help users recall their own reasoning processes. Both Cziemkiewicz et al. [Ziemkiewicz et al., 2012] and Kang et al. [Kang et al., 2011] have reported different analysis strategies observed during evaluation of visual analysis applications. These studies suggest that we can distill high level reasoning processes from user observations, but since the reported analysis strategies are application- or domain-specific, the discussions and conclusions are often limited to environments similar to the ones in those user studies.

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.

d.3 Modeling Insight Characteristics from User Interactions Our proposed insight modeling work is significant because it will create an insight model that connects a person's interaction traces 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]. Previous work has discussed 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 allow visualization designers to more easily identify interaction history segments that are indicative of interface design inefficiencies and derive interface or interaction design improvements accordingly.

d.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; 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.

d.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

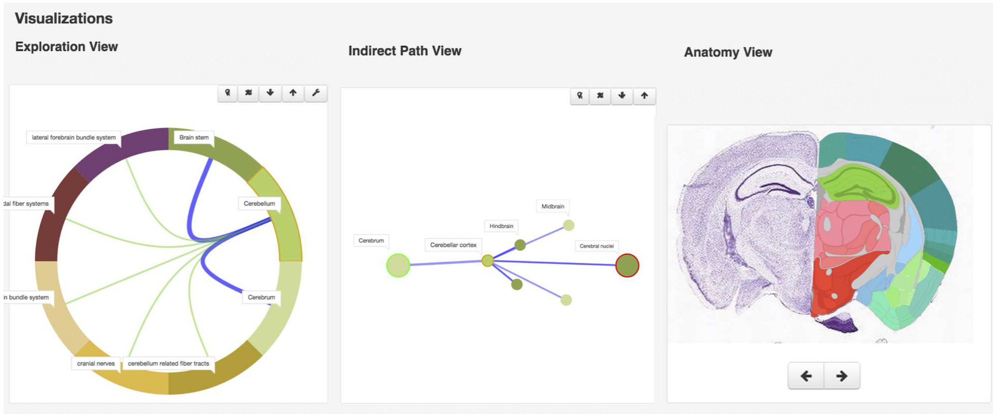

Figure 1: BraiNet, one of our driving scientific analysis applications, shows hierarchical, network, and anatomical views of connectivity in the brain. It also provides links to literature providing evidence for each connection. We are working with brain scientists to understand how combinations of these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

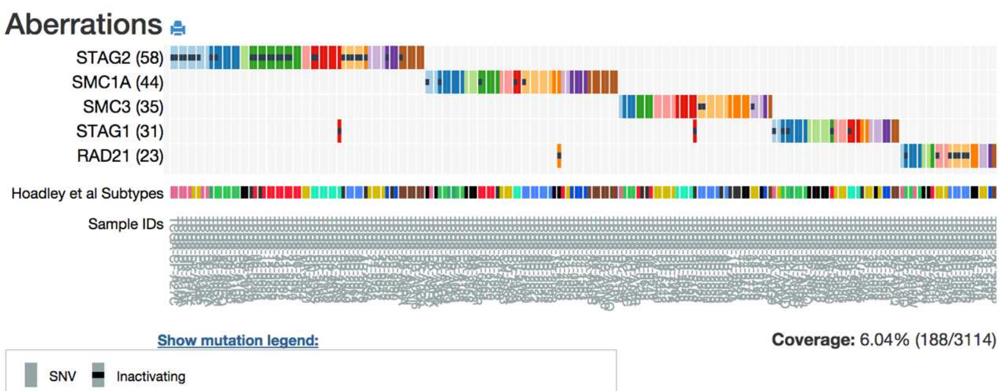

Figure 2: MAGI, our second driving scientific analysis application, displays colored cells representing how different gene mutations relate to cancer in a population. We have begun perceptual studies to choose mark size, coloring, and grouping parameters for the visualization that will help cancer genomicists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

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 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.

e 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. f 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].

e.1 Modeling Task Performance We have completed some perceptual studies to better understand how design choices for marks in visual representations affect 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. f.1, and in tasks that require integrating multiple pieces of visual information into insights, as we propose in Sec. f.3.

Collecting and analyzing keystroke-level input is important part of understanding the motor steps for interactive visualization tasks and the effect of those steps on task performance characteristics, like response time. 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.

Our experience using the KLM in the Tome project and early work with Lohse's performance model with bar and line charts [Lohse, 1993] suggests that our plan to create a perceptually-informed unit-task-level model in Sec. f.1 is feasible. Lohse's UCIE project describes an architecture for constructing the model that will inform our work.

e.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. f.3.

e.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 studies that quantify characteristics of hypothesis generation and discovery. In the domain of genomics visualizations, we performed an insight-based evaluation comparing two interactive visualizations that showed genomic rearrangements: a new method we created, GREMLIN, which uses a linear layout, and the state-of-the-art system Circos, which uses a circular layout [O'Brien et al., 2010]. We found that the linear design resulted in expert participants having more total insights, more hypothesis-driven insights, and more insights per minute with GREMLIN compared to Circos. Recently, 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; we plan to incorporate this relationship in constructing the reasoning model in Sec. f.

e.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.

f Research Plan

We propose to develop three models where each model targets visual analysis activities happening on a different scale of complexity and time: unit-task-level, reasoning-level, and insight-level. Here we describe the research plans to design and evaluate the three models and to apply them to inform the development of better visual analysis applications.

We will employ a five-stage workflow to design and evaluate each of the three models. In the first stage, we will conduct new formative studies in addition to the ones detailed in the Preliminary Work section to form qualitative observations and collect user data. In the second stage, we will construct the model, informed by both qualitative observations from our preliminary work, new formative studies, and existing theories. In the third stage, we will fit the model with data collected from the formative studies to determine specific model parameters. The fourth stage involves evaluating the model with users that differ from the ones recruited in the formative studies. Finally, we will apply the model to derive specific recommendations for improving BraiNet and MAGI, implement the recommended changes, and run in-lab studies with domain experts to evaluate the improvements.

f.1 Unit-task-level Model: Simulating task completion to generate performance prediction for unit visual-analysis tasks This model will take simple visual representations and unit tasks for these visual representations, and generate predictions for task completion time. The model will need to be instantiated for specific visual representations and tasks. Our goal is to instantiate the model for common graph tasks (e.g. find the shortest path between two nodes) and two types of graph visualizations – node-link diagrams and matrices – and have a model architecture that can be extended to other simple visual representations.

Formative studies We will first recruit participants to complete a set of graph tasks on a standard node-link diagram and a matrix visualization, respectively. Graph tasks used in the study will be based on the ones identified by Lee et al. [Lee et al., 2006]. During the study sessions, we will videotape and record user keyboard and mouse actions. 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. If a trait is shown to have a significant effect on the performance of perceptual or cognitive operations, we will include it as a parameter of the model if the effect can be quantified and use the trait to stratify participants if the effect is qualitative or categorical.

Model design and fitting To predict completion time for a task, the model will take three inputs: a specification of the visual representation rendering rules, the dataset, and the task description. Our model architecture will be similar to the architecture of UCIE, and use similar approaches to generate internal representations of visuals and transform tasks into sequences of perceptual and cognitive operations. However, our model will have a more sophisticated module for estimating the cost of each perceptual and cognitive operation. The cost estimations will be calculated using two resources. First, similar to the UCIE model, we will use a repository of cognitive engineering parameters, such as the time it takes to compare two colors in working memory or to make a saccade, as derived in previous works [Lohse, 1993]. Second, we will implement an ensemble of models that predict how specific visual properties influence the time it takes to complete certain perceptual or cognitive operations. We plan to incorporate two component models: a visual clutter model that predicts how the search time for a given visual object increases when the visual representation becomes more cluttered; and a perceptual grouping model that predicts visual object groupings and therefore how many visual groups the user needs to hold in the working memory when performing a task.

To determine parameters for the two component models, we will conduct another set of controlled studies. For these studies, we will create two sets of node-link diagrams and matrix visualizations. The first set of visual representations will be used to fit the perceptual grouping model, and we will vary 1) the number of data groups and 2) the visual encodings to distinguish data from different groups. The second set of visual representations will be used to fit the visual clutter model, and we will vary 1) the number of nodes, 2) the edge density, and 3) the number of data groups.

Model evaluation We will first evaluate the model using data collected from the formative studies and measure how well the predicted task completion time matches the ground truth completion time. We will then evaluate the model on node-link diagrams and matrix visualizations using datasets with configurations that differ from the ones used in the formative studies. We will vary graph size, edge density, and the number of data groups. This time, we will evaluate not only the prediction accuracy of the task completion time, but also how well the predicted operation sequences match with the actual sequences performed by the users. If the model yields low prediction accuracy, we will compare the predicted and actual operation sequences to identify perceptual and cognitive operations that have high uncertainty in completion time, and suggest areas in graphical perception that need further research accordingly.

Applying the model Both BraiNet and MAGI handle graph data, and one design decision to make is to choose the appropriate graph visualizations for various analysis tasks. To do so, we will first summarize the most important graph tasks that BraiNet and MAGI users need to perform respectively based on previous user interviews. We will then simulate task performance on a node-link diagram, a matrix, and a circular network diagram. We will use the simulation results to choose the graph visualizations for BraiNet and MAGI. Finally, we will implement the chosen visual representations in BraiNet and MAGI and evaluate their effectiveness through user feedback and usage data analysis.

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

Two important phases in the sensemaking process are the exploratory phase and the focused phase [Zhang et al., 2008]. This model will take as input a user's history of interactions at the level of analytic components (e.g., search, filter), and classify interaction sequences that correspond to exploratory or focused phases. These classifications will be used to design an interface component that helps analysts evaluate hypotheses by grouping and presenting data of interest that were analyzed during focused phases.

Formative studies We will first conduct a formative study with the BraiNet prototype to gather interaction data and user-reported reasoning processes. Participants will be asked to complete open-ended sensemaking tasks, such as to explore the metadata of a set of neuroscience literature and come up with 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. The naming and classification of the visualization operations will follow the interaction taxonomy proposed by Gotz and Zhou [Gotz and Zhou, 2009] to ensure generalizability. We will capture the user's reasoning process by videotaping each session and asking the participant to watch the video afterwards and recall what they were doing 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.

Model design and fitting We will code the user-reported reasoning phases from data collected in the formative study using the extended sensemaking model [Zhang et al., 2008], which describes the sensemaking process as iterations through four phases: *task analysis*, *exploratory phase*, *focused phase*, and *updates of knowledge representation*. We will extend the model to include additional phases if any emerge from the user-reported reasoning processes.

We will extract interaction sequences that temporally match either an exploratory phase or a focused phase in the user-reported reasoning processes. We will not include the task analysis phase or the knowledge representation update phase, because they are likely to involve few interactions with the visual analysis application. 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. We will then perform statistical analysis 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.

Model evaluation We will run another study with a different groups of users using BraiNet, and collect interaction data and user- reported reasoning processes similar as in the formative study. 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 ground truth data. If the model turns out to yield low prediction accuracy, we will perform additional qualitative analysis of data collected from both user studies to identify possible sources of variance that has not been captured by the model. We will then propose modifications, such as a classification of different approaches to the exploratory phase and the focused phase, to the extended sensemaking model upon which the model design is based.

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 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 the user will likely 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. 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 hypotheses. After the component is implemented and released with the applications, we will collect feedback from online users and conduct lab studies to evaluate how well the evidence tracking component supports hypothesis generation.

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

This model will connect the user's interaction history to his or her information-seeking behavior patterns to predict insight-generation characteristics.

Formative studies We will first conduct a formative study with BraiNet to collect interaction data and user-reported insights. Participants will be asked to complete open-ended sensemaking tasks, such as to explore the metadata of a set of neuroscience literature and come up with several promising research directions. We will capture high-level actions, such as "Filter documents by brain region," performed by a user. We will also ask users to think aloud during the study and code the reported insights. We will code the insights into a variety 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 two frameworks. Using an information- seeking framework [Pirolli et al., 2003], actions will be classified as either information scent-finding or scent-following actions. Using the interaction taxonomy proposed by Gotz and Zhou [Gotz and Zhou, 2009], actions will be classified as either data-exploration or visual-exploration actions. We propose to abstract actions using multiple frameworks because they capture different aspects of the semantics and utilities of the actions. When fitting and evaluating the model, all the analysis will be performed once for each abstraction. We will compare analysis results obtained using different abstractions and distill how well each abstraction reveals the correlation between information-seeking patterns and insights, and how they complement each other.

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.

Model evaluation We will evaluate the model by conducting a second user study with BraiNet utilizing a different group of users. 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. We will also evaluate the utility of the model for interface designs. We will perform qualitative analysis to see if we can locate interface designs that drive users towards or away from certain information seeking patterns and derive design recommendations accordingly.

Applying the model 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. 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.

f.4 Timeline

Our plan is divided into three phases to take place over three years. Each phase will be driven by multiple evaluatable milestones, e.g., model implementations, design reviews, and experiments. The timing is approximate; we expect the stages to overlap. We label the corresponding model for each deliverable using **UM** (our Unit-task-level model), **RM** (Reasoning-level model), and **IM** (Insight-level model).

Year 1 Deliverables

- BraiNet and MAGI prototypes fully implemented and released
- **UM 1**: Formative study with node-link diagrams and matrix visualizations conducted; user action and performance data collected; graph task completion rules coded
- **UM 2**: Controlled lab study conducted and parameters for the perceptual grouping model and visual clutter model determined
- **UM 3**: Model fully implemented and tested on a variety of datasets visualized using node-link diagrams and matrix visualizations
- **RM 1**: Formative study with BraiNet prototype conducted; user interaction data and self-reported reasoning processes collected; reported reasoning processes coded using the sensemaking framework
- **IM 1**: Formative study with BraiNet conducted; user interaction data and self-reported insights collected; insight data coded using the information-seeking framework and Gotz and Zhou's interaction taxonomy respectively

Year 2 Deliverables

- **UM 4**: Model applied on node-link diagrams, matrix visualizations, and circular network diagrams with sets of tasks tailored for BraiNet and MAGI respectively; network view designs chosen accordingly and implemented for the two applications
- **RM 2**: 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
- **RM 3**: The second user study ran with BraiNet; interaction data and self-reported reasoning processes collected; model evaluated using collected data

- **RM 4**: The evidence tracking component in BraiNet designed based on the model
- **IM 2**: 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
- **IM 3**: The second user study ran with BraiNet; interaction data and self-reported insight data collected; model evaluated using collected data
- BraiNet and MAGI interaction data collected

Year 3 Deliverables

- **RM 5**: The evidence tracking component in BraiNet evaluated using both lab studies and online interaction data analysis
- **IM 4**: MAGI interaction data collected and analyzed to identify sessions with low predicted insight generation; qualitative analysis of corresponding interaction histories conducted and design improvements identified
- MAGI design refined based on the insight analysis
- Model source code and APIs released

g Results from Prior NSF Support

Laidlaw is the PI on an active NSF award OCI-0923393, "MRI: Development of a Next-Generation Interactive Virtual-Reality Display Environment for Science" $2M, 2009-2015. **Intellectual Merit:** At this time, display construction almost completed, but there have been only a few oral presentations and one poster presentation about it. No peer-reviewed papers have been produced under this award yet. **Broader Impact:** Some of the broader impact will be in results from studies that the display enables. Additional impact will come from discoveries that will be made using the new display in fluid dynamics, biology, physics, and archaeology.

Laidlaw is also 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.

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.